



**UNIVERSIDAD AUTÓNOMA DE GUERRERO**

Unidad Académica de Matemáticas

---

---

Maestría en Matemáticas Aplicadas

**Estimación Bayesiana de los  
parámetros de la eficacia del  
Interferón en una infección viral  
*In Vitro***

**TESIS**

PARA OBTENER EL GRADO DE:

**Maestro en Matemáticas Aplicadas**

PRESENTA:

**Miguel Ángel Rodríguez Parra**

DIRECTORES DE TESIS:

M.C. Cruz Vargas de León

Dr. Flaviano Godínez Jaimes



Noviembre de 2021



---

## **Dedicatoria**

---

Dedico esta tesis especialmente a mis padres, que sé que desde el cielo están orgullosos de mi por haber culminado esta etapa de mi vida. También a mis hermanos que gracias a su apoyo desde la distancia me dieron las fuerzas para continuar y no rendirme en el camino. A toda mi familia que de una otra manera me apoyaron a lo largo de esta gran experiencia.

---

## Agradecimientos

---

A mis padres, por todos los sacrificios que hicieron para darme la oportunidad de estudiar.

A mi hermanos, porque siempre me apoyaron en los momentos más difíciles.

A mis amigos y compañeros, por todos sus consejos y su apoyo incondicional.

A mis asesores, por creer en mi y permitirme trabajar a su lado.

A mis profesores de la Maestría de Matematica Aplicada por sus consejos y valiosa enseñanza.

A la Maestría de Matemáticas Aplicadas de la Facultad de Matemáticas de la Universidad Autónoma de Guerrero por darme la oportunidad de realizar aquí mis estudios de posgrado.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT), por haberme brindado el apoyo económico necesario para realizar mis estudios de posgrado.

---

## Resumen

---

En virología, los modelos matemáticos basados en ecuaciones diferenciales han sido usado para describir la dinámica de las infecciones virales ya sea con o sin terapia antiviral y, a partir de ello, se usan como herramienta para monitorear la inhibición de la replicación viral y que conduzca a la cura de la infección o a mantener la carga viral por debajo de un umbral. En este trabajo se usa la estadística bayesiana para estimar los parámetros de dinámica celular, dinámica viral, de la eficacia del interferón (IFN), del número reproductivo básico y de otros parámetros importantes en modelos de infecciones virales, a partir de datos de cultivos celulares infectados con el Virus de Inmunodeficiencia Humana tipo 1 (VIH-1). Se presentan las estimaciones bayesianas de los modelos de crecimiento celular y la dinámica viral con o sin interferón, y se realiza la selección del mejor modelo para cada conjunto de datos. Los resultados obtenidos indican que los efectos de inhibición del interferón sobre la replicación viral y la producción de nuevas infecciones tienen la misma eficacia. Además, que el tratamiento con interferón reduce significativamente el número reproductivo básico y por ende la propagación de la infección.

---

## Abstract

---

In virology, mathematical models based on differential equations have been used to describe the dynamics of viral infections either with or without antiviral therapy and, based on this, are used as a tool to monitor replication inhibition viral and leading to cure of the infection or to keep the viral load below a threshold. In this work, Bayesian statistics are used to estimate the parameters of cell dynamics, viral dynamics, the efficacy of interferon (IFN), the basic reproductive number and other important parameters in viral infection models, from cell culture data. infected with Human Immunodeficiency Virus type 1 (HIV-1). Bayesian estimates of cell growth models and viral dynamics with or without interferon are presented, and the selection of the best model is made for each data set. The results obtained indicate that the inhibiting effects of interferon on viral replication and the production of new infections have the same efficacy. In addition, that interferon treatment significantly reduces the basic reproductive number and therefore the spread of the infection.

---

# Índice general

---

<b>1. Introducción</b>	<b>1</b>
1.1. Antecedentes . . . . .	1
1.2. Modelos de crecimiento celular . . . . .	4
1.2.1. Modelo de crecimiento de von Bertalanffy . . . . .	4
1.2.2. Modelo de crecimiento logístico . . . . .	4
1.2.3. Modelo de crecimiento de Gompertz . . . . .	5
1.3. Modelo básico de infecciones virales con dinámica celular . . . . .	5
1.4. Modelo básico de infecciones virales sin dinámica celular . . . . .	6
1.5. Modelo de infecciones virales con terapia antiviral . . . . .	7
1.6. Planteamiento del problema . . . . .	8
1.7. Objetivos . . . . .	9
1.7.1. Objetivo general . . . . .	9
1.7.2. Objetivos particulares . . . . .	9
<b>2. Marco teórico</b>	<b>10</b>
2.1. Estimación Bayesiana . . . . .	10
2.2. Cadenas de Markov Monte Carlo . . . . .	11
2.2.1. Integración Monte Carlo . . . . .	11
2.2.2. Cadenas de Markov . . . . .	12
2.3. Monte Carlo Hamiltoniano . . . . .	13
2.3.1. Algoritmo HMC . . . . .	13
2.4. Diagnósticos de convergencia de las cadenas MCMC . . . . .	14
2.4.1. Diagnósticos Gráficos . . . . .	15

2.4.2. Diagnóstico de Geweke . . . . .	16
2.4.3. Diagnóstico de Gelman-Rubin . . . . .	16
2.5. Estimador de Bayes . . . . .	17
2.6. Intervalos Creíbles . . . . .	18
2.7. Medidas de precisión predictiva . . . . .	19
2.8. Criterios de Selección de Modelos . . . . .	20
2.8.1. Criterio de Información de Watanabe-Akaike (WAIC) . . . . .	20
2.9. Método numérico de Runge-Kutta de cuarto orden . . . . .	21
<b>3. Metodología</b>	<b>22</b>
3.1. Datos . . . . .	22
3.2. Método . . . . .	22
3.3. Modelos Matemáticos . . . . .	24
3.4. Estimación Bayesiana . . . . .	26
3.4.1. Modelo Estadístico. . . . .	26
3.4.2. Función de Verosimilitud . . . . .	26
3.4.3. Información a priori . . . . .	27
3.4.4. Distribuciones a posteriori . . . . .	27
<b>4. Resultados</b>	<b>29</b>
4.1. Etapa 1: Datos de crecimiento Celular . . . . .	29
4.1.1. Modelo von Bertalanffy . . . . .	29
4.1.2. Modelo Logístico . . . . .	32
4.1.3. Modelo Gompertz . . . . .	35
4.1.4. Comparación de modelos . . . . .	38
4.2. Etapa 2: Datos de Carga Viral en ausencia de interferón . . . . .	39
4.2.1. Modelo von Bertalanffy . . . . .	40
4.2.2. Modelo Logístico . . . . .	44
4.2.3. Modelo Gompertz . . . . .	48
4.2.4. Comparación de modelos . . . . .	52
4.3. Etapa 3: Datos de Carga Viral en presencia de interferón . . . . .	53
4.3.1. Modelo von Bertalanffy . . . . .	54

4.3.2. Modelo Logístico . . . . .	58
4.3.3. Modelo Gompertz . . . . .	63
4.3.4. Comparación de modelos . . . . .	68
4.4. Número reproductivo básico y vida media de células infectadas. . . . .	69
<b>5. Conclusiones</b>	<b>72</b>
<b>Referencias</b>	<b>74</b>

---

## Índice de figuras

---

2.1. Traza de una cadena para un parámetro $\theta$ . . . . .	15
4.1. Trazas de las cadenas y densidades posteriores de los parámetros Modelo 3.1: von Bertalanffy. . . . .	31
4.2. Modelo 3.1: von Bertalanffy ajustado. . . . .	32
4.3. Trazas de las cadenas y densidades posteriores de los parámetros Modelo 3.2: Logístico. . . . .	34
4.4. Modelo 3.2: Logístico ajustado. . . . .	35
4.5. Trazas de las cadenas y densidades posteriores de los parámetros Modelo 3.3: Gompertz. . . . .	37
4.6. Modelo 3.3: Gompertz ajustado. . . . .	38
4.7. Comparación de Modelos Etapa 1. . . . .	39
4.8. Trazas de las cadenas y densidades posteriores de los parámetros Modelo 3.4: von Bertalanffy. . . . .	43
4.9. Ajuste Modelo 3.4: von Bertalanffy. . . . .	44
4.10. Trazas de las cadenas y densidades posteriores de los parámetros Modelo 3.5: Logístico. . . . .	47
4.11. Ajuste Modelo 3.5: Logístico. . . . .	48
4.12. Trazas de las cadenas y densidades posteriores de los parámetros Modelo 3.6: Gompertz. . . . .	51
4.13. Ajuste Modelo 3.6: Gompertz. . . . .	52
4.14. Comparación de Modelos Etapa 2. . . . .	53
4.15. Trazas de las cadenas y densidades posteriores de los parámetros Modelo 3.7: von Bertalanffy. . . . .	57

4.16. Ajuste Modelo 3.7: von Bertalanffy. . . . .	58
4.17. Trazas de las cadenas y densidades posteriores de los parámetros Modelo 3.8: Logístico. . . . .	62
4.18. Ajuste Modelo 3.8: Logístico. . . . .	63
4.19. Trazas de las cadenas y densidades posteriores de los parámetros Modelo 3.9: Gompertz. . . . .	67
4.20. Ajuste Modelo 3.9: Gompertz. . . . .	68
4.21. Comparación de Modelos Etapa 3. . . . .	69
4.22. Estimaciones puntuales e intervalo de alta densidad del número reproduc- tivo básico. . . . .	70
4.23. Estimaciones puntuales e intervalo de alta densidad de la vida media de las células infectadas. . . . .	71

## Introducción

---

### 1.1. Antecedentes

Los primeros trabajos realizados relacionados con el uso de modelos matemáticos para describir la dinámica del Virus de Inmunodeficiencia Humana (VIH) se remiten a la década de los 90, estos trabajos brindaron una comprensión inicial del comportamiento del virus.

En el año 1995, Ho y colaboradores describieron la dinámica de la infección viral en pacientes infectados con VIH Tipo 1 (VIH-1) bajo terapia antiviral, específicamente bajo tratamiento con inhibidores de proteasa. Sus estudios indicaron que este tratamiento hace que los niveles de VIH en plasma disminuyan exponencialmente y que los recuentos de células CD4 (glóbulos blancos que combaten infecciones) aumenten sustancialmente. Las estimaciones de la producción del virus y muerte de células objetivo indicaron que la replicación de VIH-1 es continua y altamente productiva (tiene una tasa alta) [1].

Un año mas tarde, Perelson y colaboradores, propusieron un modelo para analizar un conjunto detallado de datos de la carga viral del VIH-1, usaron datos recopilados de cinco individuos infectados después de la administración de un potente inhibidor de proteasa. La estimación de los parámetros del modelo se hizo mediante mínimos cuadrados no lineales. En este estudio se calculó la vida media de las células infectadas, la producción media de viriones por día, la duración media del ciclo de vida del virus y otros parámetros

que no solo proporcionaron una primera comprensión de la dinámica viral sino también principios teóricos para guiar el desarrollo de estrategias de tratamiento [2].

En ese mismo año, en 1996, Nowak y colaboradores presentaron el modelo básico de dinámica viral, un modelo simple para la interacción entre la replicación viral y las células huésped, este modelo se compone de un sistema de ecuaciones diferenciales no lineales con tres variables: células sanas, células infectadas y partículas virales; además describieron un segundo modelo que incluye respuestas inmunes contra las células infectadas, este modelo es una extensión del modelo básico en el que agrega una cuarta variable correspondiente a la magnitud de la respuesta de los Linfocitos Citotóxicos T (CLT), los cuales juegan un papel indispensable en la defensa antiviral atacando las células infectadas por el virus [3].

Posteriormente, en 1997 Bonhoeffer y colaboradores describieron la dinámica viral del VIH bajo terapia farmacológica al agregar al modelo básico un parámetro correspondiente a la eficacia de dos medicamentos: inhibidores de la transcriptasa inversa e inhibidores de la proteasa, los primeros previenen la infección de nuevas células, y los segundos, por otro lado, evitan que las células infectadas produzcan partículas de virus infecciosos. Además describe la dinámica viral con resistencia al medicamento mediante un modelo en el que se incluyen dos tipos de virus, el virus original y un virus que se vuelve resistente al medicamento (mutante). Así, este modelo está conformado por cinco variables: células sanas, células infectadas por el virus original, células infectadas por el virus mutante, virus original y virus mutante. Además, incluye un parámetro adicional para la tasa de cambio entre los dos tipos de virus [4].

Para el año 2002, Wodarz y Nowak, mostraron cómo se pueden usar los modelos matemáticos para comprender la dinámica de la infección y la terapia del VIH. En su trabajo describieron un modelo básico de infección por virus y mostraron cómo se usó para obtener algunas ideas cruciales sobre la dinámica durante la fase asintomática de la enfermedad. Exploraron cómo la evolución del VIH puede impulsar la progresión de la enfermedad, y cómo se pueden usar modelos matemáticos para diseñar tratamientos específicos que puedan aumentar la inmunidad antiviral e inducir el control del virus a largo plazo [5].

En el mismo año, Callaway y Perelson hicieron una revisión profunda de los modelos

de dinámica del VIH, desde el modelo más básico bajo terapia antiviral hasta modelos más complejos en los que hacen uso de heterogeneidades en la eficacia del fármaco. En su trabajo concluyeron que los modelos que mejor describen la terapia antiviral incluyen la tasa de eliminación de células infectadas como una función de su densidad [6].

También en el año 2002, Perelson hizo una revisión profunda sobre los modelos de dinámica del VIH con y sin terapia antiviral. Además presentó debates sobre la dinámica e interacciones con el Virus de la Hepatitis C (VHC), el Virus de la Hepatitis B (VHB), el Citomegalovirus (CMV) y el Virus de la Coriomeningitis Linfocítica (LCMV) [7].

Al revisar la literatura se encontraron trabajos relacionados con otros tipos de virus, como el VHC y la Influenza que valen la pena ser mencionados.

En 1998, Neumann y colaboradores estudiaron la dinámica del VHC y el efecto antiviral del interferón (sustancias producidas por los glóbulos blancos y otras células del cuerpo, pero también se pueden producir en el laboratorio para el tratamiento de diferentes enfermedades). Analizaron la disminución viral en 23 pacientes durante la terapia con un modelo matemático de ecuaciones diferenciales con tres variables, células sanas e infectadas y carga viral. El modelo empleado es una extensión del modelo básico en el que se incluye el efecto del interferón. Sus hallazgos mostraron que la infección con VHC es altamente dinámica y que el monitoreo temprano de la carga viral puede ayudar a guiar la terapia [8]. En el año 2007, Dahari y colaboradores extendieron el modelo formulado originalmente por Neumann y colaboradores en 1998, al incluir en él la proliferación de hepatocitos [9].

Baccam y colaboradores en el 2006, utilizaron una serie de modelos matemáticos de complejidad creciente, que incorporan la limitación de células objetivo y la respuesta innata de interferón, para examinar la cinética del virus de la influenza A en las vías respiratorias superiores de adultos infectados experimentalmente. Usaron regresión no lineal y estimación por mínimos cuadrados para ajustar los modelos a un conjunto de datos que provienen de un estudio de infección experimental del virus de la gripe porcina [10].

Dos años después Beauchemin y colaboradores analizaron la dinámica de una infección viral de influenza A (H3N2), utilizando un conjunto de modelos matemáticos que

destacan las diferencias entre la infección in vivo e in vitro. Utilizaron variaciones de los modelos propuestos por Baccam y colaboradores en 2006 y mediante métodos de estimación de parámetros no lineales, describieron la cinética de la dinámica del virus y los efectos de la dosificación del fármaco sobre la infección utilizando un conjunto de datos experimentales [11].

## 1.2. Modelos de crecimiento celular

### 1.2.1. Modelo de crecimiento de von Bertalanffy

La ecuación diferencial de von Bertalanffy viene dada por la expresión:

$$P'(t) = r(k - P(t)), \quad (1.1)$$

donde  $P(t)$  es el tamaño de la población al tiempo  $t$ ,  $P'(t)$  es su tasa de cambio,  $r$  la tasa de crecimiento intrínseca y  $k$  la capacidad de carga del medio, es decir, el número máximo de individuos que la población admite. Esta ecuación puede resolverse analíticamente y sus soluciones vienen dadas por:

$$P(t) = k - Ae^{-rt}, \quad (1.2)$$

donde  $A$  es un número real positivo que depende de la población inicial  $P(0)$ ,  $A = k - P(0)$

### 1.2.2. Modelo de crecimiento logístico

La ecuación diferencial logística o de Verhulst está dada por la expresión:

$$P'(t) = rP(t)\left(1 - \frac{P(t)}{k}\right), \quad (1.3)$$

donde  $P(t)$  es el tamaño de la población al tiempo  $t$ ,  $r$  la tasa de crecimiento y  $k$  la capacidad de carga del medio. Las soluciones de esta ecuación vienen dadas por:

$$P(t) = \frac{k}{1 + Ae^{-rt}}, \quad (1.4)$$

donde  $A = \frac{k}{P(0)} - 1$  y  $P(0)$  es el tamaño de la población inicial.

### 1.2.3. Modelo de crecimiento de Gompertz

La ecuación diferencial de Gompertz está dada por la expresión:

$$P'(t) = rP(t) \ln \left( \frac{k}{P(t)} \right), \quad (1.5)$$

donde  $P(t)$  es el tamaño de la población al tiempo  $t$ ,  $r$  la tasa de crecimiento intrínseca y  $k$  la capacidad de carga del medio. Esta ecuación puede resolverse explícitamente y sus soluciones vienen dadas por:

$$P(t) = ke^{-Ae^{-rt}}, \quad (1.6)$$

donde  $A = \ln \left( \frac{k}{P(0)} \right)$ , es un número real positivo que desplaza el modelo a la izquierda o derecha y  $P(0)$  es el tamaño de la población inicial.

## 1.3. Modelo básico de infecciones virales con dinámica celular

Cuando una célula es infectada por algún tipo de virus, ésta libera partículas virales que infectan a células sanas, a medida que las células se infectan con el virus, abandonan el grupo de células sanas y se unen a la población de células infectadas. Esta dinámica viral puede describirse con el Modelo Básico de Infecciones Virales, un modelo simple que involucra tres variables: células no infectadas, células infectadas y partículas virales liberadas. El Modelo Básico de Infecciones Virales está dado por el siguiente sistema de ecuaciones diferenciales [3, 4]:

$$\begin{aligned} x'(t) &= \lambda - \mu_x x(t) - \beta x(t)v(t), \\ y'(t) &= \beta x(t)v(t) - \mu_y y(t), \\ v'(t) &= py(t) - \gamma v(t). \end{aligned} \quad (1.7)$$

donde,

$x(t)$  = Concentración de células no infectadas.

$y(t)$  = Concentración de células infectadas.

$v(t)$  = Concentración de partículas virales libres.

$\lambda$  = Tasa de proliferación de células sanas.

$\mu_x$  = Tasa de muerte de células sanas.

$\beta$  = Tasa de infección viral.

$\mu_y$  = Tasa de muerte de células infectadas.

$p$  = Tasa de replicación del virus.

$\gamma$  = Tasa de declinación del virus.

El tiempo de vida medio de las células infectadas y el virus liberado se calcula por  $(\ln 2)/\mu_y$  y  $(\ln 2)/\gamma$ , respectivamente. El número promedio de partículas de virus producidas durante la vida útil de una sola célula infectada es  $p/\mu_y$  [7].

Antes de la infección, tenemos que  $y = 0$ ,  $v = 0$  y las células no infectadas están en equilibrio,  $x_0 = \lambda/\mu_x$ . Denotando como  $t = 0$ , al momento en el cual ocurre la infección [7, 5].

Para cualquier modelo de infección, el número promedio de células infectadas por una célula infectada típica durante su vida útil al comienzo de la infección se denomina número reproductivo básico,  $R_0$ . Para el modelo básico, el número reproductivo básico está dado por:

$$R_0 = \frac{\lambda\beta p}{\mu_x\mu_y\gamma}$$

Si  $R_0 < 1$ , entonces el virus no se propagará, ya que cada célula infectada producirá en promedio menos de una nueva célula infectada y, por tanto, el sistema regresará al estado de no infección donde  $x_0 = \lambda/\mu_x$ ,  $y_0 = 0$ , y  $v_0 = 0$ .

Si, por el contrario,  $R_0 > 1$ , cada célula infectada producirá en promedio más de una nueva célula infectada. La reacción en cadena generará una multiplicación explosiva de virus. El crecimiento del virus no continuará indefinidamente, porque el suministro de células no infectadas es limitado. Habrá un pico en la carga de virus y luego tenderá a estabilizarse [3, 5, 4, 12].

## 1.4. Modelo básico de infecciones virales sin dinámica celular

A diferencia del modelo anterior, en este modelo no existe dinámica poblacional de las células susceptibles, es decir,  $\lambda = 0$  y  $\mu_x = 0$ . Por tanto, para este modelo el sistema de

ecuaciones diferenciales es:

$$\begin{aligned}x'(t) &= -\beta x(t)v(t), \\y'(t) &= \beta x(t)v(t) - \mu_y y(t), \\v'(t) &= py(t) - \gamma v(t).\end{aligned}\tag{1.8}$$

para este modelo, el número reproductivo básico está dado por:

$$R_0 = \frac{\beta p T_0}{\mu_y \gamma}$$

donde  $T_0$  es el número inicial de células susceptibles, es decir, el número de células susceptibles, antes de la infección [10].

## 1.5. Modelo de infecciones virales con terapia antiviral

Este modelo es una extensión del modelo básico con dinámica celular (Ecuación 1.7) en el que se agrega el efecto de los medicamentos antivirales que bloquean la infección y la producción de partículas virales. El modelo está dado por [4, 8]:

$$\begin{aligned}x'(t) &= \lambda - \mu_x x(t) - (1 - \eta)\beta x(t)v(t), \\y'(t) &= (1 - \eta)\beta x(t)v(t) - \mu_y y(t), \\v'(t) &= (1 - \epsilon)py(t) - \gamma v(t).\end{aligned}\tag{1.9}$$

donde,

$\eta$  = Efecto de los medicamentos antivirales que bloquean la infección.

$\epsilon$  = Efecto de los medicamentos antivirales que bloquean la producción de partículas virales.

Los valores de  $\eta$  y  $\epsilon$  varían entre 0 y 1, dependiendo de la eficacia del medicamento. Si  $\eta = 0$  y  $\epsilon = 0$ , quiere decir que el fármaco no tiene efecto, mientras que si  $\eta = 1$  y  $\epsilon = 1$ , significa que el tratamiento es 100% efectivo (fármaco perfecto) [13]. Antes del tratamiento,  $\eta = \epsilon = 0$ . Una vez que se inicia el tratamiento,  $\eta > 0$  y/o  $\epsilon > 0$  [8].

## 1.6. Planteamiento del problema

Las enfermedades virales infecciosas son un problema de salud mundial. Para ampliar el desarrollo de medicamentos y tratamientos para contrarrestar estas enfermedades es de gran importancia la comprensión de la dinámica de la infección viral, es decir, conocer el comportamiento del crecimiento del virus en sus distintas etapas.

La modelación matemática ha demostrado ser una herramienta valiosa para estudiar problemas en distintas áreas del conocimiento como la medicina, la biología, la fisiología, la bioquímica, la epidemiología, la economía, entre otras; sus objetivos primordiales son describir, explicar y predecir fenómenos y procesos en dichas áreas.

En virología, los modelos matemáticos son usados para describir la dinámica de las infecciones virales ya sea con o sin tratamiento y, a partir de ello, se usan como herramienta para tomar decisiones de prevención y erradicación de enfermedades.

Generalmente se ha usado el método de mínimos cuadrados para estimar los parámetros de estos modelos. Una desventaja de este método es que únicamente se puede hacer estimación puntual. La estadística bayesiana, por otro lado, no presenta este inconveniente ya que permite determinar las distribuciones de los parámetros de interés y, por tanto, se pueden calcular probabilidades, intervalos de credibilidad y pruebas de hipótesis para los parámetros de interés o para funciones de los parámetros de interés.

Otra ventaja que tiene el uso de estadística bayesiana es que además de la información de la muestra, utiliza información previa del parámetro esto hace que sea viable trabajar con muestras pequeñas.

En este trabajo se usará la estadística bayesiana para estimar los parámetros de dinámica celular, dinámica viral y de la eficacia del interferón en modelos ajustados a un conjunto de datos experimentales de infección por el Virus de Inmunodeficiencia Humana tipo 1 (VIH-1).

## **1.7. Objetivos**

### **1.7.1. Objetivo general**

Estimar los parámetros de la eficacia del interferón en modelos de infecciones virales *in vitro* bajo el enfoque Bayesiano.

### **1.7.2. Objetivos particulares**

- Estimar los parámetros de la dinámica celular en modelos de infecciones virales.
- Estimar los parámetros de la dinámica viral en modelos de infecciones virales.
- Estimar los parámetros de la eficacia de los interferones en modelos de infecciones virales.
- Seleccionar el modelo que se ajuste mejor a datos de crecimiento celular y viral.

### Marco teórico

---

#### 2.1. Estimación Bayesiana

La inferencia bayesiana consiste en ajustar un modelo de probabilidad a un conjunto de datos y resumir el resultado mediante una distribución de probabilidad en los parámetros del modelo y en cantidades no observadas, como predicciones para nuevas observaciones.

La inferencia bayesiana sobre un vector de parámetros  $\theta$  o datos no observados,  $\tilde{D}$ , se hacen en términos de probabilidades. Estas probabilidades están condicionadas a los datos observados, y se escriben como  $\pi(\theta|D)$  y  $\pi(\tilde{D}|D)$ , respectivamente.

Para hacer declaraciones de probabilidad sobre  $\theta$  dado un conjunto de datos observados  $D$ , se debe tener un modelo de distribución de probabilidad conjunta para  $\theta$  y  $D$ . La función de probabilidad conjunta se puede escribir como un producto de dos densidades, *distribución a priori*  $\pi(\theta)$  y *distribución de los datos*  $\pi(D|\theta)$ , esto es:

$$\pi(\theta, D) = \pi(\theta)\pi(D|\theta)$$

A la distribución de probabilidad del vector de parámetros  $\theta$  dados los datos, se le conoce como *distribución a posteriori*  $\pi(\theta|D)$ . Usando la propiedad básica de probabilidad condicional conocida como regla de Bayes, la distribución o priori se puede expresar de

la siguiente manera:

$$\pi(\boldsymbol{\theta}|\mathbf{D}) = \frac{\pi(\boldsymbol{\theta}, \mathbf{D})}{\pi(\mathbf{D})} = \frac{\pi(\boldsymbol{\theta})\pi(\mathbf{D}|\boldsymbol{\theta})}{\pi(\mathbf{D})} \quad (2.1)$$

donde  $\pi(\mathbf{D}) = \int \pi(\boldsymbol{\theta})\pi(\mathbf{D}|\boldsymbol{\theta}) d\boldsymbol{\theta}$  y se le conoce como constante de normalización, pues no depende de  $\boldsymbol{\theta}$ , entonces (2.1) se puede escribir como:

$$\pi(\boldsymbol{\theta}|\mathbf{D}) \propto \pi(\boldsymbol{\theta})\pi(\mathbf{D}|\boldsymbol{\theta}) \quad (2.2)$$

De modo que para obtener la distribución a posteriori únicamente se necesita la distribución a priori y la distribución de los datos o función de verosimilitud. A partir de  $\pi(\boldsymbol{\theta}|\mathbf{D})$  se hace la inferencia sobre  $\boldsymbol{\theta}$ .

## 2.2. Cadenas de Markov Monte Carlo

A partir de la distribución a posteriori se pueden calcular para  $\boldsymbol{\theta}$  cantidades de interés como momentos, cuantiles, probabilidades, etc. Todas estas cantidades se pueden expresar como esperanzas posteriores de funciones de  $\boldsymbol{\theta}$ . La esperanza posterior de una función  $f(\boldsymbol{\theta})$  es:

$$E[f(\boldsymbol{\theta})|\mathbf{D}] = \frac{\int f(\boldsymbol{\theta})\pi(\boldsymbol{\theta})\pi(\mathbf{D}|\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int \pi(\boldsymbol{\theta})\pi(\mathbf{D}|\boldsymbol{\theta}) d\boldsymbol{\theta}}$$

En ocasiones estas integrales no se pueden calcular analíticamente, por tanto se deben usar métodos computacionales de aproximación. Uno de ellos es el método Monte Carlo en el que se generan muestras aleatorias independientes e idénticamente distribuidas para aproximar integrales, sin embargo, cuando las distribuciones son complejas no es posible muestrear directamente, es aquí donde se usan las Cadenas de Markov. Esencialmente los métodos de Cadenas de Markov Monte Carlo (MCMC) consiste en la integración Monte Carlo usando Cadenas de Markov.

### 2.2.1. Integración Monte Carlo

La integración Monte Carlo aproxima  $E[f(X)]$  extrayendo muestras  $\{X_t, t = 1, \dots, n\}$  de una función de distribución de probabilidad  $\pi(\cdot)$ . Esta aproximación es:

$$E[f(X)] \approx \frac{1}{n} \sum_{t=1}^n f(X_t)$$

Cuando las muestras  $\{X_t\}$  son independientes, la ley de los grandes números asegura que esta aproximación puede hacerse tan precisa como se desee incrementando el tamaño muestral  $n$  [14].

## 2.2.2. Cadenas de Markov

Las cadenas de Markov son procesos estocásticos donde la probabilidad de un estado depende únicamente del estado inmediatamente anterior [15]. Es decir, es una secuencia de variables aleatorias  $X_1, X_2, \dots, X_t, X_{t+1}$  donde el valor de  $X_t$  es el estado del proceso en el tiempo  $t$ ; esta secuencia de variables cumplen con:

$$P[X_{t+1}|X_t, X_{t-1}, X_{t-2}, \dots] = P[X_{t+1}|X_t]$$

A  $P[\cdot|\cdot]$  se le conoce como *kernel de transición*. Cuando estas probabilidades no dependen de  $t$ , se denominan probabilidades homogéneas.

Para que una cadena converja a una distribución fija  $\phi(x)$ , aquí llamada *distribución estacionaria*, deben satisfacer las siguientes propiedades:

1. Irreducibilidad. Es decir, para todo punto inicial de la cadena, la cadena puede alcanzar un conjunto de estados cualquiera con probabilidad positiva en un número finito de iteraciones. De manera más formal, una cadena de Markov se denomina irreducible si para todo  $i, j$ , existe  $t > 0$ , tal que  $P_{ij}(t) > 0$ , donde  $P_{ij}(t)$  es la matriz de transición de la cadena de un estado  $i$  a un estado  $j$ .
2. Aperiodicidad. Es decir, la cadena no puede estar oscilando en un número finito de iteraciones entre dos estados.
3. La cadena debe ser positivo-recurrente. Es decir, si un valor de la cadena es de la distribución estacionaria  $\phi(x)$ , entonces la cadena convergerá a dicha distribución.

## 2.3. Monte Carlo Hamiltoniano

Los problemas de las bajas tasas de aceptación de las propuestas de técnicas de Metropolis y el bajo rendimiento del algoritmo de Gibbs en problemas multidimensionales llevaron al surgimiento de una nueva técnica MCMC utilizando dinámica hamiltoniana (en honor al físico irlandés William Rowan Hamilton (1805-1865)). El nombre de esta técnica es Monte Carlo Hamiltoniano (HMC).

HMC es una adaptación de la técnica de Metropolis y emplea un esquema guiado para generar nuevas propuestas: esto mejora la tasa de aceptación de la propuesta y, en consecuencia, la eficiencia. Más específicamente, HMC utiliza el gradiente logarítmico posterior para dirigir la cadena de Markov a regiones de mayor densidad posterior, donde se recolectan la mayoría de las muestras. Como resultado, una cadena de Markov con el algoritmo HMC bien ajustado aceptará propuestas a un ritmo mucho más alto que el algoritmo tradicional de Metropolis. Metropolis tiene una tasa de aceptación de aproximadamente el 23 %, mientras que HMC tiene un 65 % [16]. Además de explorar mejor la distribución posterior, HMC es mucho más eficiente que Metropolis y no sufre el problema de correlación de parámetros de Gibbs.

Para cada componente  $\theta_j$ , HMC agrega una variable de momento  $\phi_j$ . La densidad posterior  $P(\theta | y)$  se incrementa mediante una distribución independiente  $P(\phi)$  del momento, definiendo así una distribución conjunta:

$$P(\theta, \phi | y) = P(\phi) \cdot P(\theta | y)$$

con  $\phi_j \sim \text{Normal}(0, M_{jj})$ , donde  $M$  es una matriz diagonal y corresponde a la matriz de masa.

### 2.3.1. Algoritmo HMC

El algoritmo HMC es muy similar al algoritmo Metropolis pero con la inclusión de  $\phi$  como forma de cuantificar el gradiente de la distribución posterior:

1. Muestrar  $\phi$  a partir de una distribución  $N(\mathbf{0}, \mathbf{M})$

2. Muestrear simultáneamente  $\theta$  y  $\phi$  con  $L$  pasos de salto cada uno escalado por un factor  $\epsilon$ . En un paso de salto, tanto  $\theta$  como  $\phi$  cambian, en relación con los demás. Repetir los siguientes pasos  $L$  veces:

a) Usar el gradiente del logaritmo posterior de  $\theta$  para producir un paso intermedio de  $\phi$ :

$$\phi \leftarrow \phi + \frac{1}{2}\epsilon \frac{d \log p(\theta | y)}{d\theta}$$

b) Usar  $\phi$  para actualizar  $\theta$ :

$$\theta \leftarrow \theta + \epsilon \mathbf{M}^{-1} \phi$$

c) Usar de nuevo el gradiente del logaritmo posterior de  $\theta$  para producir otro paso intermedio de  $\phi$ :

$$\phi \leftarrow \phi + \frac{1}{2}\epsilon \frac{d \log p(\theta | y)}{d\theta}$$

3. Asignar  $\theta^{t-1}$  y  $\phi^{t-1}$  como los valores del vector de parámetros y el vector de momento, respectivamente, al comienzo del proceso de salto (paso 2) y  $\theta^*$  y  $\phi^*$  como los valores después de  $L$  pasos. Como regla de aceptación / rechazo, calcular:

$$r = \frac{p(\theta^* | y)p(\phi^*)}{p(\theta^{t-1} | y)p(\phi^{t-1})}$$

4. Asignar:

$$\theta^t = \begin{cases} \theta^* & \text{con probabilidad } \min(r, 1) \\ \theta^{t-1} & \text{otro caso} \end{cases}$$

## 2.4. Diagnósticos de convergencia de las cadenas MCMC

Como ya se ha mencionado, se espera que las cadenas de Markov converjan a una distribución estacionaria, la cual corresponde a la distribución posterior de interés  $\pi(\theta | \mathbf{D})$ .

Existen algunos diagnósticos para verificar dicha convergencia entre los cuales se encuentran los diagnósticos gráficos, el diagnóstico de Gelman-Rubin y el diagnóstico de Geweke.

### 2.4.1. Diagnósticos Gráficos

Un primer enfoque empírico para el control de la convergencia es hacer gráficos de la salida de cadenas simuladas para detectar comportamientos de no estacionariedad.

Una primera gráfica es dibujar la secuencia de  $\theta^{(t)}$  contra  $t$ . Si los estados de la cadena se mueven lentamente sobre el rango de valores posibles, no existe convergencia a la distribución estacionaria [17].

En la Figura 2.1 se muestran dos casos de un diagnóstico grafico para una cadena de un parámetro  $\theta$ . Se puede observar que a diferencia del gráfico 2.1a, en el grafico 2.1b, en algunas partes los valores de la cadena cambian lentamente, lo cual indica que la cadena no converge a la distribución estacionaria.

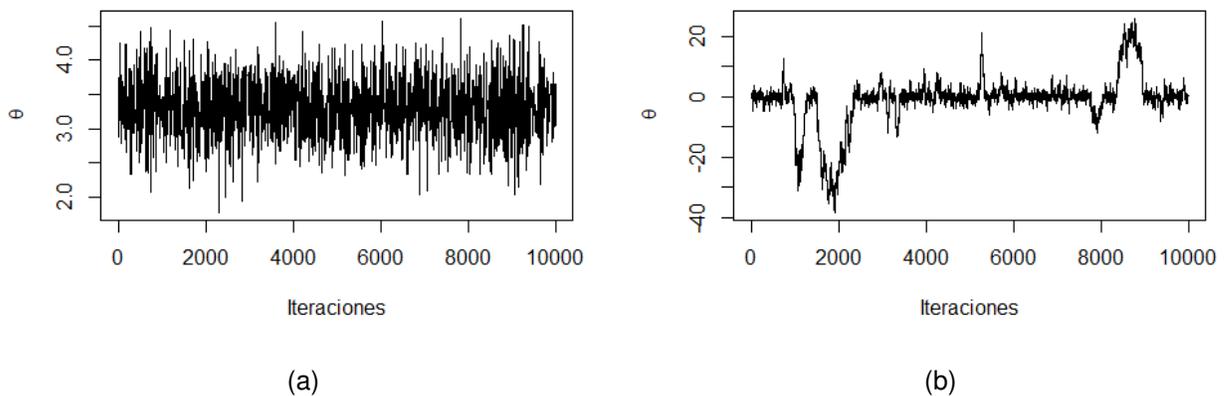


Figura 2.1: Traza de una cadena para un parámetro  $\theta$ .

Otro diagnóstico gráfico natural apropiado es verificar la estabilización de la función de densidad acumulada empírica derivado de las cadenas de Markov. Una falta de estabilidad significa que requiere más simulaciones [17].

## 2.4.2. Diagnóstico de Geweke

Geweke (1992) usa la *densidad espectral* de  $h(\theta^{(t)})$ ,

$$S_h(w) = \frac{1}{2\pi} \sum_{t=-\infty}^{t=\infty} \text{cov}\left(h(\theta^{(0)}), h(\theta^{(t)})\right) e^{tw}$$

para derivar la varianza asintótica de (8.1) como  $\gamma_h^2 = S_h^2(0)$ .

Geweke toma las primeras  $T_A$  y las últimas  $T_B$  observaciones de una secuencia de longitud  $T$  para obtener

$$\delta_A = \frac{1}{T_A} \sum_{t=1}^{T_A} h(\theta^{(t)}), \quad \delta_B = \frac{1}{T_B} \sum_{t=T-T_B+1}^T h(\theta^{(t)}),$$

y las estimaciones  $\sigma_A^2$  y  $\sigma_B^2$  de  $S_h(0)$  basadas en ambas submuestras, respectivamente.

La estadística de prueba es la denominada puntuación  $Z$ :

$$Z = \frac{\sqrt{T}(\delta_A - \delta_B)}{\sqrt{\frac{\sigma_A^2}{\tau_A} + \frac{\sigma_B^2}{\tau_B}}}$$

es asintóticamente normal con  $T_A = \tau_A T$ ,  $T_B = \tau_B T$  y  $\tau_A + \tau_B < 1$ .

De modo que valores grandes de  $Z$  indican falta de convergencia. Sin embargo, valores pequeños no significan que existe convergencia. Luego, la decisión debe ser tomada conjuntamente con otros criterios [17].

## 2.4.3. Diagnóstico de Gelman-Rubin

Usando  $M$  cadenas paralelas  $\{\theta_m^{(t)}\}_t$  ( $1 \leq m \leq M, 1 \leq t \leq T$ ), posiblemente transformadas en  $\xi_m^{(t)} = h(\theta_m^{(t)})$ , la regla de detención de Gelman y Rubin (1992) se basa en la diferencia entre un estimador ponderado de la varianza y la varianza de los estimadores de las diferentes cadenas.

La varianza entre cadenas se define como la varianza de los promedios

$$B_T = \frac{1}{M-1} \sum_{m=1}^M (\bar{\xi}_m - \bar{\xi})^2$$

mientras que la varianza dentro de la cadena es el promedio de las varianzas

$$W_T = \frac{1}{M-1} \sum_{m=1}^M s_m^2 = \frac{1}{M-1} \sum_{m=1}^M \frac{1}{T-1} \sum_{t=1}^T (\xi_m^{(t)} - \bar{\xi}_m)^2$$

con

$$\bar{\xi}_m = \frac{1}{T} \sum_{t=1}^T \xi_m^{(t)}, \quad \bar{\xi} = \frac{1}{M} \sum_{m=1}^M \bar{\xi}_m$$

Un primer estimador de la varianza posterior de  $\xi$  es

$$\hat{\sigma}_T^2 = \frac{T-1}{T} W_T + B_T$$

El criterio recomendado de Gelman y Rubin es monitorear

$$R_T^2 = \frac{\hat{\sigma}_T^2 + \frac{B_T}{M} V_T + 1}{W_T V_T + 3}$$

donde  $V_T = 2(\hat{\sigma}_T^2 + \frac{B_T}{M})^2 / W_T$  son los grados de libertad estimados. Mientras que la distribución aproximada de  $R_T^2$  se puede obtener de una aproximación F,  $T B_T / W_T \sim F(M-1, 2W_T^2 / \bar{w}_T)$ , con

$$\bar{w}_T = \frac{1}{M^2} \left[ \sum_{m=1}^M s_m^4 - \frac{1}{M} \left( \sum_{m=1}^M s_m^2 \right)^2 \right]$$

El factor  $R_T^2$  disminuye a 1 a medida que la longitud de la cadena tiende al infinito, por lo que  $R_T^2$  debe estar cerca de 1 si las cadenas han convergido aproximadamente a la distribución objetivo. El valor de  $R_T$  debería ser inferior a 1.1 o 1.2 [18]

## 2.5. Estimador de Bayes

Un enfoque de la teoría de decisiones para la estimación de un parámetro desconocido  $\theta$  introduce la función de pérdida  $L(\theta, a)$  que, en términos generales, produce el costo de decidir que el parámetro tiene el valor  $a$ , cuando de hecho es igual a  $\theta$ . La estimación  $a$  puede elegirse para minimizar la pérdida esperada posterior,

$$E(L(a|y)) = \int L(\theta, a) \pi(\theta|y) d\theta$$

Esta elección óptima de  $a$  se denomina estimador de Bayes para la función de pérdida  $L$  [19].

Las funciones de pérdida más usuales son:

- Función de pérdida cuadrática:

$$L(\theta, a) = (\theta - a)^2$$

- Función de pérdida lineal absoluta:

$$L(\theta, a) = |\theta - a|$$

- Función de pérdida 0-1:

$$L(\theta, a) = \begin{cases} 0 & \text{si } \theta = a \\ 1 & \text{si } \theta \neq a \end{cases}$$

La media, mediana y moda son estimadores de Bayes bajo las funciones de pérdida cuadrática, lineal absoluta y 0-1, respetivamente.

## 2.6. Intervalos Creíbles

Formalmente, un conjunto creíble para la distribución posterior se define de la siguiente manera. Sea  $C$  un subconjunto del espacio de parámetros,  $\Theta$ , tal que un intervalo creíble del  $100(1 - \alpha)\%$  cumpla la condición:

$$1 - \alpha = \int_C \pi(\theta|y) d\theta \quad (2.3)$$

Un defecto importante de un conjunto creíble es que no especifica si los valores de  $\theta$  dentro del conjunto son más probables que los valores fuera del conjunto. Es preferible elegir un conjunto de valores creíbles con la densidad posterior más alta. Por lo tanto, una extensión del conjunto creíble es el intervalo de densidad posterior más alta (HPDI, por sus siglas en inglés) que cubre al menos una probabilidad de  $1 - \alpha$  y contiene los valores más probables de  $\theta$ ,  $P(C|y) \geq 1 - \alpha$ , y para  $\theta_1 \in C$  y  $\theta_2 \notin C$  se cumple que  $P(\theta_1|y) \geq P(\theta_2|y)$ , es decir, la densidad posterior de cada punto dentro del intervalo HPD es mayor que la de cada punto fuera del intervalo.

Para un parámetro multidimensional, el conjunto creíble se define como una región creíble en lugar de un intervalo creíble. De la misma manera, la región HPD tiene la propiedad de que tiene el menor volumen posible en el espacio de parámetros de  $\theta$ . Para una distribución posterior unimodal simétrica, la región HPD es igual a un conjunto creíble de cola igual que se define tomando los cuantiles  $\alpha/2$  y  $1 - \alpha/2$  de la distribución posterior como la región creíble con probabilidad  $1 - \alpha$  [20].

## 2.7. Medidas de precisión predictiva

Una forma de evaluar un modelo es mediante la precisión de sus predicciones. A veces nos preocupamos por esta precisión por sí misma, como cuando evaluamos un pronóstico. En otros entornos, la precisión predictiva se valora para comparar diferentes modelos y no por sí misma.

Preferiblemente, la medida de precisión predictiva se adapta específicamente a la aplicación en cuestión y mide de la manera más correcta posible el beneficio (o costo) de predecir datos futuros con el modelo. Ejemplos de medidas específicas de la aplicación son la precisión de la clasificación y el costo monetario. Sin embargo, para muchos análisis de datos, no se dispone de información explícita sobre costos o beneficios, y el rendimiento predictivo de un modelo se evalúa mediante funciones y reglas de puntuación genéricas [16].

En la *predicción puntual* (estimación puntual predictiva o pronóstico puntual) un solo valor es reportado como una predicción de la observación futura desconocida. Medidas de precisión predictiva para la predicción puntual se denominan funciones de puntuación. Consideramos el error cuadrático como un ejemplo, ya que es la función de puntuación más común en la literatura sobre predicción [16].

*Error cuadrático medio.* El ajuste de un modelo a nuevos datos se puede resumir en predicción puntual por el error cuadrático medio,  $\frac{1}{n} \sum_{i=1}^n (y_i - E(y_i | \theta))^2$  o una versión ponderada como  $\frac{1}{n} \sum_{i=1}^n (y_i - E(y_i | \theta))^2 / \text{var}(y_i | \theta)$ . Estas medidas tienen la ventaja de ser fáciles de calcular y, más importante, de interpretar, pero la desventaja de ser menos apropiadas para modelos que están muy lejos de la distribución normal [16].

*Log densidad predictiva o log-verosimilitud.* La puntuación logarítmica de las predic-

ciones es el logaritmo de densidad predictiva  $\log p(y | \theta)$ , que es proporcional al error cuadrático medio si el modelo es normal con varianza constante. La log densidad predictiva también se denomina a veces log-verosimilitud. Debido a su generalidad, se utiliza la log densidad predictiva para medir la precisión predictiva [16].

### *Evaluación de la precisión predictiva para un modelo ajustado*

En la práctica el parámetro  $\theta$  no se conoce, así que no podemos conocer la log densidad predictiva  $\log p(y | \theta)$ . Se utiliza la distribución a posteriori,  $p_{\text{post}}(\theta) = p(\theta | y)$ , para resumir la precisión predictiva del modelo ajustado a los datos por [16]:

$$\begin{aligned} \text{lppd} &= \text{log densidad predictiva puntual} \\ &= \log \prod_{i=1}^n p_{\text{post}}(y_i) = \sum_{i=1}^n \log \int p(y_i | \theta) p_{\text{post}}(\theta) d\theta \end{aligned} \quad (2.4)$$

## **2.8. Criterios de Selección de Modelos**

### **2.8.1. Criterio de Información de Watanabe-Akaike (WAIC)**

WAIC es un enfoque completamente Bayesiano para estimar la densidad predictiva puntual logarítmica esperada y se define como [16]:

$$WAIC = -2(\text{lppd} - p_{WAIC})$$

donde  $\text{lppd}$  es la log-densidad predictiva puntual y  $p_{WAIC}$  es el número efectivo de parámetros.

En la práctica, la log-densidad predictiva puntual  $\text{lppd}$ , se calcula usando muestras de  $p_{\text{post}}(\theta)$ , las simulaciones posteriores habituales, que etiquetamos  $\theta^s$ ,  $s = 1, \dots, S$ :

$$\begin{aligned} \text{lppd calculada} &= \text{log densidad predictiva puntual calculada} \\ &= \sum_{i=1}^n \log \left( \frac{1}{S} \sum_{s=1}^S p(y_i | \theta^s) \right) \end{aligned} \quad (2.5)$$

El cálculo del número efectivo de parámetros ( $p_{WAIC}$ ) se hace mediante:

$$p_{WAIC} = \sum_{i=1}^n V_{s=1}^S (\log p(y_i | \theta^s)) \quad (2.6)$$

## 2.9. Método numérico de Runge-Kutta de cuarto orden

Es el miembro más conocido de la familia Runge-Kutta generalmente se conoce como “RK4”, el “método clásico de Runge-Kutta”, o simplemente como “el método Runge-Kutta”.

Especificando un problema de valor inicial de la siguiente manera:

$$\frac{dy}{dt} = f(t, y), \quad y(t_0) = y_0$$

donde  $y$  es una función desconocida (escalar o vectorial) del tiempo  $t$ , que se quiere aproximar;  $\frac{dy}{dt}$ , la velocidad a la que  $y$  cambia, es una función de  $t$  y  $y$ . En el momento inicial  $t_0$  el correspondiente valor de  $y$  es  $y_0$ . La función  $f$  y las condiciones iniciales  $t_0, y_0$  están dadas.

Eligiendo un tamaño de paso  $h > 0$  y definiendo [21]:

$$\begin{aligned} y_{n+1} &= y_n + \frac{1}{6}h(k_1 + 2k_2 + 2k_3 + k_4), \\ t_{n+1} &= t_n + h. \end{aligned} \tag{2.7}$$

con

$$\begin{aligned} k_1 &= f(t_n, y_n), \\ k_2 &= f\left(t_n + \frac{h}{2}, y_n + h\frac{k_1}{2}\right), \\ k_3 &= f\left(t_n + \frac{h}{2}, y_n + h\frac{k_2}{2}\right), \\ k_4 &= f(t_n + h, y_n + hk_3). \end{aligned} \tag{2.8}$$

Aquí  $y_{n+1}$  corresponde a la aproximación RK4 de  $y(t_{n+1})$ , y el valor siguiente ( $y_{n+1}$ ) está determinado por el valor presente ( $y_n$ ) más el promedio ponderado de cuatro incrementos, donde cada incremento es el producto del tamaño del intervalo,  $h$ , y una pendiente estimada especificada por la función  $f$  en el lado derecho de la ecuación diferencial.

- $k_1$  es la pendiente al comienzo del intervalo, usando  $y$  ( Método de Euler );
- $k_2$  es la pendiente en el punto medio del intervalo, utilizando  $y$  y  $k_1$ ;
- $k_3$  es nuevamente la pendiente en el punto medio, pero ahora usando  $y$  y  $k_2$ ;
- $k_4$  es la pendiente al final del intervalo, utilizando  $y$  y  $k_3$ .

### Metodología

---

#### 3.1. Datos

Para este trabajo se cuenta con tres conjuntos de datos tomados de Ikeda y colaboradores [22], estos datos fueron obtenidos de un experimento realizado en laboratorio, en el que se sembraron células MT4C5 en dos mililitros de un medio de cultivo. El primer conjunto de datos corresponde al recuento de células MT4C5 durante los primeros 10 días. Estos datos se muestran en el Cuadro 3.1.

El segundo conjunto de datos corresponde a la cantidad de proteína viral medida durante los primeros 6 días en un segundo experimento en el que las células MT4C5 se inocularon con 10 ng de proteína viral de VIH-1.

El tercer conjunto de datos corresponde a la cantidad de proteína viral medida durante los primeros 6 días en un tercer experimento en el que las células MT4C5 se inocularon con 10 ng de proteína viral de VIH-1 en presencia de interferón (IFN).

El segundo y tercer conjunto de datos se muestran en el Cuadro 3.2.

#### 3.2. Método

La metodología que se usa en este trabajo básicamente es la estimación bayesiana de los parámetros de un modelo estadístico considerado para cada conjunto de datos y

Cuadro 3.1: Recuento de Células en cultivo sin infección por VIH-1.

Tiempo (días)	Células Vivas
0	559827
1	847413
2	975689
3	1155689
4	1244655
5	1312931
6	1379137
7	1447413
8	1455689
9	1435000
10	1432931

comprende tres etapas:

- Etapa 1: Proponer distribuciones no informativas y estimar los parámetros de la dinámica celular para el primer conjunto de datos.
- Etapa 2: Para el segundo conjunto de datos, estimar los parámetros de la dinámica celular usando las distribuciones a posteriores obtenidas de la Etapa 1 como distribuciones a priori en esta etapa. Proponer distribuciones a priori no informativas para la estimación de los parámetros de la dinámica viral.
- Etapa 3: Para el tercer conjunto de datos, estimar los parámetros de la dinámica celular y dinámica viral, usando las distribuciones a posteriores obtenidas en la Etapa 2 como distribuciones a priori en esta etapa. Proponer distribuciones a priori no informativas para la estimación de los parámetros de la eficacia del interferón.

Cuadro 3.2: Carga viral en cultivos con infección por VIH-1.

Tiempo (días)	Carga Viral (ng)	
	Sin IFN	Con IFN
0	4.47	3.05
1	490.20	73.25
2	1785.07	176.77
3	1486.64	420.62
4	478.56	1302.52
5	58.71	1503.24
6	6.71	549.82

### 3.3. Modelos Matemáticos

Para describir el comportamiento poblacional de las células en el primer conjunto de datos se usarán los siguientes modelos matemáticos:

#### Modelo de von Bertalanffy

$$S'(t) = r(k - S(t)) \quad (3.1)$$

#### Modelo Logístico

$$S'(t) = rS(t) \left[ 1 - \frac{S(t)}{k} \right] \quad (3.2)$$

#### Modelo de Gompertz

$$S'(t) = rS(t) \ln \left[ \frac{k}{S(t)} \right] \quad (3.3)$$

donde  $S(t)$  es el número de células en el tiempo  $t$ , y los parámetros  $r$  y  $k$  representan la tasa de crecimiento de las células y la capacidad de carga (número máximo de células), respectivamente.

Para describir el segundo conjunto de datos se usarán los siguientes sistemas de ecuaciones diferenciales:

### Modelo de von Bertalanffy

$$\begin{aligned}S'(t) &= r(k - S(t)) - \beta S(t)V(t), \\I'(t) &= \beta S(t)V(t) - \delta I(t), \\V'(t) &= \rho I(t) - \gamma V(t).\end{aligned}\tag{3.4}$$

### Modelo Logístico

$$\begin{aligned}S'(t) &= rS(t) \left[ 1 - \frac{S(t)+I(t)}{k} \right] - \beta S(t)V(t), \\I'(t) &= \beta S(t)V(t) - \delta I(t), \\V'(t) &= \rho I(t) - \gamma V(t).\end{aligned}\tag{3.5}$$

### Modelo de Gompertz

$$\begin{aligned}S'(t) &= rS(t) \ln \left[ \frac{k}{S(t)+I(t)} \right] - \beta S(t)V(t), \\I'(t) &= \beta S(t)V(t) - \delta I(t), \\V'(t) &= \rho I(t) - \gamma V(t).\end{aligned}\tag{3.6}$$

donde  $S(t)$ ,  $I(t)$  y  $V(t)$  representan el número de células sanas, células infectadas y carga viral en el tiempo  $t$ , respectivamente. Los parámetros  $\beta$ ,  $\delta$ ,  $\rho$  y  $\gamma$  representan la tasa de infección, la tasa de mortalidad de las células infectadas, la tasa de producción de virus y la tasa de eliminación de viriones, respectivamente.

Para describir el tercer conjunto de datos, a los sistema de ecuaciones diferenciales anteriores se agrega el efecto del interferón:

### Modelo de von Bertalanffy

$$\begin{aligned}S'(t) &= r(k - S(t)) - (1 - \eta)\beta S(t)V(t), \\I'(t) &= (1 - \eta)\beta S(t)V(t) - \delta I(t), \\V'(t) &= (1 - \epsilon)\rho I(t) - \gamma V(t).\end{aligned}\tag{3.7}$$

### Modelo Logístico

$$\begin{aligned}S'(t) &= rS(t) \left[ 1 - \frac{S(t)+I(t)}{k} \right] - (1 - \eta)\beta S(t)V(t), \\I'(t) &= (1 - \eta)\beta S(t)V(t) - \delta I(t), \\V'(t) &= (1 - \epsilon)\rho I(t) - \gamma V(t).\end{aligned}\tag{3.8}$$

## Modelo de Gompertz

$$\begin{aligned}S'(t) &= rS(t) \ln \left[ \frac{k}{S(t)+I(t)} \right] - (1 - \eta)\beta S(t)V(t), \\I'(t) &= (1 - \eta)\beta S(t)V(t) - \delta I(t), \\V'(t) &= (1 - \epsilon)\rho I(t) - \gamma V(t).\end{aligned}\tag{3.9}$$

donde los parámetros  $\eta$  y  $\epsilon$  representan el efecto del interferón sobre la producción de células infectadas y de viriones, respectivamente.

## 3.4. Estimación Bayesiana

### 3.4.1. Modelo Estadístico.

Para la estimación de los parámetros de los modelos propuestos para cada conjunto de datos en este estudio, se considera el siguiente modelo estadístico:

$$y_i = X_\theta(t_i) + \varepsilon(t_i), \quad i = 1, \dots, n \tag{3.10}$$

- $y_i$  representa el número de células en el tiempo  $i$  para el primer conjunto de datos o la carga viral medida en el tiempo  $i$  para el segundo y tercer conjunto de datos.
- $\theta$  es el vector de parámetros:  $\theta=(r, k)$  en los Modelos (3.1), (3.2) y (3.3),  $\theta=(r, k, \beta, \delta, \rho)$  para los Modelos (3.4), (3.5) y (3.6) y  $\theta=(r, k, \beta, \delta, \rho, \eta, \epsilon)$  para los Modelos (3.7), (3.8) y (3.9).
- $X_\theta(t_i)$  es la solución numérica de las ecuaciones (3.1), (3.2), (3.3) y la solución numérica de la ecuación correspondiente a la variable de carga viral de los sistemas de ecuaciones (3.4), (3.5), (3.6), (3.7), (3.8) y (3.9).
- $\varepsilon(t_i)$  es el error aleatorio en el tiempo  $i$ .

### 3.4.2. Función de Verosimilitud

Se asume que el error aleatorio de la ecuación (3.10),  $\varepsilon(t_i)$ , tiene distribución normal, con media 0 y varianza  $\sigma^2$ . Por tanto  $y_i \sim N(X_\theta(t_i), \sigma^2)$ .

En consecuencia, la función de verosimilitud de los datos bajo el Modelo (3.10) está dada por:

$$L(\boldsymbol{\theta}|y_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(y_i - X_{\boldsymbol{\theta}}(t_i))^2\right]$$

### 3.4.3. Información a priori

En la primera etapa se proponen distribuciones uniformes para los parámetros de dinámica celular  $r$  y  $k$  cuyo soporte se eligió de acuerdo a la literatura consultada y para darle sentido biológico [22]. Para el parámetro  $\sigma^2$  se asignó una distribución Gamma Inversa no informativa como a priori [16] (Cuadro 3.3)

Cuadro 3.3: Distribuciones a priori Etapa 1.

Parámetro	Distribución
$r$	<i>Uniforme</i> (0, 1)
$k$	<i>Uniforme</i> ( $1 \times 10^6$ , $2 \times 10^6$ )
$\sigma^2$	<i>Gamma – Inversa</i> ( $1 \times 10^{-3}$ , $1 \times 10^{-3}$ )

En la segunda etapa, para los parámetros de dinámica celular  $r$  y  $k$  se usan como a priori distribuciones gamma, cuyos hiperparámetros se obtienen al ajustar una distribución gamma a las muestras obtenidas mediante MCMC en la primera etapa para el parámetro correspondiente del modelo respectivo. Para los parámetros de dinámica viral  $\beta$ ,  $\delta$  y  $\rho$  se proponen distribuciones a priori uniformes. Se hizo un proceso análogo para las distribuciones a priori de la tercera etapa.

### 3.4.4. Distribuciones a posteriori

Para no agregar información se supone independencia entre los parámetros, luego la distribución conjunta se puede expresar como producto de las distribuciones marginales de los parámetros, es decir,  $\pi(\boldsymbol{\theta}, \sigma^2) = \pi(r)\pi(k)\pi(\sigma^2)$ . Por tanto, la distribución a posteriori conjunta es:

$$\begin{aligned}\pi(\boldsymbol{\theta}, \sigma^2 | \mathbf{y}, X_\theta) &\propto \pi(\boldsymbol{\theta}, \sigma^2) p(\mathbf{y} | \boldsymbol{\theta}, \sigma^2) \\ &\propto \pi(r) \pi(k) \pi(\sigma^2) p(\mathbf{y} | X_\theta, \boldsymbol{\theta}, \sigma^2)\end{aligned}$$

Debido a que no es posible encontrar expresiones analíticas para las distribuciones posteriores marginales de los parámetros, se obtienen aproximaciones mediante muestreo MCMC. Se usa como estimador Bayesiano la media de la distribución marginal posterior de cada parámetro y los intervalos de alta densidad posterior (HPDI) como estimación por intervalo.

La estimación de los parámetros de los modelos de la primera etapa se obtuvo mediante el uso de la librería *brms* [23] del software estadístico R [24]. Para la segunda y tercera etapa se utiliza la librería *turing* [25] del software Julia [26]. Para cada modelo se generaron tres cadenas, cada una con 10000 iteraciones.

Con el propósito de evaluar la convergencia de las cadenas MCMC, se realizan los diagnósticos de convergencia de Gelman-Rubin y Geweke, que se encuentran implementados en la librería *coda* [27] del software R. Además, se realizan diagnósticos gráficos para evaluar la convergencia de las cadenas de Markov para cada uno de los parámetros.

### Resultados

---

En este capítulo se muestran las estimaciones bayesianas puntuales y por intervalo obtenidos para los parámetros de los modelos ajustados en cada una de las etapas, usando como estimador Bayesiano la media posterior; además los diagnósticos de convergencia y se hace la selección del modelo que se ajusta mejor a los datos de dinámica celular y viral usados.

#### 4.1. Etapa 1: Datos de crecimiento Celular

En esta etapa se usó el primer conjunto de datos correspondiente a los datos de crecimiento celular del primer cultivo para ajustar los Modelos (3.1), (3.2) y (3.3) (modelos con función de crecimiento von Bertalanffy, Logística y Gompertz respectivamente).

##### 4.1.1. Modelo von Bertalanffy

###### Distribuciones a priori

Las distribuciones a priori usadas se muestran en el Cuadro 3.3.

## Estimación

En el Cuadro 4.1 se muestran las estimaciones puntuales e intervalos de alta densidad posterior para cada parámetro en donde se puede observar que: La estimación bayesiana puntual para la tasa de crecimiento de las células,  $r$ , es de 0.33 y que con una probabilidad de 0.95 esta tasa se encuentra entre 0.29 y 0.37; la estimación bayesiana puntual para la capacidad de carga,  $k$ , es de 1500654.1, esto quiere decir que el número máximo de células que puede haber en el medio es 1500654 y que con una probabilidad de 0.95 este número se encuentra entre 1464524 y 1541804.

Cuadro 4.1: Estimaciones Modelo 3.1: von Bertalanffy.

Parámetro	Media	L. Inferior HPDI	L. Superior HPDI
$r$	0.331	0.291	0.373
$k$	1500654	1464523	1541804
$\sigma$	29280	17595	49723

## Diagnósticos de convergencia

Los resultados del diagnóstico de Gelman-Rubin mostraron un valor menor que 1.1 del factor de reducción de escala potencial para cada parámetro y un límite superior de su intervalo creíble por debajo de 1.2 en todos los casos, esto indica que las cadenas han convergido a la distribución estacionaria (Cuadro 4.2).

Cuadro 4.2: Diagnóstico de Gelman-Rubin Modelo 3.1: von Bertalanffy.

Parámetro	Estimación puntual	L. Superior
$r$	1.00	1.00
$k$	1.00	1.00
$\sigma$	1.00	1.00

Los resultados del diagnóstico de Geweke muestran que la mayoría de los valores de la estadística de prueba  $Z$  para cada parámetro y cada cadena se encuentran entre -2 y

2, lo cual indica que las medias de las primeras y las últimas iteraciones de cada cadena son estadísticamente iguales, y por tanto existe convergencia de las mismas (Cuadro 4.3).

Cuadro 4.3: Diagnóstico de Geweke Modelo 3.1: von Bertalanffy.

Cadena	Parámetros		
	$r$	$k$	$\sigma$
1	0.1929	1.1916	0.3479
2	0.6067	-0.7051	0.6424
3	2.1329	-1.9726	0.8622

Los diagnósticos de convergencia gráficos como las trazas de las cadenas y las densidades posteriores de los parámetros son congruentes con los resultados anteriores. Se puede observar un buen mezclado de las cadenas y además que recorren todo el espacio paramétrico de su distribución posterior indicando convergencia de su distribución posterior (Figura 4.1).

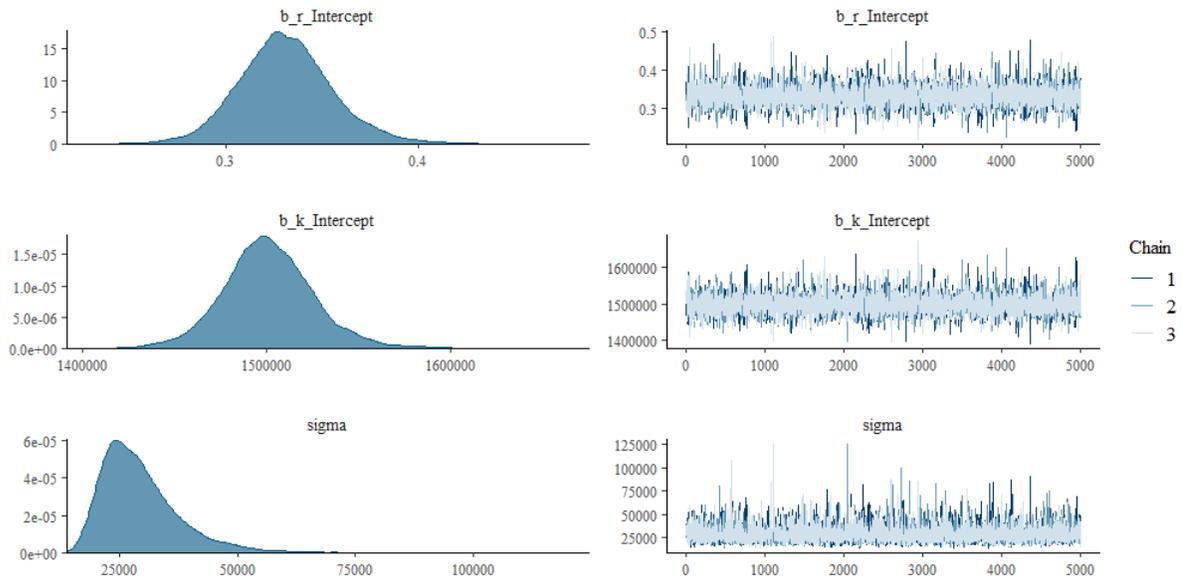


Figura 4.1: Trazas de las cadenas y densidades posteriores de los parámetros Modelo 3.1: von Bertalanffy.

## Ajuste del modelo

Se usaron las estimaciones puntuales y los intervalos creíbles correspondientes para calcular la solución numérica del Modelo (3.1). En la Figura 4.2 se puede observar un buen ajuste del modelo ya que la curva ajustada pasa muy cerca de los datos de crecimiento celular, además todas las observaciones caen dentro de la banda de credibilidad del 95%.

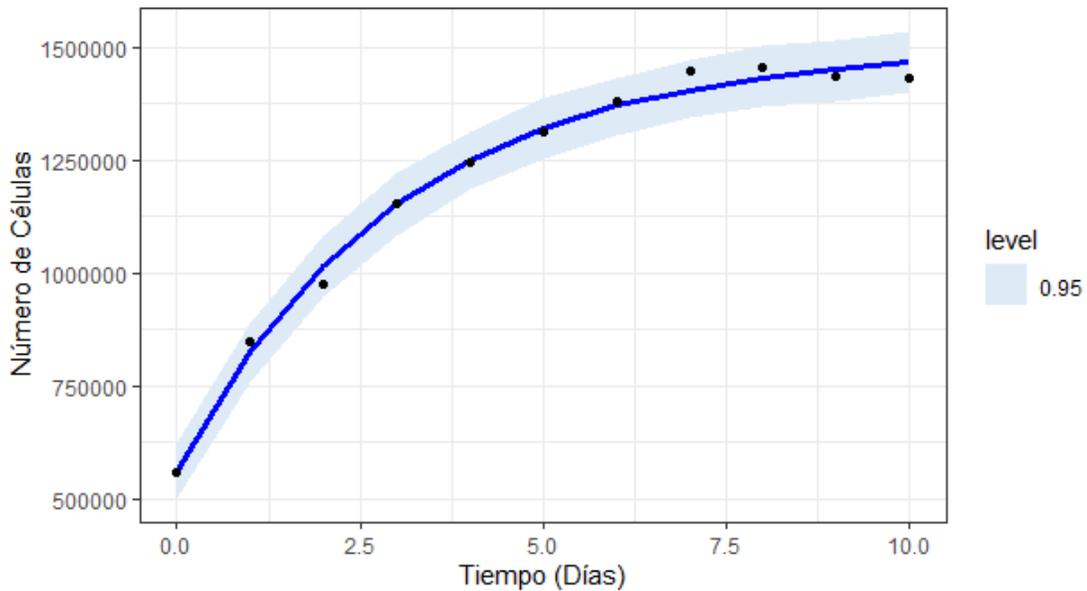


Figura 4.2: Modelo 3.1: von Bertalanffy ajustado.

### 4.1.2. Modelo Logístico

#### Distribuciones a priori

Las distribuciones a priori usadas se muestran en el Cuadro 3.3.

#### Estimación

En el Cuadro 4.4 se muestran las estimaciones puntuales e intervalos de alta densidad posterior para cada parámetro en donde se puede observar que: La estimación bayesiana puntual para la tasa de crecimiento de las células,  $r$ , es de 0.61 y que con una probabilidad de 0.95 esta tasa se encuentra entre 0.54 y 0.69; la estimación bayesiana puntual para la capacidad de carga,  $k$ , es de 1447203.9, esto quiere decir que el número máximo de

células que puede haber en el medio es 1447204 y que con una probabilidad de 0.95 este número se encuentra entre 1407225 y 1489354.

Cuadro 4.4: Estimaciones Modelo 3.2: Logístico.

Parámetro	Media	L. Inferior HPDI	L. Superior HPDI
$r$	0.610	0.538	0.694
$k$	1447204	1407224	1489354
$\sigma$	36231	22182	62419

### Diagnósticos de convergencia

Los resultados del diagnóstico de Gelman-Rubin mostraron un valor menor que 1.1 del factor de reducción de escala potencial para cada parámetro y un límite superior de su intervalo creíble por debajo de 1.2 en todos los casos, esto indica que las cadenas han convergido a la distribución estacionaria (Cuadro 4.5).

Cuadro 4.5: Diagnóstico de Gelman-Rubin Modelo 3.2: Logístico.

Parámetro	Estimación puntual	L. Superior
$r$	1.00	1.00
$k$	1.00	1.00
$\sigma$	1.00	1.00

Los resultados del diagnóstico de Geweke muestran que la mayoría de los valores de la estadística de prueba  $Z$  para cada parámetro y cada cadena se encuentran entre -2 y 2, lo cual indica que las medias de las primeras y las últimas iteraciones de cada cadena son estadísticamente iguales, y por tanto existe convergencia de las mismas (Cuadro 4.6).

Cuadro 4.6: Diagnóstico de Geweke Modelo 3.2: Logístico.

Cadena	Parámetros		
	$r$	$k$	$\sigma$
1	0.6693	-0.5480	-0.7836
2	2.0772	-1.8485	0.6481
3	-1.0995	1.2113	1.4363

Los diagnósticos de convergencia gráficos como las trazas de las cadenas y las densidades posteriores de los parámetros son congruentes con los resultados anteriores. Se puede observar un buen mezclado de las cadenas y además que recorren todo el espacio paramétrico de su distribución posterior indicando convergencia (Figura 4.3).

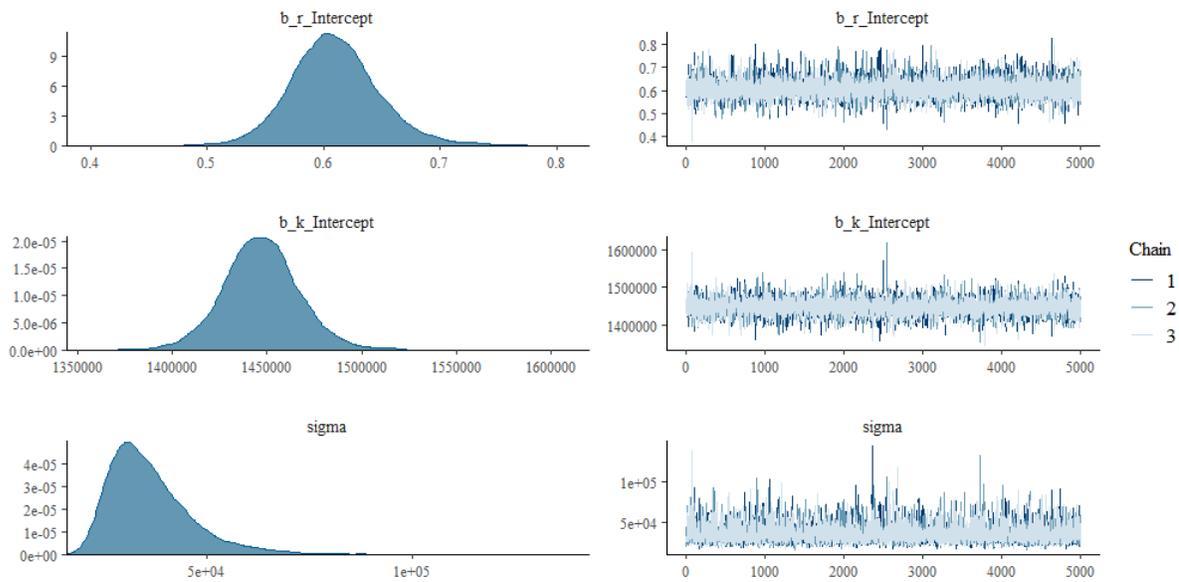


Figura 4.3: Trazas de las cadenas y densidades posteriores de los parámetros Modelo 3.2: Logístico.

### Ajuste del modelo

Se usaron las estimaciones puntuales y los intervalos creíbles correspondientes para calcular la solución numérica del Modelo (3.2). En la Figura 4.4 se puede observar un buen ajuste del modelo ya que la curva ajustada pasa muy cerca de los datos de crecimiento celular, además todas las observaciones caen dentro de la banda de credibilidad

del 95 %.

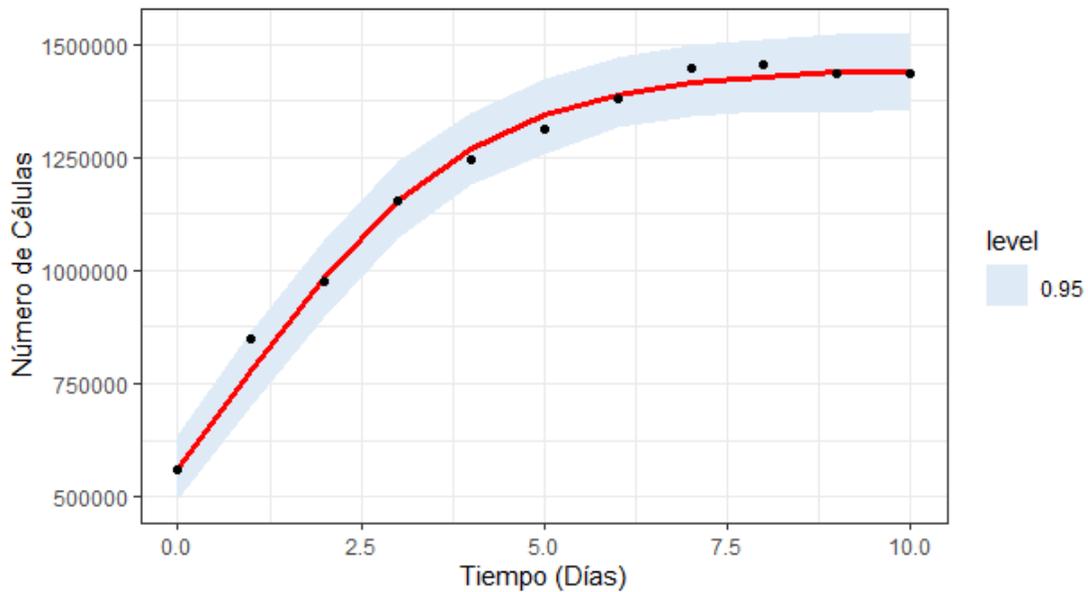


Figura 4.4: Modelo 3.2: Logístico ajustado.

### 4.1.3. Modelo Gompertz

#### Distribuciones a priori

Las distribuciones a priori usadas se muestran en el Cuadro 3.3.

#### Estimación

En el Cuadro 4.7 se muestran las estimaciones puntuales e intervalos de alta densidad posterior para cada parámetro en donde se puede observar que: La estimación bayesiana puntual para la tasa de crecimiento de las células,  $r$ , es de 0.46 y que con una probabilidad de 0.95 esta tasa se encuentra entre 0.41 y 0.52; la estimación bayesiana puntual para la capacidad de carga,  $k$ , es de 1467772.7, esto quiere decir que el número máximo de células que puede haber en el medio es 1467773 y que con una probabilidad de 0.95 este número se encuentra entre 1428181 y 1509613.

#### Diagnósticos de convergencia

Los resultados del diagnóstico de Gelman-Rubin mostraron un valor menor que 1.1 del factor de reducción de escala potencial para cada parámetro y un límite superior de

Cuadro 4.7: Estimaciones Modelo 3.3: Gompertz.

Parámetro	Media	L. Inferior HPDI	L. Superior HPDI
$r$	0.462	0.407	0.524
$k$	1467773	1428181	1509613
$\sigma$	29888	18187	51249

su intervalo creíble por debajo de 1.2 en todos los casos, esto indica que las cadenas han convergido a la distribución estacionaria (Cuadro 4.8).

Cuadro 4.8: Diagnóstico de Gelman-Rubin Modelo 3.3: Gompertz.

Parámetro	Estimación puntual	L. Superior
$r$	1.00	1.00
$k$	1.00	1.00
$\sigma$	1.00	1.00

Los resultados del diagnóstico de Geweke muestran que la mayoría de los valores de la estadística de prueba  $Z$  para cada parámetro y cada cadena se encuentran entre -2 y 2, lo cual indica que las medias de las primeras y las últimas iteraciones de cada cadena son estadísticamente iguales, y por tanto existe convergencia de las mismas (Cuadro 4.9).

Cuadro 4.9: Diagnóstico de Geweke Modelo 3.3: Gompertz.

Cadena	Parámetros		
	$r$	$k$	$\sigma$
1	0.5459	-0.5844	-1.0853
2	0.9671	-0.8759	-1.2048
3	1.3452	-1.2452	-1.3082

Los diagnósticos de convergencia gráficos como las trazas de las cadenas y las densidades posteriores de los parámetros son congruentes con los resultados anteriores. Se

puede observar un buen mezclado de las cadenas y además que recorren todo el espacio paramétrico de su distribución posterior indicando convergencia (Figura 4.5).

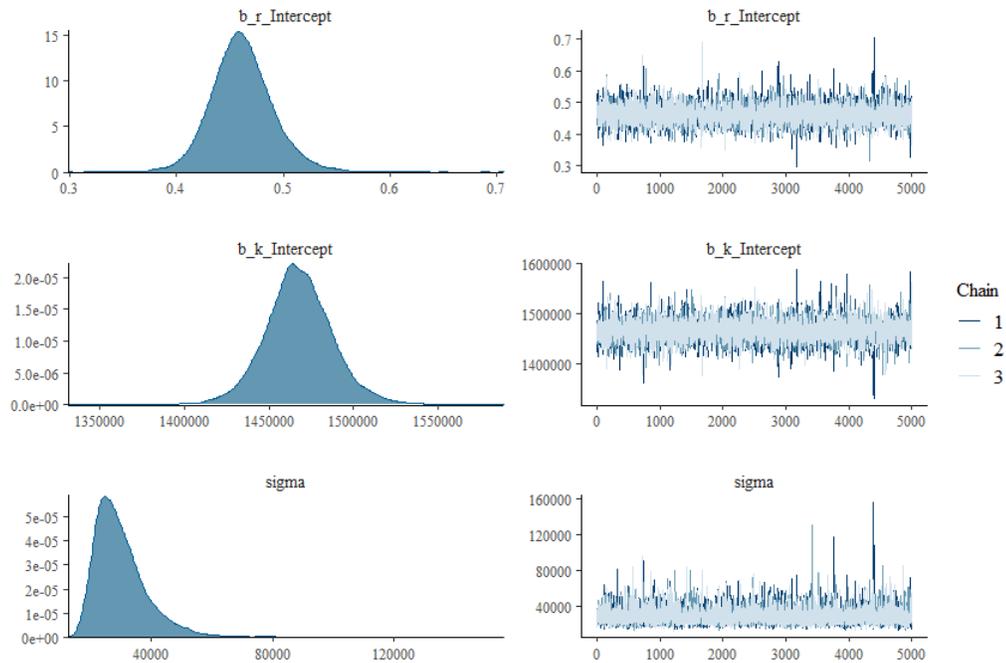


Figura 4.5: Trazas de las cadenas y densidades posteriores de los parámetros Modelo 3.3: Gompertz.

### Ajuste del modelo

Se usaron las estimaciones puntuales y los intervalos creíbles correspondientes para calcular la solución numérica del Modelo (3.3). En la Figura 4.6 se puede observar un buen ajuste del modelo ya que la curva ajustada pasa muy cerca de los datos de crecimiento celular, además todas las observaciones caen dentro de la banda de credibilidad del 95 %.

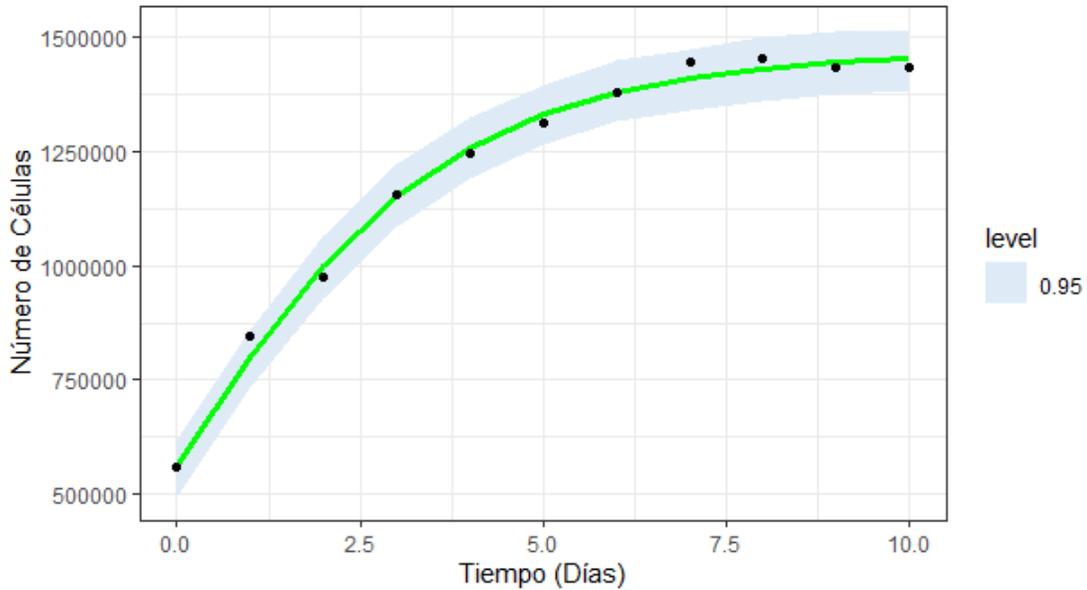


Figura 4.6: Modelo 3.3: Gompertz ajustado.

#### 4.1.4. Comparación de modelos

Con el propósito de seleccionar el modelo que se ajuste mejor a los datos de crecimiento celular, en primer lugar se hace una comparación gráfica de los modelos ajustados. En la Figura 4.7 se puede observar que las curvas de los ajustes son muy similares, y por tanto es necesario utilizar otros criterios de comparación como el criterio de Información de Watanabe-Akaike (WAIC). En el Cuadro 4.10 se muestran los resultados de esta comparación donde se puede observar que el modelo que tiene menor WAIC es el Modelo Gompertz y el Modelo von Bertalanffy. Esto significa que el ajuste es muy similar, sin embargo elegimos el Modelo Gompertz.

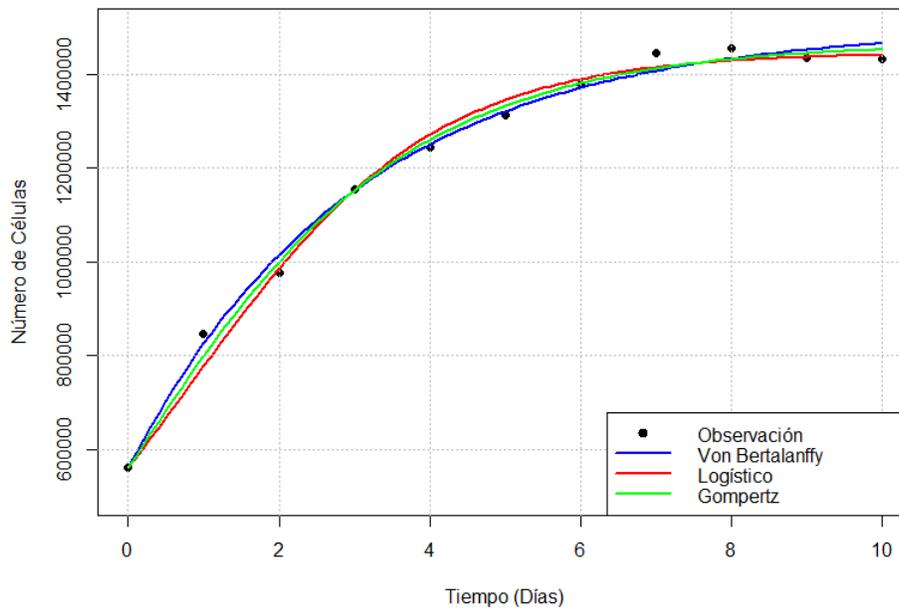


Figura 4.7: Comparación de Modelos Etapa 1.

Cuadro 4.10: Criterios de Comparación de Modelos.

Modelo	WAIC
von Bertalanffy	259.2
Logístico	264.0
Gompertz	258.1

## 4.2. Etapa 2: Datos de Carga Viral en ausencia de interferón

En esta etapa se usó el segundo conjunto de datos correspondiente a los datos de carga viral del segundo cultivo para ajustar los Modelos (3.4), (3.5) y (3.6) (modelos con función de crecimiento von Bertalanffy, Logística y Gompertz).

Los hiperparámetros de la distribución Gamma para los parámetros de dinámica celular  $r$  y  $k$  se obtuvieron al ajustar una distribución gamma a las muestras obtenidas

mediante MCMC en la primera etapa para el parámetro correspondiente. Se asignaron distribuciones uniformes para los parámetros de dinámica viral  $\beta$ ,  $\delta$  y  $\rho$  cuyo soporte se eligió de acuerdo a la literatura consultada y para darle sentido biológico [22]. El parámetro  $\gamma$  se fijó en 2.3. Para el parámetro  $\sigma^2$  se asignó una distribución Gamma Inversa no informativa.

## 4.2.1. Modelo von Bertalanffy

### Distribuciones a priori

Las distribuciones a priori usadas se muestran en el Cuadro 4.11.

Cuadro 4.11: Distribuciones a priori usadas en el Modelo 3.4: von Bertalanffy.

Parámetro	Distribución
$r$	$Gamma(158, 473)$
$k$	$Gamma(3 \times 10^3, 2 \times 10^{-3})$
$\beta$	$Uniforme(0, 1 \times 10^{-2})$
$\delta$	$Uniforme(0, 2)$
$\rho$	$Uniforme(0, 1 \times 10^{-2})$
$\sigma^2$	$Gamma - Inversa(1 \times 10^{-3}, 1 \times 10^{-3})$

### Estimación

En el Cuadro 4.12 se muestran las estimaciones puntuales e intervalos de alta densidad posterior para cada parámetro en donde se puede observar que: La estimación bayesiana puntual para la tasa de crecimiento de las células,  $r$ , es de 0.0254 y que con una probabilidad de 0.95 esta tasa se encuentra entre 0.0004 y 0.0317; la estimación bayesiana puntual para la capacidad de carga,  $k$ , es de 1461026.01, esto quiere decir que el número máximo de células que puede haber en el medio es 14610267 y que con una probabilidad de 0.95 este número se encuentra entre 1018568 y 1710053; la tasa de infección de células sanas,  $\beta$ , es de 0.0013 y con una probabilidad de 0.95 se encuentra entre 0.0011 y 0.0014; la tasa de mortalidad de células infectadas,  $\delta$ , se estima en 1.49 y con una probabilidad de 0.95 se encuentra entre 1.01 y 1.61; la tasa de producción de

virus,  $\rho$ , se estima en 0.0088 y con una probabilidad de 0.95 se encuentra entre 0.0058 y 0.0096.

Cuadro 4.12: Estimaciones Modelo 3.4: von Bertalanffy.

Parámetro	Media	L. Inferior HPDI	L. Superior HPDI
$r$	0.0254	0.0004	0.0317
$k$	1461026	1018568	1710052
$\beta$	0.0013	0.0011	0.0014
$\delta$	1.4890	1.0089	1.6089
$\rho$	0.0088	0.0058	0.0096
$\sigma$	102	41	138

### Diagnósticos de convergencia

Los resultados del diagnóstico de Gelman-Rubin mostraron un valor menor que 1.1 del factor de reducción de escala potencial para cada parámetro y un límite superior de su intervalo creíble por debajo de 1.2 en todos los casos, esto indica que las cadenas han convergido a la distribución estacionaria (Cuadro 4.13).

Cuadro 4.13: Diagnóstico de Gelman-Rubin Modelo 3.4: von Bertalanffy.

Parámetro	Estimación puntual	L. Superior
$r$	1.00	1.00
$k$	1.00	1.00
$\beta$	1.00	1.00
$\delta$	1.00	1.00
$\rho$	1.00	1.00
$\sigma$	1.00	1.00

Los resultados del diagnóstico de Geweke muestran que la totalidad de los valores de la estadística de prueba  $Z$  para cada parámetro y cada cadena se encuentran entre -2 y 2, lo cual indica que las medias de las primeras y las últimas iteraciones de cada cadena

son estadísticamente iguales, y por tanto existe convergencia de las mismas (Cuadro 4.14).

Cuadro 4.14: Diagnóstico de Geweke Modelo 3.4: von Bertalanffy.

Cadena	Parámetros					
	$r$	$k$	$\beta$	$\delta$	$\rho$	$\sigma$
1	-1.2206	0.2777	-0.1612	-0.4283	-0.2311	-1.0046
2	-0.9109	0.5290	-1.8911	1.4887	1.4388	-1.4825
3	-0.7071	-0.0429	-0.8230	-0.5825	-0.0598	-0.6869

Los diagnósticos de convergencia gráficos como las trazas de las cadenas y las densidades posteriores de los parámetros son congruentes con los resultados anteriores. Se puede observar un buen mezclado de las cadenas y además que recorren todo el espacio paramétrico de su distribución posterior indicando convergencia (Figura 4.8).

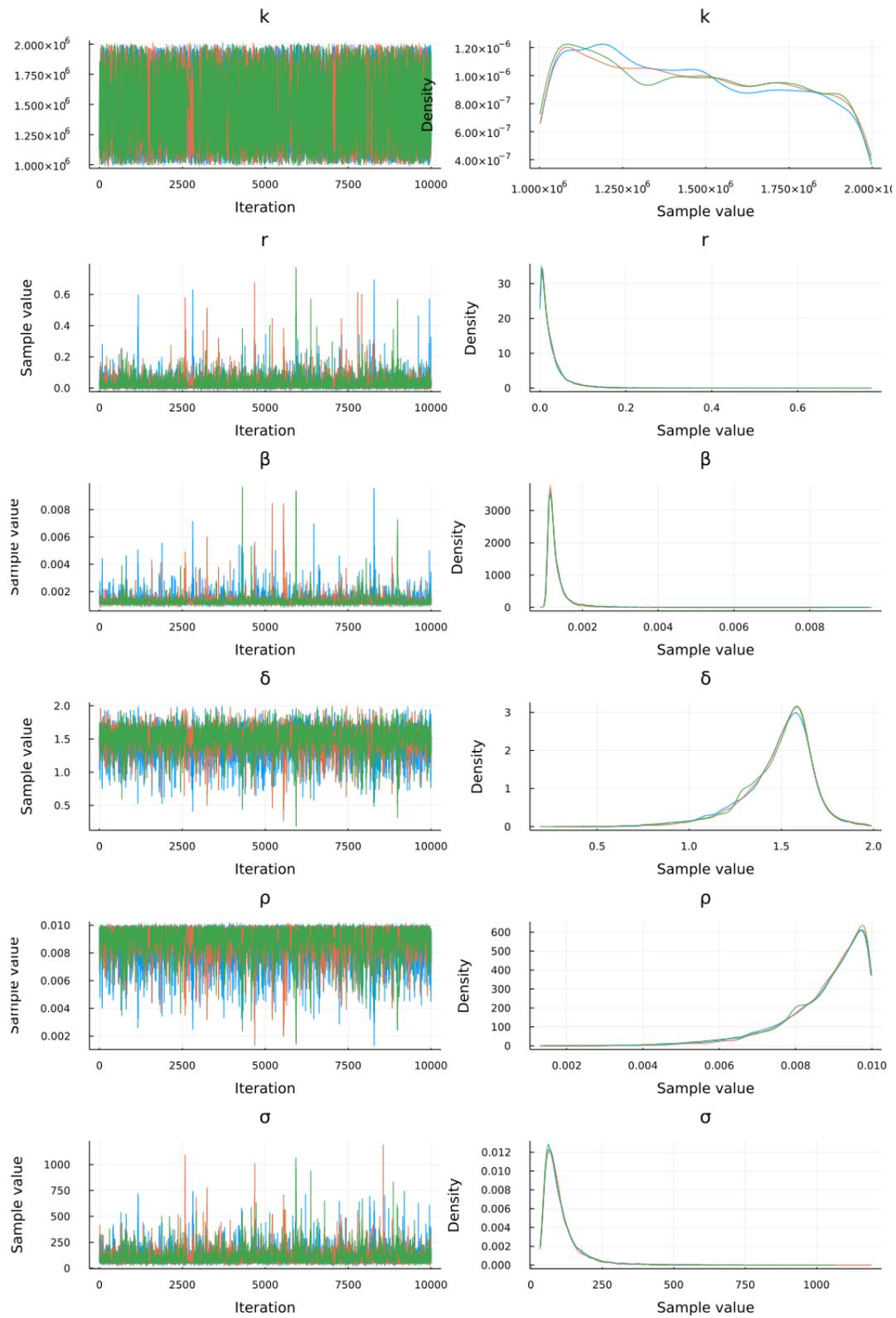
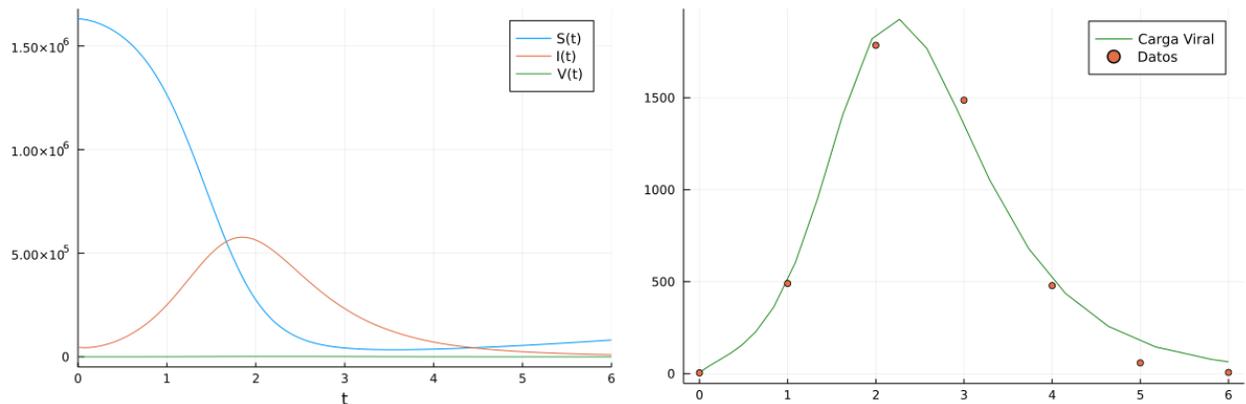


Figura 4.8: Trazas de las cadenas y densidades posteriores de los parámetros Modelo 3.4: von Bertalanffy.

## Ajuste del modelo

Se usaron las estimaciones puntuales correspondientes para calcular la solución numérica del Modelo (3.4). En la Figura 4.9b se puede observar un buen ajuste del modelo ya que la curva ajustada pasa muy cerca de los datos de virus.



(a) Modelo ajustado a las tres poblaciones (Células Sanas, Células Infeccionadas y Carga Viral).

(b) Modelo ajustado a los datos de Carga Viral.

Figura 4.9: Ajuste Modelo 3.4: von Bertalanffy.

## 4.2.2. Modelo Logístico

### Distribuciones a priori

Las distribuciones a priori usadas se muestran en el Cuadro 4.15.

Cuadro 4.15: Distribuciones a priori usadas en el Modelo 3.5: Logístico.

Parámetro	Distribución
$r$	$Gamma(232, 379)$
$k$	$Gamma(4,8 \times 10^3, 3,3 \times 10^{-3})$
$\beta$	$Uniforme(0, 1 \times 10^{-2})$
$\delta$	$Uniforme(0, 2)$
$\rho$	$Uniforme(0, 1 \times 10^{-2})$
$\sigma^2$	$Gamma - Inversa(1 \times 10^{-3}, 1 \times 10^{-3})$

## Estimación

En el Cuadro 4.16 se muestran las estimaciones puntuales e intervalos de alta densidad posterior para cada parámetro en donde se puede observar que: La estimación bayesiana puntual para la tasa de crecimiento de las células,  $r$ , es de 0.61 y que con una probabilidad de 0.95 esta tasa se encuentra entre 0.53 y 0.64; la estimación bayesiana puntual para la capacidad de carga,  $k$ , es de 1443221.9, esto quiere decir que el número máximo de células que puede haber en el medio es 1443222 y que con una probabilidad de 0.95 este número se encuentra entre 1403126 y 1457086; la tasa de infección de células sanas,  $\beta$ , es de 0.0012 y con una probabilidad de 0.95 se encuentra entre 0.0010 y 0.0013; la tasa de mortalidad de células infectadas,  $\delta$ , se estima en 1.64 y con una probabilidad de 0.95 se encuentra entre 1.08 y 1.83; la tasa de producción de virus,  $\rho$ , se estima en 0.0068 y con una probabilidad de 0.95 se encuentra entre 0.0068 y 0.0109.

Cuadro 4.16: Estimaciones Modelo 3.5: Logístico.

Parámetro	Media	L. Inferior HPDI	L. Superior HPDI
$r$	0.6088	0.5311	0.6356
$k$	1443222	1403126	1457086
$\beta$	0.0012	0.0010	0.0013
$\delta$	1.6399	1.0808	1.9811
$\rho$	0.0099	0.0068	0.0109
$\sigma$	82	38	115

## Diagnósticos de convergencia

Los resultados del diagnóstico de Gelman-Rubin mostraron un valor menor que 1.1 del factor de reducción de escala potencial para cada parámetro y un límite superior de su intervalo creíble por debajo de 1.2 en todos los casos, esto indica que las cadenas han convergido a la distribución estacionaria (Cuadro 4.17).

Cuadro 4.17: Diagnóstico de Gelman-Rubin Modelo 3.5: Logístico.

Parámetro	Estimación puntual	L. Superior
$r$	1.00	1.00
$k$	1.00	1.00
$\beta$	1.01	1.01
$\delta$	1.00	1.00
$\rho$	1.00	1.00
$\sigma$	1.00	1.01

Los resultados del diagnóstico de Geweke muestran que todos los valores de la estadística de prueba  $Z$  para cada parámetro y cada cadena se encuentran entre -2 y 2, lo cual indica que las medias de las primeras y las últimas iteraciones de cada cadena son estadísticamente iguales, y por tanto existe convergencia de las mismas (Cuadro 4.18).

Cuadro 4.18: Diagnóstico de Geweke Modelo 3.5: Logístico.

Cadena	Parámetros					
	$r$	$k$	$\beta$	$\delta$	$\rho$	$\sigma$
1	-0.1112	-1.1232	-1.0163	0.8062	0.8832	-0.3823
2	0.5102	0.5903	0.0993	-0.3634	-0.4543	-0.0405
3	-0.8821	-0.3142	0.6028	-0.6086	-0.7466	-0.5731

Los diagnósticos de convergencia gráficos como las trazas de las cadenas y las densidades posteriores de los parámetros son congruentes con los resultados anteriores. Se puede observar un buen mezclado de las cadenas y además que recorren todo el espacio paramétrico de su distribución posterior indicando convergencia (Figura 4.10).

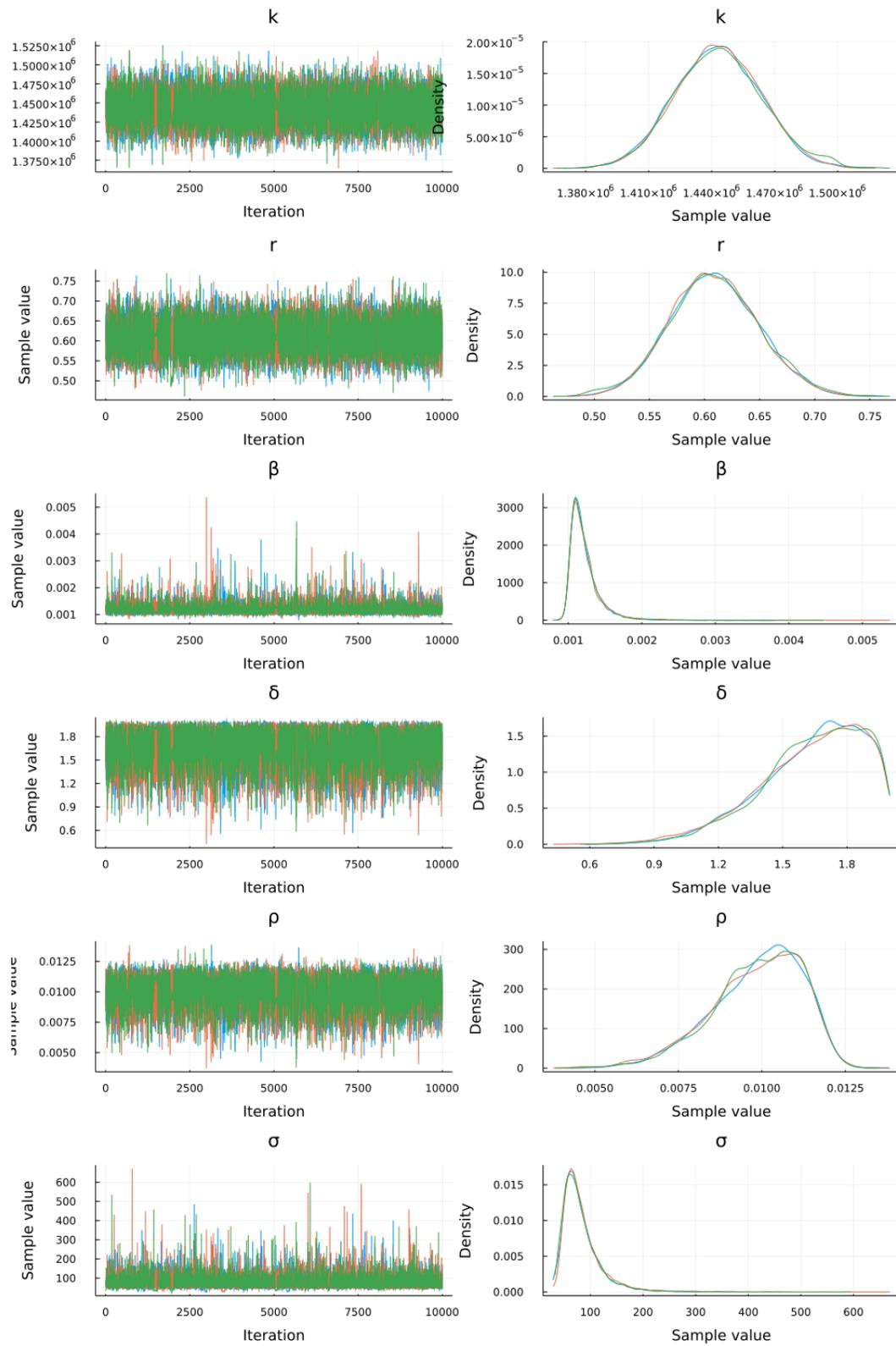
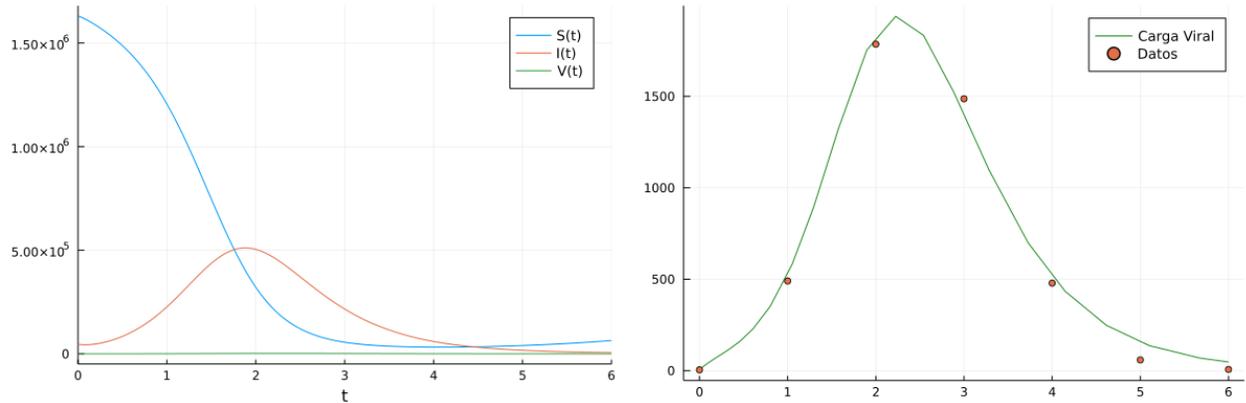


Figura 4.10: Trazas de las cadenas y densidades posteriores de los parámetros Modelo 3.5: Logístico.

## Ajuste del modelo

Se usaron las estimaciones puntuales correspondientes para calcular la solución numérica del Modelo (3.5). En la Figura 4.11b se puede observar un buen ajuste del modelo ya que la curva ajustada pasa muy cerca de los datos de virus.



(a) Modelo ajustado a las tres poblaciones (Células Sanas, Células Infectadas y Carga Viral).

(b) Modelo ajustado a los datos de Carga Viral.

Figura 4.11: Ajuste Modelo 3.5: Logístico.

### 4.2.3. Modelo Gompertz

#### Distribuciones a priori

Las distribuciones a priori usadas se muestran en el Cuadro 4.19.

Cuadro 4.19: Distribuciones a priori usadas en el Modelo 3.6: Gompertz.

Parámetro	Distribución
$r$	$Gamma(247, 536)$
$k$	$Gamma(5,3 \times 10^3, 3,6 \times 10^{-3})$
$\beta$	$Uniforme(0, 1 \times 10^{-2})$
$\delta$	$Uniforme(0, 2)$
$\rho$	$Uniforme(0, 1 \times 10^{-2})$
$\sigma^2$	$Gamma - Inversa(1 \times 10^{-3}, 1 \times 10^{-3})$

## Estimación

En el Cuadro 4.20 se muestran las estimaciones puntuales e intervalos de alta densidad posterior para cada parámetro en donde se puede observar que: La estimación bayesiana puntual para la tasa de crecimiento de las células,  $r$ , es de 0.46 y que con una probabilidad de 0.95 esta tasa se encuentra entre 0.40 y 0.48; la estimación bayesiana puntual para la capacidad de carga,  $k$ , es de 1473170.85, esto quiere decir que el número máximo de células que puede haber en el medio es 1473171 y que con una probabilidad de 0.95 este número se encuentra entre 1434124 y 1486818; la tasa de infección de células sanas,  $\beta$ , es de 0.0015 y con una probabilidad de 0.95 se encuentra entre 0.0012 y 0.0016; la tasa de mortalidad de células infectadas,  $\delta$ , se estima en 1.37 y con una probabilidad de 0.95 se encuentra entre 0.83 y 1.55; la tasa de producción de virus,  $\rho$ , se estima en 0.0082 y con una probabilidad de 0.95 se encuentra entre 0.0054 y 0.0091.

Cuadro 4.20: Estimaciones Modelo 3.6: Gompertz.

Parámetro	Media	L. Inferior HPDI	L. Superior HPDI
$r$	0.4575	0.4022	0.4768
$k$	1473171	1434124	1486818
$\beta$	0.0015	0.0012	0.0016
$\delta$	1.3706	0.8307	1.5516
$\rho$	0.0082	0.0054	0.0091
$\sigma$	115	56	143

## Diagnósticos de convergencia Modelo 3.6: Gompertz

Los resultados del diagnóstico de Gelman-Rubin mostraron un valor menor que 1.1 del factor de reducción de escala potencial para cada parámetro y un límite superior de su intervalo creíble por debajo de 1.2 en todos los casos, esto indica que las cadenas han convergido a la distribución estacionaria (Cuadro 4.21).

Cuadro 4.21: Diagnóstico de Gelman-Rubin Modelo 3.6: Gompertz.

Parámetro	Estimación puntual	L. Superior
$r$	1.00	1.00
$k$	1.00	1.00
$\beta$	1.03	1.04
$\delta$	1.00	1.00
$\rho$	1.00	1.00
$\sigma$	1.01	1.01

Los resultados del diagnóstico de Geweke muestran que la totalidad de los valores de la estadística de prueba  $Z$  para cada parámetro y cada cadena se encuentran entre -2 y 2, lo cual indica que las medias de las primeras y las últimas iteraciones de cada cadena son estadísticamente iguales, y por tanto existe convergencia de las mismas (Cuadro 4.22).

Cuadro 4.22: Diagnóstico de Geweke Modelo 3.6: Gompertz.

Cadena	Parámetros					
	$r$	$k$	$\beta$	$\delta$	$\rho$	$\sigma$
1	-1.5245	0.0063	-1.9447	2.0050	1.8003	-1.6845
2	-1.6195	-0.2503	-2.2966	1.2802	1.3096	-0.7721
3	-0.1134	-0.3185	0.1296	0.5413	0.4910	-0.3961

Los diagnósticos de convergencia gráficos como las trazas de las cadenas y las densidades posteriores de los parámetros son congruentes con los resultados anteriores. Se puede observar un buen mezclado de las cadenas y además que recorren todo el espacio paramétrico de su distribución posterior indicando convergencia (Figura 4.12).

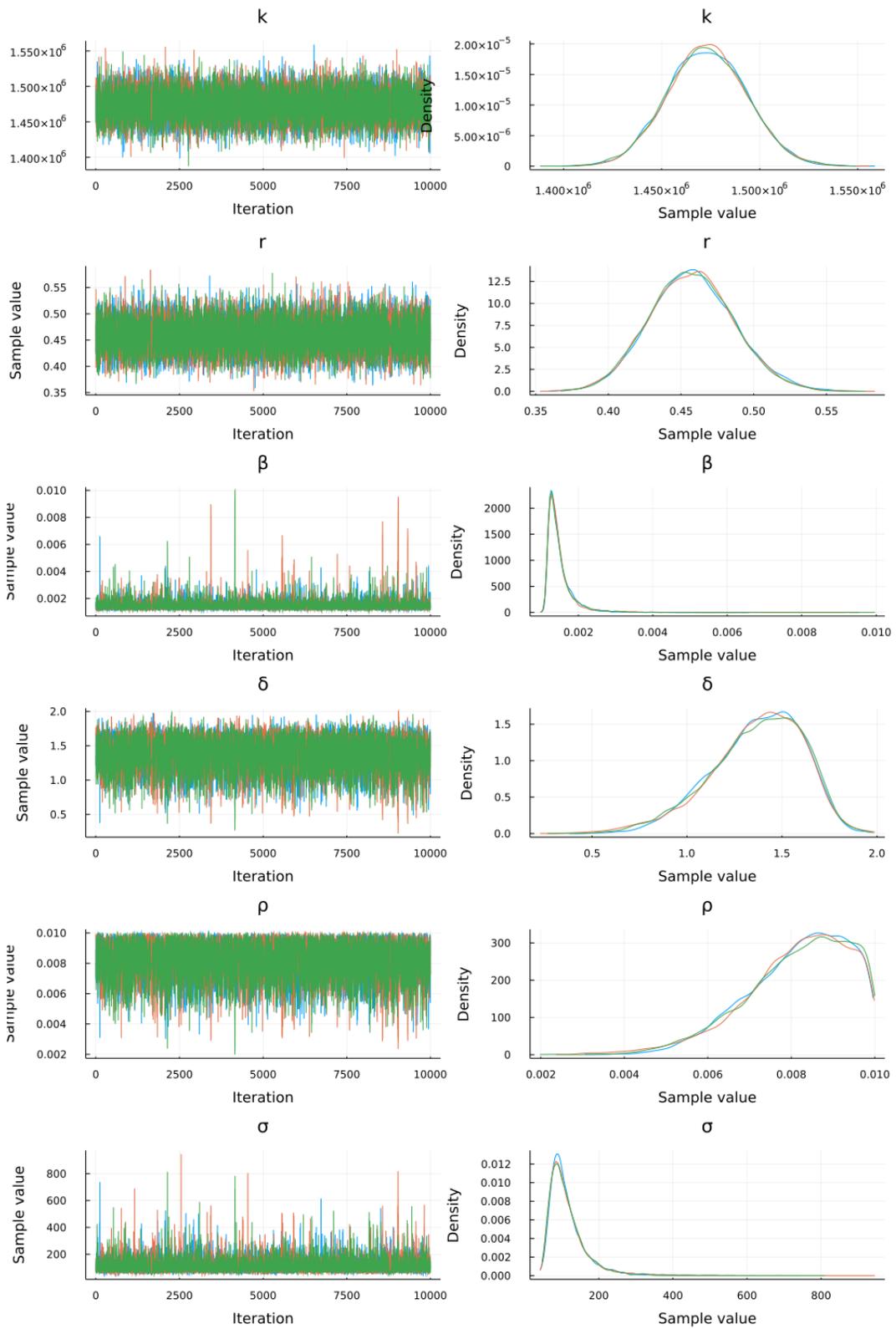
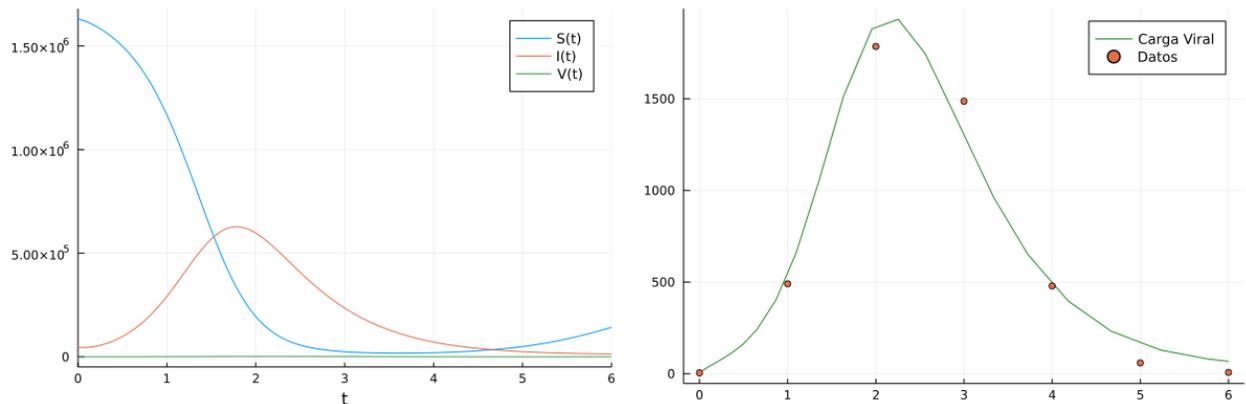


Figura 4.12: Trazas de las cadenas y densidades posteriores de los parámetros Modelo 3.6: Gompertz.

## Ajuste del modelo

Se usaron las estimaciones puntuales correspondientes para calcular la solución numérica del Modelo (3.6). En la Figura 4.13b se puede observar un buen ajuste del modelo ya que la curva ajustada pasa muy cerca de los datos de virus.



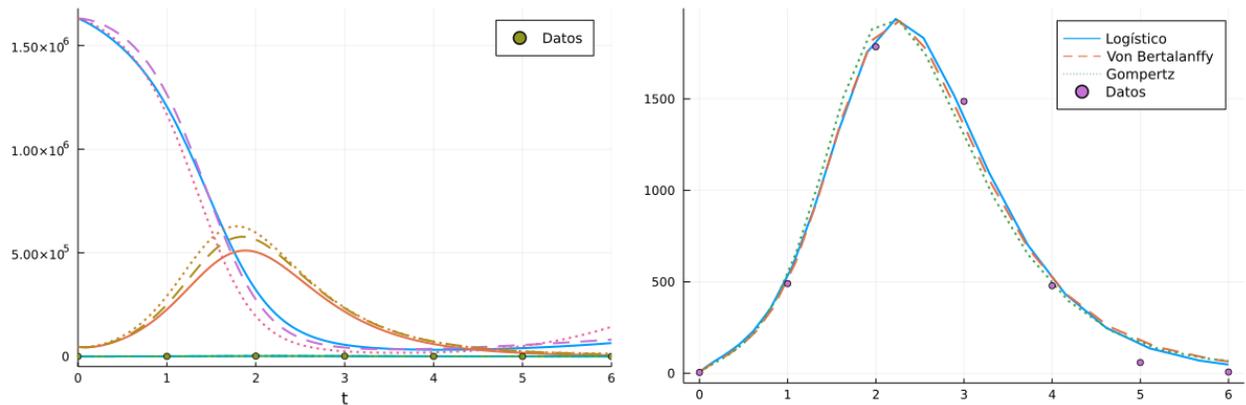
(a) Modelo ajustado a las tres poblaciones (Células Sanas, Células Infeccionadas y Carga Viral).

(b) Modelo ajustado a los datos de Carga Viral.

Figura 4.13: Ajuste Modelo 3.6: Gompertz.

### 4.2.4. Comparación de modelos

Con el propósito de seleccionar el modelo que se ajuste mejor a los datos de carga viral en ausencia de interferon, en primer lugar se hace una comparación gráfica de los modelos ajustados. En la Figura 4.14a se muestran los ajustes de los Modelos Logístico (Línea sólida), von Bertalanffy (Línea a rayas) y Gompertz (Línea punteada) a las tres poblaciones. En la Figura 4.14b se puede observar que las curvas de los ajustes son muy similares, y por tanto es necesario utilizar otros criterios de comparación como el criterio de Información de Watanabe-Akaike (WAIC). En el Cuadro 4.23 se muestran los resultados de esta comparación, donde se puede observar que el modelo que presenta menor WAIC y por tanto se ajusta mejor a los datos, es el Modelo Logístico.



(a) Modelos ajustados a las tres poblaciones (Células Sanas, Células Infectadas y Carga Viral). (b) Modelos ajustados a los datos de carga viral.

Figura 4.14: Comparación de Modelos Etapa 2.

Cuadro 4.23: Criterios de Comparación de Modelos.

Modelo	WAIC
von Bertalanffy	171.8
Logístico	144.6
Gompertz	167.8

### 4.3. Etapa 3: Datos de Carga Viral en presencia de interferón

En esta etapa se usó el tercer conjunto de datos correspondiente a los datos de carga viral del tercer cultivo para ajustar los Modelos (3.7), (3.8) y (3.9) (modelos con función de crecimiento von Bertalanffy, Logística y Gompertz).

Los hiperparámetros de la distribución Gamma para los parámetros de dinámica celular  $r$  y  $k$  y de dinámica viral  $\beta$ ,  $\delta$  y  $\rho$  se obtuvieron al ajustar una distribución gamma a las muestras obtenidas mediante MCMC en la segunda etapa para el parámetro correspondiente. Se asignaron distribuciones uniformes para los parámetros del efecto del Interferón  $\eta$  y  $\epsilon$  cuyo soporte se eligió de acuerdo a la literatura consultada y para darle

sentido biológico [22]. El parámetro  $\gamma$  se fijó en 2.3. Para el parámetro  $\sigma^2$  se asignó una distribución Gamma Inversa no informativa.

### 4.3.1. Modelo von Bertalanffy

#### Distribuciones a priori

Las distribuciones a priori usadas se muestran en el Cuadro 4.24.

Cuadro 4.24: Distribuciones a priori usadas en el Modelo 3.7: von Bertalanffy.

Parámetro	Distribución
$r$	$Gamma(158, 473)$
$k$	$Gamma(256, 3, 1,8 \times 10^{-5})$
$\beta$	$Gamma(65, 53911)$
$\delta$	$Gamma(129, 85)$
$\rho$	$Gamma(145, 15779)$
$\eta$	$Uniforme(0, 1)$
$\epsilon$	$Uniforme(0, 1)$
$\sigma^2$	$Gamma - Inversa(1 \times 10^{-3}, 1 \times 10^{-3})$

#### Estimación

En el Cuadro 4.25 se muestran las estimaciones puntuales e intervalos de alta densidad posterior para cada parámetro en donde se puede observar que: La estimación bayesiana puntual para la tasa de crecimiento de las células,  $r$ , es de 0.17 y que con una probabilidad de 0.95 esta tasa se encuentra entre 0.07 y 0.24; la estimación bayesiana puntual para la capacidad de carga,  $k$ , es de 1646152.85, esto quiere decir que el número máximo de células que puede haber en el medio es 1646153 y que con una probabilidad de 0.95 este número se encuentra entre 1153781 y 1922844; la tasa de infección de células sanas,  $\beta$ , es de 0.0015 y con una probabilidad de 0.95 se encuentra entre 0.0012 y 0.0016; la tasa de mortalidad de células infectadas,  $\delta$ , se estima en 1.49 y con una probabilidad de 0.95 se encuentra entre 1.25 y 1.58; la tasa de producción de virus,  $\rho$ , se estima en 0.0101 y con una probabilidad de 0.95 se encuentra entre 0.0087 y 0.0109;

el efecto del interferón sobre la producción de células infectadas,  $\eta$ , es del 6.09 % y se encuentra entre 0.140 % y 10.32 % con una probabilidad de 0.95; el efecto del interferón sobre la producción de partículas virales,  $\epsilon$ , es del 6.31 % y con una probabilidad de 0.95 este porcentaje se encuentra entre el 0.170 % y el 11.52 %.

Cuadro 4.25: Estimaciones Modelo 3.7: von Bertalanffy.

Parámetro	Media	L. Inferior HPDI	L. Superior HPDI
$r$	0.1673	0.0665	0.2369
$k$	1646152	1153781	1922844
$\beta$	0.0015	0.0012	0.0016
$\delta$	1.4934	1.2492	1.5794
$\rho$	0.0101	0.0087	0.0109
$\eta$	0.0609	0.0014	0.1032
$\epsilon$	0.0631	0.0017	0.1152
$\sigma$	181	94	251

### Diagnósticos de convergencia

Los resultados del diagnóstico de Gelman-Rubin mostraron un valor menor que 1.1 del factor de reducción de escala potencial para cada parámetro y un límite superior de su intervalo creíble por debajo de 1.2 en todos los casos, esto indica que las cadenas han convergido a la distribución estacionaria (Cuadro 4.26).

Cuadro 4.26: Diagnóstico de Gelman-Rubin Modelo 3.7: von Bertalanffy.

Parámetro	Estimación puntual	L. Superior
$r$	1.00	1.00
$k$	1.00	1.00
$\beta$	1.00	1.00
$\delta$	1.00	1.00
$\rho$	1.00	1.00
$\eta$	1.00	1.01
$\epsilon$	1.00	1.00
$\sigma$	1.00	1.00

Los resultados del diagnóstico de Geweke muestran que la totalidad de los valores de la estadística de prueba  $Z$  para cada parámetro y cada cadena se encuentran entre -2 y 2, lo cual indica que las medias de las primeras y las últimas iteraciones de cada cadena son estadísticamente iguales, y por tanto existe convergencia de las mismas (Cuadro 4.27).

Cuadro 4.27: Diagnóstico de Geweke Modelo 3.7: von Bertalanffy.

Cadena	Parámetros					
	$r$	$k$	$\beta$	$\delta$	$\rho$	$\sigma$
1	-1.2206	0.2777	-0.1612	-0.4283	-0.2311	-1.0046
2	-0.9109	0.5290	-1.8911	1.4887	1.4388	-1.4825
3	-0.7071	-0.0429	-0.8230	-0.5825	-0.0598	-0.6869

Los diagnósticos de convergencia gráficos como las trazas de las cadenas y las densidades posteriores de los parámetros son congruentes con los resultados anteriores. Se puede observar un buen mezclado de las cadenas y además que recorren todo el espacio paramétrico de su distribución posterior indicando convergencia (Figura 4.15).

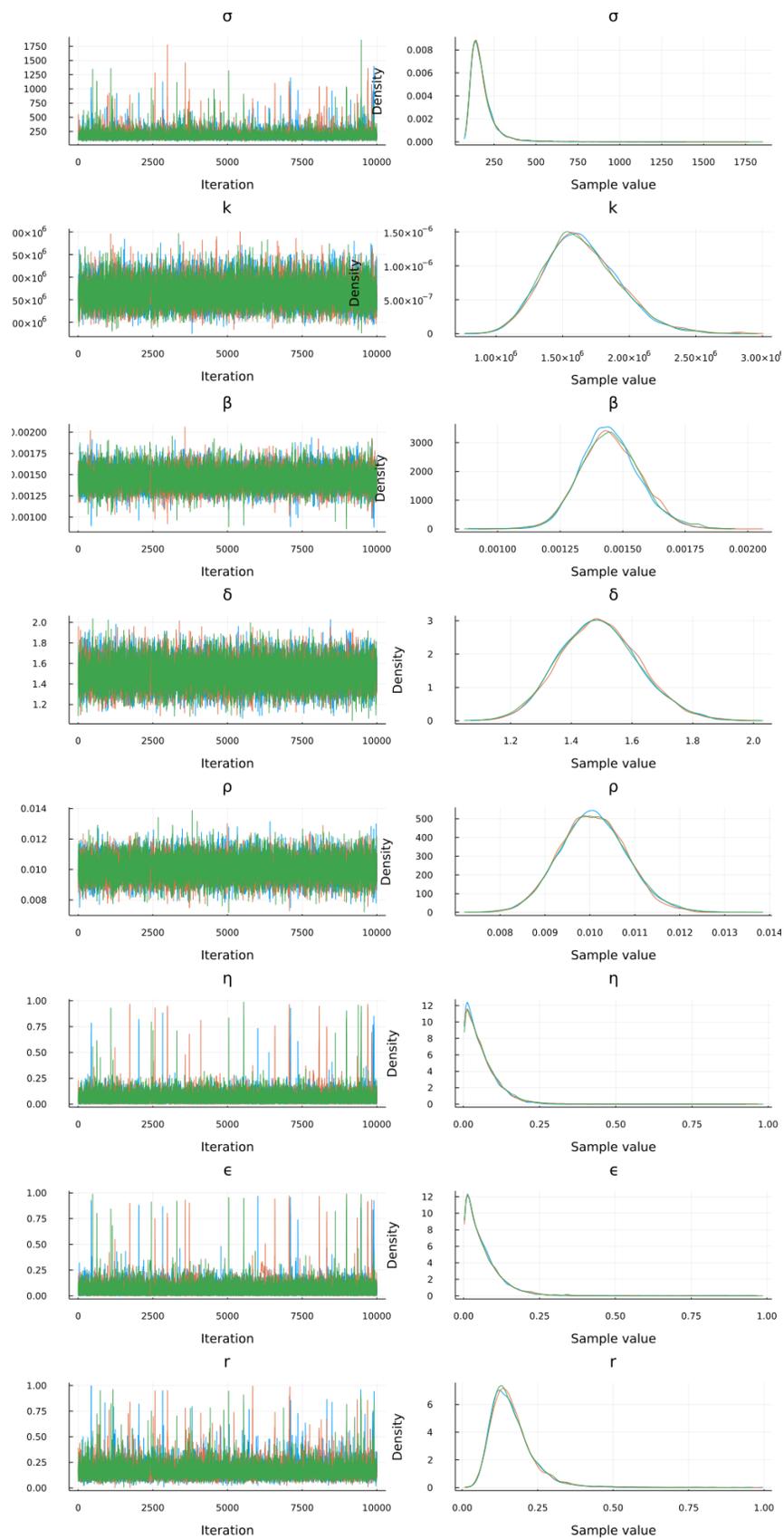
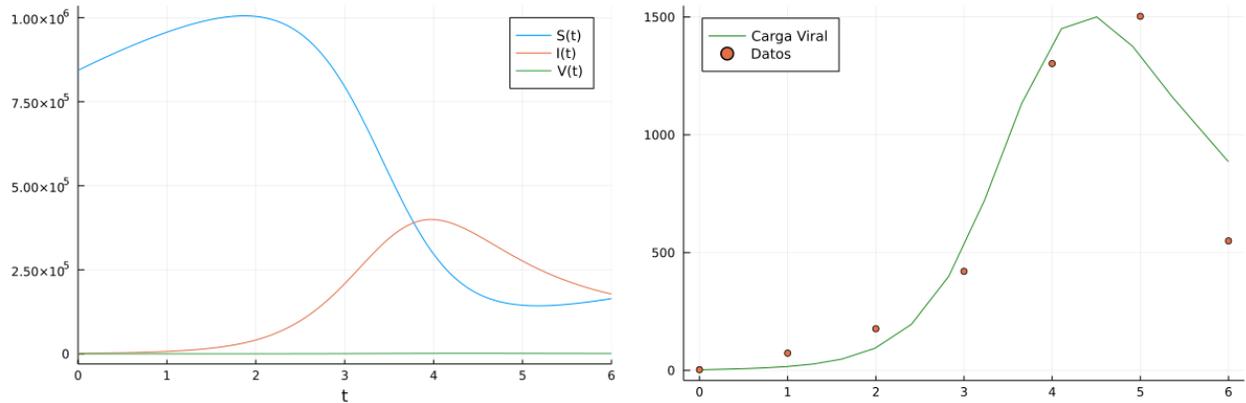


Figura 4.15: Trazas de las cadenas y densidades posteriores de los parámetros Modelo 3.7: von Bertalanffy.

## Ajuste del modelo

Se usaron las estimaciones puntuales correspondientes para calcular la solución numérica del Modelo (3.7). En la Figura 4.16b se puede observar un buen ajuste del modelo ya que la curva ajustada pasa muy cerca de los datos de virus.



(a) Modelo ajustado a las tres poblaciones (Células Sanas, Células Infectadas y Carga Viral).

(b) Modelo ajustado a los datos de Carga Viral.

Figura 4.16: Ajuste Modelo 3.7: von Bertalanffy.

### 4.3.2. Modelo Logístico

#### Distribuciones a priori

Las distribuciones a priori usadas se muestran en el Cuadro 4.28.

Cuadro 4.28: Distribuciones a priori usadas en el Modelo 3.8: Logístico

Parámetro	Distribución
$r$	$Gamma(235, 387)$
$k$	$Gamma(4771, 3,3 \times 10^{-3})$
$\beta$	$Gamma(50, 42104)$
$\delta$	$Gamma(60, 36)$
$\rho$	$Gamma(70, 6994)$
$\eta$	$Uniforme(0, 1)$
$\epsilon$	$Uniforme(0, 1)$
$\sigma^2$	$Gamma - Inversa(1 \times 10^{-3}, 1 \times 10^{-3})$

### Estimación

En el Cuadro 4.29 se muestran las estimaciones puntuales e intervalos de alta densidad posterior para cada parámetro en donde se puede observar que: La estimación bayesiana puntual para la tasa de crecimiento de las células,  $r$ , es de 0.61 y que con una probabilidad de 0.95 esta tasa se encuentra entre 0.54 y 0.69; la estimación bayesiana puntual para la capacidad de carga,  $k$ , es de 1447744.78, esto quiere decir que el número máximo de células que puede haber en el medio es 1447745 y que con una probabilidad de 0.95 este número se encuentra entre 1406889 y 1487427; la tasa de infección de células sanas,  $\beta$ , es de 0.0013 y con una probabilidad de 0.95 se encuentra entre 0.0011 y 0.0016; la tasa de mortalidad de células infectadas,  $\delta$ , se estima en 1.60 y con una probabilidad de 0.95 se encuentra entre 1.29 y 1.91; la tasa de producción de virus,  $\rho$ , se estima en 0.0111 y con una probabilidad de 0.95 se encuentra entre 0.0093 y 0.0130; el efecto del interferón sobre la producción de células infectadas,  $\eta$ , es del 9.93 % y se encuentra entre 0.39 % y 26.49 % con una probabilidad de 0.95; el efecto del interferón sobre la producción de partículas virales,  $\epsilon$ , es del 8.12 % y con una probabilidad de 0.95 este porcentaje se encuentra entre el 0.30 % y el 23.07 %.

Cuadro 4.29: Estimaciones Modelo 3.8: Logístico.

Parámetro	Media	L. Inferior HPDI	L. Superior HPDI
$r$	0.6102	0.5354	0.6887
$k$	1447745	1406889	1487427
$\beta$	0.0013	0.0011	0.0016
$\delta$	1.6004	1.2915	1.9148
$\rho$	0.0111	0.0093	0.0130
$\eta$	0.0993	0.0039	0.2649
$\epsilon$	0.0812	0.0030	0.2307
$\sigma$	105	55	171

### Diagnósticos de convergencia

Los resultados del diagnóstico de Gelman-Rubin mostraron un valor menor que 1.1 del factor de reducción de escala potencial para cada parámetro y un límite superior de su intervalo creíble por debajo de 1.2 en todos los casos, esto indica que las cadenas han convergido a la distribución estacionaria (Cuadro 4.30).

Cuadro 4.30: Diagnóstico de Gelman-Rubin Modelo 3.8: Logístico.

Parámetro	Estimación puntual	L. Superior
$r$	1.00	1.00
$k$	1.00	1.00
$\beta$	1.01	1.01
$\delta$	1.00	1.00
$\rho$	1.00	1.00
$\eta$	1.01	1.02
$\epsilon$	1.00	1.00
$\sigma$	1.00	1.00

Los resultados del diagnóstico de Geweke muestran que todos los valores de la estadística de prueba  $Z$  para cada parámetro y cada cadena se encuentran entre -2 y 2, lo

cual indica que las medias de las primeras y las últimas iteraciones de cada cadena son estadísticamente iguales, y por tanto existe convergencia de las mismas (Cuadro 4.31).

Cuadro 4.31: Diagnóstico de Geweke Modelo 3.8: Logístico.

Cadena	Parámetros					
	$r$	$k$	$\beta$	$\delta$	$\rho$	$\sigma$
1	-0.1112	-1.1232	-1.0163	0.8062	0.8832	-0.3823
2	0.5102	0.5903	0.0993	-0.3634	-0.4543	-0.0405
3	-0.8821	-0.3142	0.6028	-0.6086	-0.7466	-0.5731

Los diagnósticos de convergencia gráficos como las trazas de las cadenas y las densidades posteriores de los parámetros son congruentes con los resultados anteriores. Se puede observar un buen mezclado de las cadenas y además que recorren todo el espacio paramétrico de su distribución posterior indicando convergencia (Figura 4.17).

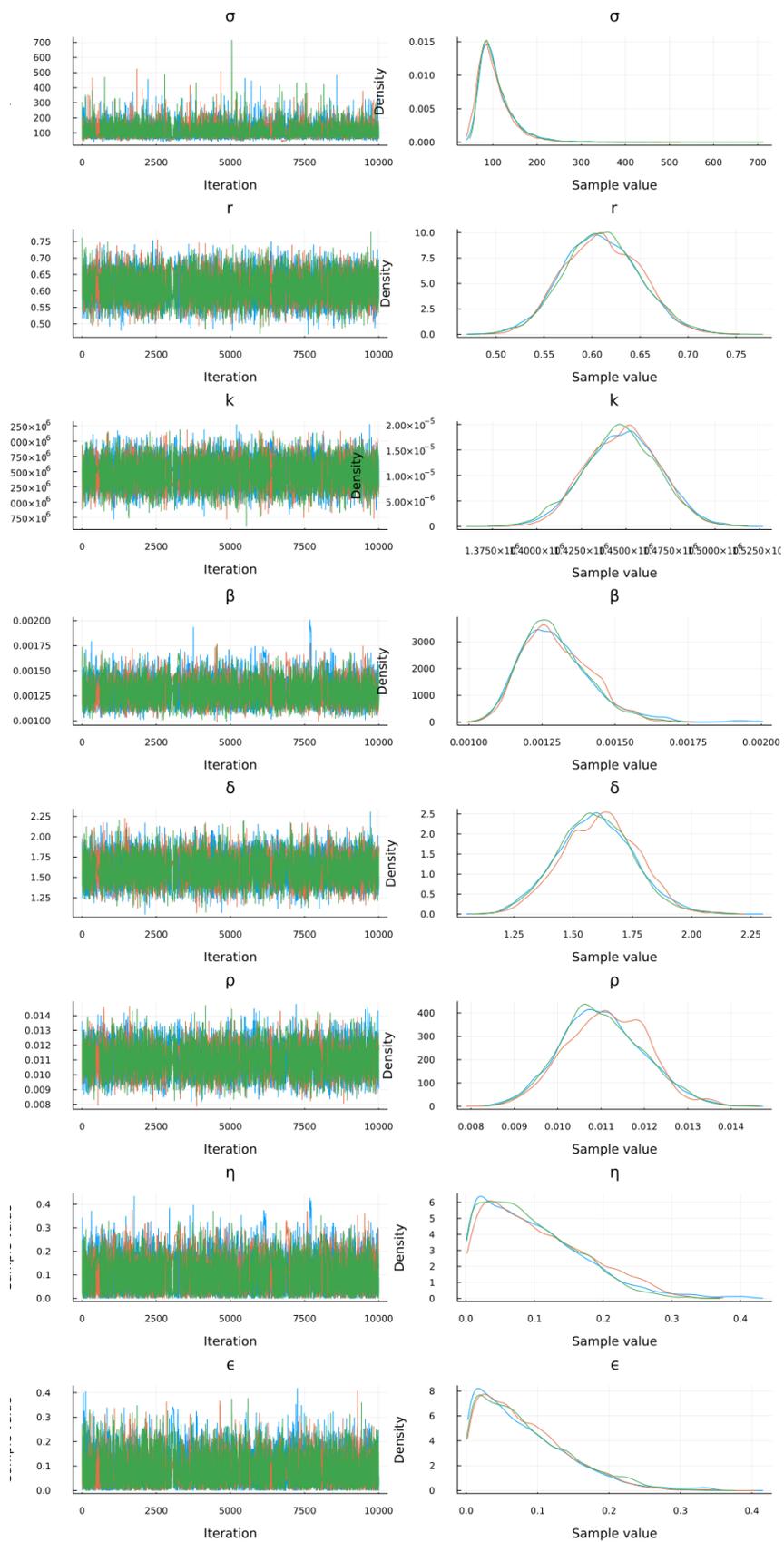
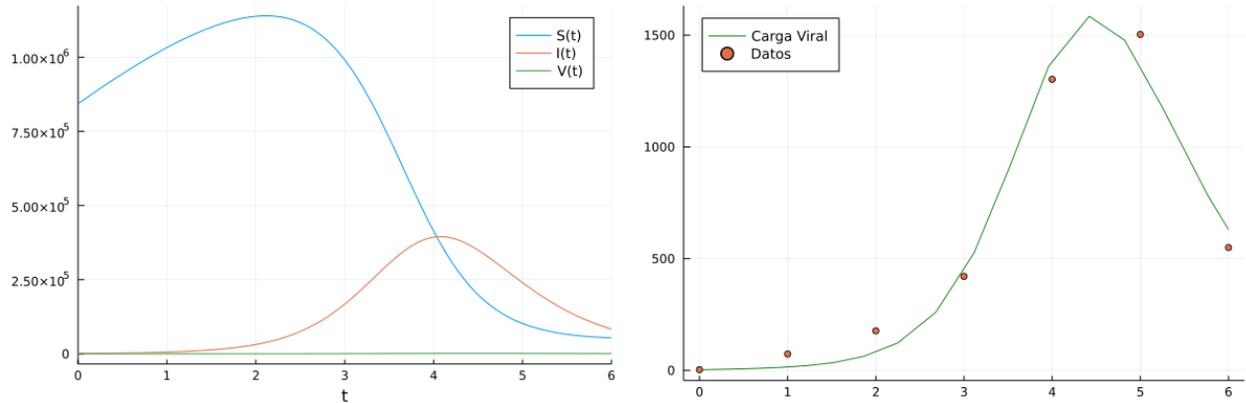


Figura 4.17: Trazas de las cadenas y densidades posteriores de los parámetros Modelo 3.8: Logístico.

## Ajuste del modelo

Se usaron las estimaciones puntuales correspondientes para calcular la solución numérica del Modelo (3.8). En la Figura 4.18b se puede observar un buen ajuste del modelo ya que la curva ajustada pasa muy cerca de los datos de virus.



(a) Modelo ajustado a las tres poblaciones (Células Sanas, Células Infectadas y Carga Viral).

(b) Modelo ajustado a los datos de Carga Viral.

Figura 4.18: Ajuste Modelo 3.8: Logístico.

### 4.3.3. Modelo Gompertz

#### Distribuciones a priori

Las distribuciones a priori usadas se muestran en el Cuadro 4.32.

Cuadro 4.32: Distribuciones a priori usadas en el Modelo 3.9: Gompertz.

Parámetro	Distribución
$r$	$Gamma(247, 544)$
$k$	$Gamma(5148, 3,5 \times 10^{-3})$
$\beta$	$Gamma(29, 20290)$
$\delta$	$Gamma(46, 33)$
$\rho$	$Gamma(64, 7573)$
$\eta$	$Uniforme(0, 1)$
$\epsilon$	$Uniforme(0, 1)$
$\sigma^2$	$Gamma - Inversa(1 \times 10^{-3}, 1 \times 10^{-3})$

### Estimación

En el Cuadro 4.33 se muestran las estimaciones puntuales e intervalos de alta densidad posterior para cada parámetro en donde se puede observar que: La estimación bayesiana puntual para la tasa de crecimiento de las células,  $r$ , es de 0.59 y que con una probabilidad de 0.95 esta tasa se encuentra entre 0.19 y 0.96; la estimación bayesiana puntual para la capacidad de carga,  $k$ , es de 1717349.11, esto quiere decir que el número máximo de células que puede haber en el medio es 1717350 y que con una probabilidad de 0.95 este número se encuentra entre 1201709 y 1989034; la tasa de infección de células sanas,  $\beta$ , es de 0.0016 y con una probabilidad de 0.95 se encuentra entre 0.0013 y 0.0020; la tasa de mortalidad de células infectadas,  $\delta$ , se estima en 1.42 y con una probabilidad de 0.95 se encuentra entre 0.81 y 1.94; la tasa de producción de virus,  $\rho$ , se estima en 0.0090 y con una probabilidad de 0.95 se encuentra entre 0.0071 y 0.0111; el efecto del interferón sobre la producción de células infectadas,  $\eta$ , es del 13.47 % y se encuentra entre 0.63 % y 33.98 % con una probabilidad de 0.95; el efecto del interferón sobre la producción de partículas virales,  $\epsilon$ , es del 13.31 % y con una probabilidad de 0.95 este porcentaje se encuentra entre el 0.48 % y el 37.06 %.

Cuadro 4.33: Estimaciones Modelo 3.9: Gompertz.

Parámetro	Media	L. Inferior HPDI	L. Superior HPDI
$r$	0.5874	0.1915	0.9648
$k$	1717350	1201709	1989034
$\beta$	0.0016	0.0013	0.0020
$\delta$	1.4179	0.8095	1.9446
$\rho$	0.0090	0.0071	0.0111
$\eta$	0.1347	0.0063	0.3398
$\epsilon$	0.1331	0.0048	0.3706
$\sigma$	104	59	139

### Diagnósticos de convergencia

Los resultados del diagnóstico de Gelman-Rubin mostraron un valor menor que 1.1 del factor de reducción de escala potencial para cada parámetro y un límite superior de su intervalo creíble por debajo de 1.2 en todos los casos, esto indica que las cadenas han convergido a la distribución estacionaria (Cuadro 4.34).

Cuadro 4.34: Diagnóstico de Gelman-Rubin Modelo 3.9: Gompertz.

Parámetro	Estimación puntual	L. Superior
$r$	1.00	1.00
$k$	1.00	1.00
$\beta$	1.00	1.00
$\delta$	1.00	1.00
$\rho$	1.00	1.00
$\eta$	1.00	1.00
$\epsilon$	1.00	1.00
$\sigma$	1.00	1.01

Los resultados del diagnóstico de Geweke muestran que la totalidad de los valores de la estadística de prueba  $Z$  para cada parámetro y cada cadena se encuentran entre -2 y

2, lo cual indica que las medias de las primeras y las últimas iteraciones de cada cadena son estadísticamente iguales, y por tanto existe convergencia de las mismas (Cuadro 4.35).

Cuadro 4.35: Diagnóstico de Geweke Modelo 3.9: Gompertz.

Cadena	Parámetros					
	$r$	$k$	$\beta$	$\delta$	$\rho$	$\sigma$
1	-1.5245	0.0063	-1.9447	2.0050	1.8003	-1.6845
2	-1.6195	-0.2503	-2.2966	1.2802	1.3096	-0.7721
3	-0.1134	-0.3185	0.1296	0.5413	0.4910	-0.3961

Los diagnósticos de convergencia gráficos como las trazas de las cadenas y las densidades posteriores de los parámetros son congruentes con los resultados anteriores. Se puede observar un buen mezclado de las cadenas y además que recorren todo el espacio paramétrico de su distribución posterior indicando convergencia (Figura 4.19).

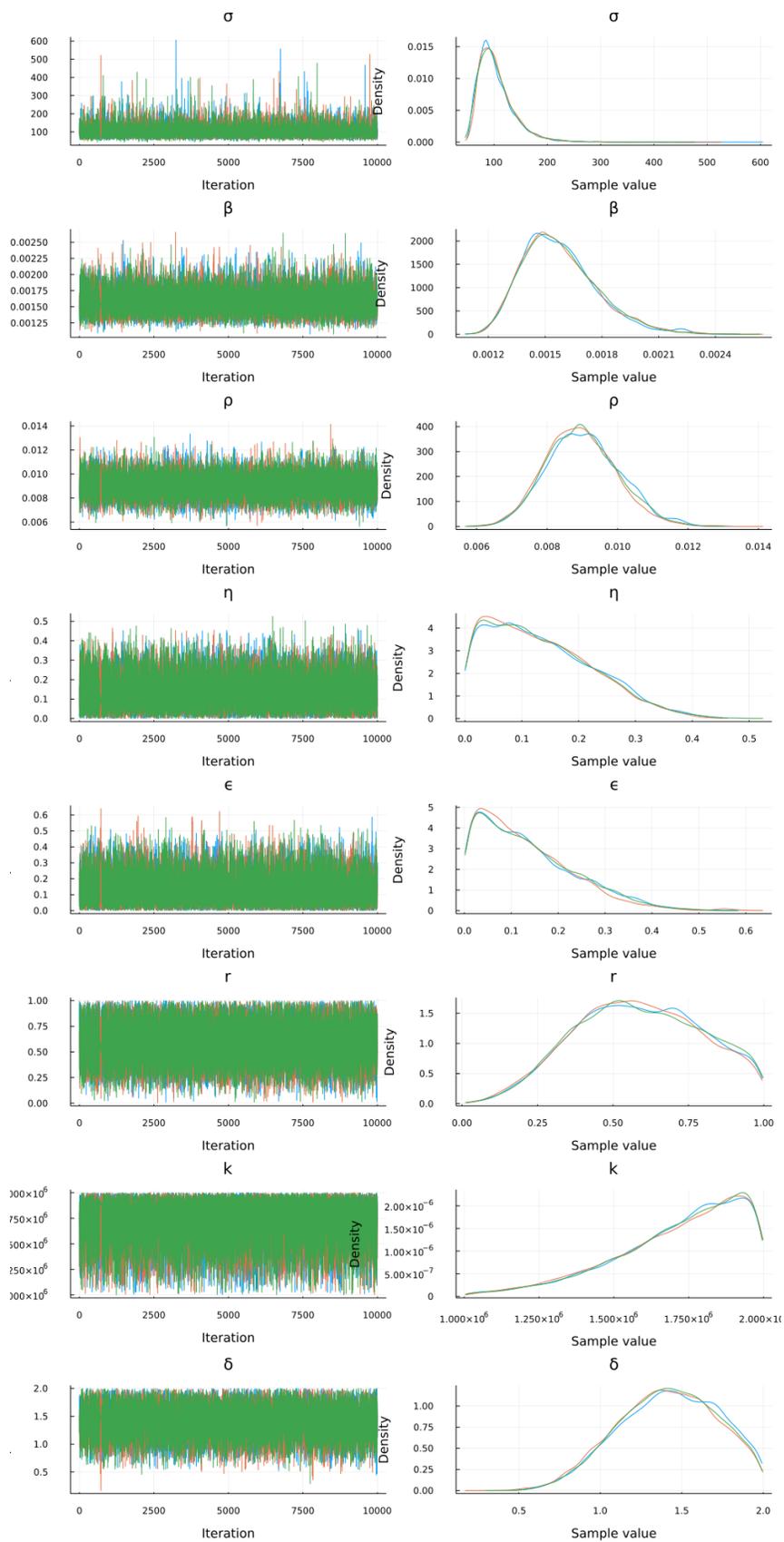
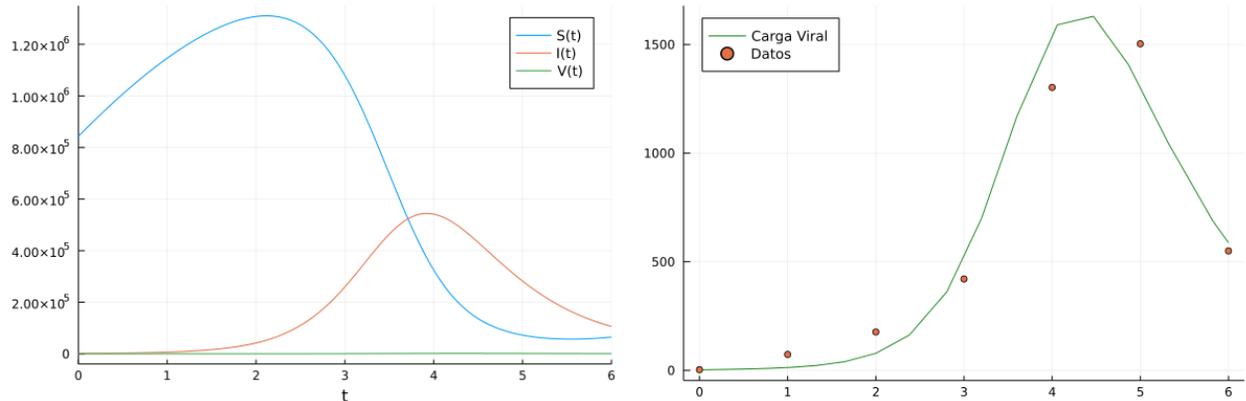


Figura 4.19: Trazas de las cadenas y densidades posteriores de los parámetros Modelo 3.9: Gompertz.

## Ajuste del modelo

Se usaron las estimaciones puntuales correspondientes para calcular la solución numérica del Modelo (3.9). En la Figura 4.20b se puede observar un buen ajuste del modelo ya que la curva ajustada pasa cerca de los datos de virus.



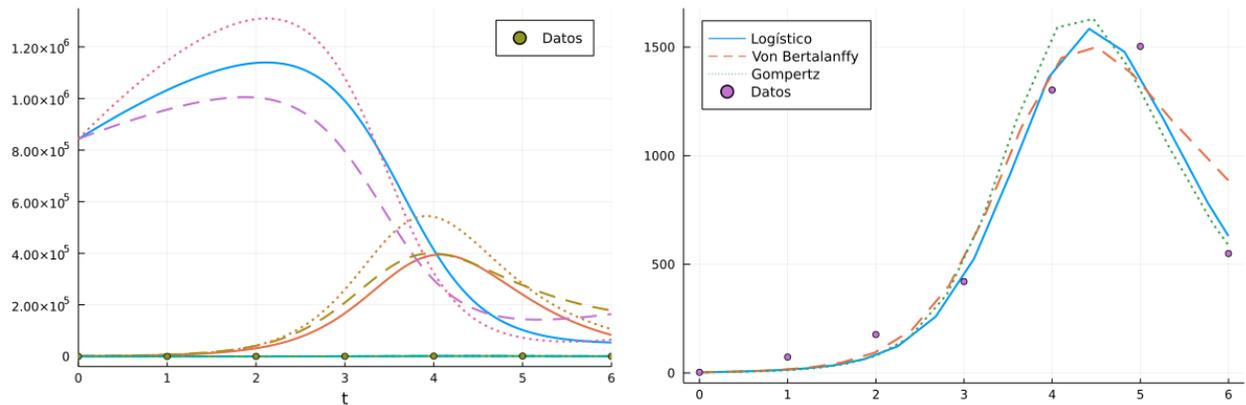
(a) Modelo ajustado a las tres poblaciones (Células Sanas, Células Infeccionadas y Carga Viral).

(b) Modelo ajustado a los datos de Carga Viral.

Figura 4.20: Ajuste Modelo 3.9: Gompertz.

### 4.3.4. Comparación de modelos

Con el propósito de seleccionar el modelo que se ajuste mejor a los datos de carga viral en presencia de interferón, en primer lugar se hace una comparación gráfica de los modelos ajustados. En la Figura 4.21a se muestran los ajustes de los Modelos Logístico (Línea sólida), von Bertalanffy (Línea a rayas) y Gompertz (Línea punteada) a las tres poblaciones. En la Figura 4.21b se muestra el ajuste de los modelos a los datos de Carga Viral donde se puede observar que la curva del Modelo von Bertalanffy no describe muy bien los datos, sin embargo, las curvas de los ajustes para los Modelos Logístico y Gompertz son muy similares, y por tanto es necesario utilizar otros criterios de comparación como el criterio de Información de Watanabe-Akaike (WAIC). En el Cuadro 4.36 se muestran los resultados de esta comparación donde se puede observar que el modelo que presenta menor WAIC, y por tanto se ajusta mejor a los datos es el Modelo Logístico.



(a) Modelos ajustados a las tres poblaciones (Células Sanas, Células Infectadas y Carga Viral). (b) Modelos ajustados a los datos de carga viral.

Figura 4.21: Comparación de Modelos Etapa 3.

Cuadro 4.36: Criterios de Comparación de Modelos.

Modelo	WAIC
von Bertalanffy	99.9
Logístico	87.6
Gompertz	91.1

#### 4.4. Número reproductivo básico y vida media de células infectadas.

A partir de las muestras MCMC de los parámetros de los modelos se obtuvieron las estimaciones puntuales e intervalos creíbles de algunos parámetros relevantes para los Modelos (3.4), (3.5), (3.6), (3.7), (3.8) y (3.9), como lo son el número reproductivo básico ( $R_0$ ) y la vida media de las células infectadas  $((\ln 2)/\delta)$ .

Las estimaciones para el  $R_0$  son mayores a 2 para todos los modelos, sin embargo, para los modelos con tratamiento de interferón son mas pequeñas que para los modelos con ausencia de interferón en todos los casos. Gráficamente se puede apreciar que los intervalos creíbles no se traslapan entre sí y por ende son estadísticamente diferentes. Es

decir que el tratamiento con interferón reduce significativamente el número reproductivo básico (Figura 4.22).

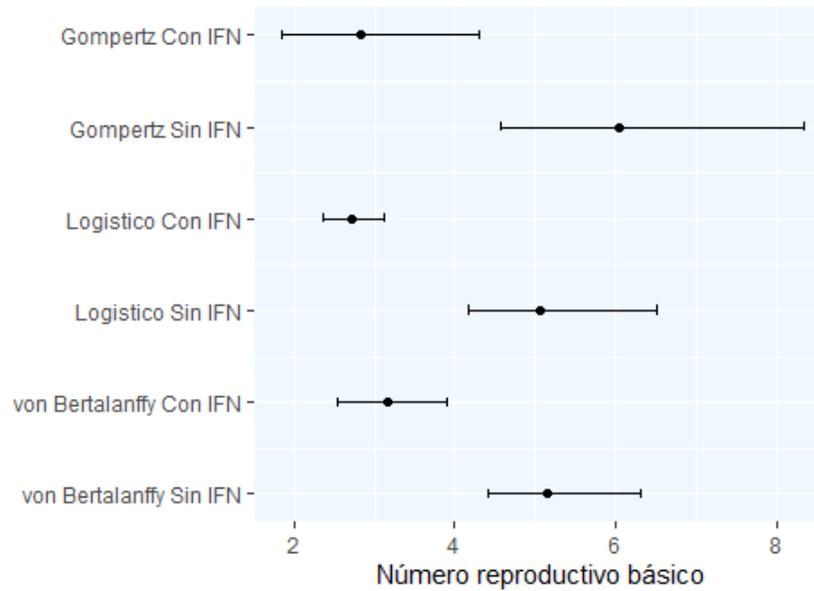


Figura 4.22: Estimaciones puntuales e intervalo de alta densidad del número reproductivo básico.

La estimación puntual de la vida media de las células infectadas es muy similar en todos los modelos y se encuentra alrededor de 0.45 días para todos los modelos, excepto para el modelo Gompertz con tratamiento de interferón (Modelo 3.9), que es ligeramente superior a 0.5 días (Figura 4.23). Gráficamente se aprecia que los intervalos creíbles se traslapan entre sí, lo cual quiere decir que el interferón no tiene efecto significativo en el tiempo de vida de las células infectadas, esto es debido a que la tasa de mortalidad de las células infectadas se modela de igual manera en todos los modelos.

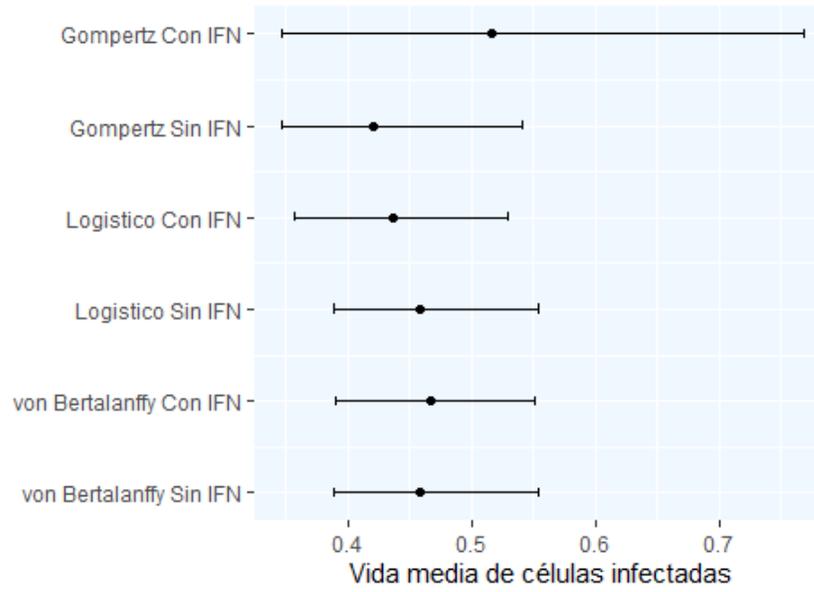


Figura 4.23: Estimaciones puntuales e intervalo de alta densidad de la vida media de las células infectadas.

### Conclusiones

---

La estadística bayesiana además de ser un enfoque para la estimación de parámetros, permite obtener distribuciones de probabilidad posteriores, no sólo de los parámetros sino también de medidas relevantes en este tipo de modelos de dinámica viral, como lo son el número reproductivo básico, la vida media de las células infectadas, entre otras. Además, permite incluir conocimiento empírico sobre los parámetros a partir de las distribuciones a priori.

Tanto las estimaciones de los parámetros de dinámica celular: tasa de crecimiento y capacidad de carga, como las estimaciones de los parámetros de dinámica viral: tasa de infección, tasa de mortalidad de las células infectadas y la tasa de producción de virus de los correspondientes modelos ajustados son muy similares, esto indica consistencia en dichas estimaciones.

De acuerdo al criterio de información de Watanabe-Akaike, el modelo que ajustó mejor a los datos de crecimiento celular usados en la primera etapa fue el modelo con función de crecimiento de Gompertz y el que menos ajustó fue el modelo con función de crecimiento Logístico. Para los modelos ajustados sin y con tratamiento de interferón (Etapas 2 y 3, respectivamente), el modelo con función de crecimiento Logística fue el que describió mejor los datos de carga viral y el que menos se ajustó fue el modelo con función

de crecimiento de von Bertalanffy.

El efecto de inhibición estimado del interferón sobre la infección celular en el modelo seleccionado, es decir, en el modelo con función de crecimiento de Logística fue aproximadamente del 9.9% y con una probabilidad de 0.95 este efecto se encuentra entre el 0.4% y el 26.5%. El efecto de inhibición estimado del interferón sobre la producción de partículas virales en el modelo seleccionado (modelo con función de crecimiento de Logística) fue aproximadamente del 8.1% y con una probabilidad de 0.95 se encuentra entre el 0.3% y el 23.1%. Lo anterior indica que estos efectos de inhibición son similares en la prevención de nuevas infecciones como en la replicación viral.

El número reproductivo básico estimado para los modelos con tratamiento de interferón fue mas pequeño que para los modelos sin tratamiento, se puede concluir que el tratamiento con interferón reduce significativamente el número reproductivo básico y por ende la propagación de la infección. Mientras que para el tiempo de vida medio de las células infectadas, el tratamiento con interferón no tiene efecto significativo.

---

## Referencias

---

- [1] Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature*. 1995;373(6510):123–126.
- [2] Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science*. 1996;271(5255):1582–1586.
- [3] Nowak MA, Bangham CR. Population dynamics of immune responses to persistent viruses. *Science*. 1996;272(5258):74–79.
- [4] Bonhoeffer S, May RM, Shaw GM, Nowak MA. Virus dynamics and drug therapy. *Proceedings of the National Academy of Sciences*. 1997;94(13):6971–6976.
- [5] Wodarz D, Nowak MA. Mathematical models of HIV pathogenesis and treatment. *BioEssays*. 2002;24(12):1178–1187.
- [6] Callaway DS, Perelson AS. HIV-1 infection and low steady state viral loads. *Bulletin of mathematical biology*. 2002;64(1):29–64.
- [7] Perelson AS. Modelling viral and immune system dynamics. *Nature Reviews Immunology*. 2002;2(1):28–36.
- [8] Neumann AU, Lam NP, Dahari H, Gretch DR, Wiley TE, Layden TJ, et al. Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon- $\alpha$  therapy. *Science*. 1998;282(5386):103–107.
- [9] Dahari H, Lo A, Ribeiro RM, Perelson AS. Modeling hepatitis C virus dynamics: liver regeneration and critical drug efficacy. *Journal of theoretical biology*. 2007;247(2):371–381.
- [10] Baccam P, Beauchemin C, Macken CA, Hayden FG, Perelson AS. Kinetics of influenza A virus infection in humans. *Journal of virology*. 2006;80(15):7590–7599.

- [11] Beauchemin CA, McSharry JJ, Drusano GL, Nguyen JT, Went GT, Ribeiro RM, et al. Modeling amantadine treatment of influenza A virus in vitro. *Journal of theoretical biology.* 2008;254(2):439–451.
- [12] Petracic J, Wilson DP. Simulating the entire natural course of HIV infection by extending the basic viral dynamics equations to include declining viral clearance. *Pathogens and Disease.* 2019;77(4):ftz043.
- [13] Zitzmann C, Kaderali L. Mathematical analysis of viral replication dynamics and antiviral treatment strategies: from basic models to age-based multi-scale modeling. *Frontiers in microbiology.* 2018;9:1546.
- [14] Gilks WR, Richardson S, Spiegelhalter DJ. Introducing markov chain monte carlo. *Markov chain Monte Carlo in practice.* 1996;1:19.
- [15] Jiménez Luna J. Métodos Monte Carlo basados en cadenas de Markov. A. 2015;.
- [16] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian data analysis.* CRC press; 2013.
- [17] Robert CP, Casella G, Casella G. *Introducing monte carlo methods with r.* vol. 18. Springer; 2010.
- [18] Rizzo ML. *Statistical computing with R.* CRC Press; 2007.
- [19] Casella G, Berger RL. *Statistical Inference.* Brooks. Col Publish Co. 1990;.
- [20] Fox JP. *Bayesian item response modeling: Theory and applications.* Springer Science & Business Media; 2010.
- [21] Chapra SC, Canale RP. *Numerical methods for engineers.* vol. 2. Mcgraw-hill New York; 2011.
- [22] Ikeda H, Godinho-Santos A, Rato S, Vanwalscappel B, Clavel F, Aihara K, et al. Quantifying the antiviral effect of IFN on HIV-1 replication in cell culture. *Scientific reports.* 2015;5(1):1–8.
- [23] Bürkner PC. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software.* 2017;80(1):1–28.
- [24] Team RC. *R: A Language and Environment for Statistical Computing.* Vienna, Austria; 2020. Available from: <https://www.R-project.org/>.

- [25] Ge H, Xu K, Ghahramani Z. Turing: a language for flexible probabilistic inference. In: International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain; 2018. p. 1682–1690. Available from: <http://proceedings.mlr.press/v84/ge18b.html>.
- [26] Bezanson J, Edelman A, Karpinski S, Shah VB. Julia: A fresh approach to numerical computing. *SIAM review*. 2017;59(1):65–98. Available from: <https://doi.org/10.1137/141000671>.
- [27] Plummer M, Best N, Cowles K, Vines K. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*. 2006;6(1):7–11. Available from: <https://journal.r-project.org/archive/>.