

INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO



CAZANDO INCERTIDUMBRE:
UN MODELO PRESA-DEPREDADOR BAYESIANO
PARA LA SUPERVIVENCIA DEL MÁS APTO

POR

DAVID DAMIÁN ARBEU
BRANDON HERNÁNDEZ TRONCOSO

CDMX.

2023

Índice

1. Introducción	4
2. Descriptivo de modelo	5
2.1. Enfoque de sistemas dinámicos	5
2.2. El papel de la estadística bayesiana	6
3. Análisis exploratorio	9
4. Resultados	11
4.1. Chequeos predictivos previos.	11
4.2. Diagnósticos MCMC.	15
4.3. Calibración basada en simulación	17
4.4. Estimación posterior	20
4.5. Contando con incertidumbre: Inferencia del tamaño de las poblaciones	21
4.6. Número efectivo de parámetros	26
5. Conclusiones	27
5.1. Trabajo futuro	28
6. Apéndice	29
6.1. Definiciones y resultados.	29
6.2. Análisis dinámico	30
7. Bibliografía	32

Índice de figuras

1.	Liebres (presas) y lince canadienses (depredadores). Tomada de Carpenter (2018)	4
2.	Número de pieles de lince y liebre (en miles) recolectadas entre 1900 y 1920.	9
3.	Diferencia absoluta del número de pieles de lince y liebre (en miles) recolectadas entre 1900 y 1920.	10
4.	Número de pieles de lince contra número de pieles de liebre recolectadas de 1900 a 1920.	10
5.	Flujo de trabajo bayesiano. Tomada de Garbuno-Iñigo (2023b)	11
6.	Diagnósticos gráficos MCMC	16
7.	Gráfico de coordenadas paralelas para los parámetros $\alpha, \beta, \gamma, \delta$	16
8.	<i>Rank plots</i> de cada cadena para cada parámetro.	17
9.	En rojo: valor real de parámetros $\alpha, \beta, \gamma, \delta, \sigma_1, \sigma_2$ bajo la previa; en verde: predicción puntual posterior y línea negra: intervalos al 95 %.	18
10.	En rojo: valor real de condiciones iniciales $u(0), v(0)$ bajo la previa; en verde: predicción puntual posterior y línea negra: intervalos al 95 %.	19
11.	Datos replicados de la predictiva previa.	19
12.	Predicción puntual e intervalos de confianza al 95 % para los parámetros del sistema dinámico	20
13.	Chequeo predictivo posterior con datos replicados para el número de pieles y tamaño poblacional.	22
14.	Prueba <i>lineup</i> de datos replicados para la evolución de las pieles recolectadas entre 1900 y 1920.	23
15.	Dinámica del número de pieles (en miles) a partir del <i>solver</i> de Stan considerando 100 muestras posteriores de θ	24
16.	Número de pieles de lince y liebre (en miles) esperadas, considerando 100 muestras de los parámetros obtenidos de la distribución posterior.	24
17.	Predicción media de liebre y lince 50 años atrás y adelante a partir de la predictiva posterior.	25
18.	Espacio fase para diferentes soluciones del modelo Lotka-Volterra.	30

Índice de cuadros

1.	Cantidad de pieles recogidas por la Compañía de la Bahía de Hudson en los años 1900-1920	9
2.	Diagnóstico de las muestras generadas	15
3.	Parámetros simulados de la previa	17
4.	Datos simulados de la predictiva posterior	18
5.	Estimación posterior usando datos falsos.	18
6.	Resumen global de las muestras generadas	20
7.	Métricas WAIC para liebres (izquierda) y lince (derecha)	26

1. Introducción

Life's most important problems are, for the most part, problems of probability. (Laplace, Pierre-Simon; 1812)

En 1920 Lotka publicó un artículo titulado “Apunte analítico sobre ciertas relaciones en los sistemas orgánicos”. En él, propuso un **sistema dinámico**¹ que describe el tamaño poblacional de dos especies. El ejemplo que consideró fue el de una población de herbívoros (**depredadores**) cuyo principal recurso de sobrevivencia son plantas (**presas**) del mismo hábitat y cuyas poblaciones crecen, en cada unidad de tiempo, a una velocidad:

$$\begin{aligned}\frac{d}{dt}u &= u(\alpha - \beta v) \\ \frac{d}{dt}v &= v(-\gamma + \delta u)\end{aligned}\tag{1}$$

en donde $u(t)$ y $v(t)$ denotan el tamaño de la población de hervíboros y plantas al tiempo $t \geq 0$ respectivamente. Los parámetros α , β , γ y δ son estimados a menudo a partir de datos de poblaciones *presa-depredador* durante un período determinado.

En este trabajo, abordamos un **problema inverso**: Si el tamaño poblacional es conocido en un periodo de tiempo ¿cuáles son los parámetros que dieron lugar a tal comportamiento en el tamaño de las poblaciones?. Para ello se realiza una estimación del vector de parámetros $\theta = (\alpha, \beta, \gamma, \delta)$ desde una perspectiva de Estadística Bayesiana usando **Stan**.

La estimación de los parámetros es fundamental para comprender la dinámica de algún par de poblaciones específico así como el comportamiento por su interacción mutua. Profundizamos sobre la interpretación de los mismos en la siguiente sección.

Para conseguir nuestro objetivo nos basamos en el artículo de Carpenter (2018) disponible en el sitio web de **Stan**. En particular, utilizamos los datos de la cantidad de pieles de línices canadienses y liebres recogidas por la Compañía de la Bahía de Hudson en los años 1900-1920 (Howard, P. 2009), que son proporcionados en este mismo artículo. Estos datos son relevantes para nuestro estudio ya que nos permiten analizar la relación entre la población de depredadores y presas en un ecosistema, y validar nuestro modelo teórico mediante la comparación con resultados empíricos.



Figura 1: Liebres (presas) y línices canadienses (depredadores). Tomada de Carpenter (2018)

En nuestro trabajo, hemos mejorado y agregado originalidad al flujo de trabajo bayesiano presentado en el artículo base. Esto se logró mediante la inclusión de etapas adicionales, tales como chequeos previos, la calibración basada en simulación, además de *forecasting* y *backcasting*, y un análisis detallado del sistema dinámico definido en la ecuación (1).

Gracias a estas adiciones, hemos logrado obtener una comprensión más profunda y completa del comportamiento del sistema, lo cual no fue abordado en el artículo base de Carpenter (2018). Esperamos que el lector también pueda beneficiarse de estas mejoras y comprender mejor el sistema que hemos analizado.

¹véase definición 1

2. Descriptivo de modelo

En la presente sección, se describe con mayor detalle la especificación del modelo y se provee al lector de algunas referencias bibliográficas relevantes para consultar información adicional.

2.1. Enfoque de sistemas dinámicos

Volterra, V. (1926) propuso el mismo modelo que Lotka, pero con un enfoque diferente. Mientras que Lotka se enfocó en el estudio de la dinámica de las poblaciones animales, Volterra abordó la pregunta de si sería posible explicar las fluctuaciones que se habían observado en la población de peces del mar Adriático y que preocupaban mucho a los pescadores en épocas de baja población de peces.

Modelo 1(Lotka-Volterra)

$$\begin{aligned}\frac{d}{dt}u &= u(\alpha - \beta v) \\ \frac{d}{dt}v &= v(-\gamma + \delta u)\end{aligned}\tag{2}$$

2.1.1. Supuestos

El modelo de Lotka Volterra asume los siguientes puntos, según lo indicado por Brauer and Castillo-Chavez, C. (2012) p.123

Supuestos del modelo LV

- La fuente principal de alimento de las presas es ilimitada y por lo tanto la tasa de crecimiento per capita de la población de presas, en ausencia de depredadores, es constante.
- La principal fuente de alimentación de los depredadores son las presas y en un escenario de ausencia de presas, la tasa de mortalidad de los depredadores es constante.
- En presencia de presas, la tasa de crecimiento de los depredadores incrementa de manera lineal en relación a la población de presas.
- En presencia de depredadores, la población de presas disminuye de manera lineal en relación a la población de depredadores.

2.1.2. Interpretación

Es importante hacer notar que, bajo los supuestos del modelo LV, se puede dar la siguiente interpretación a los parámetros de (2).

Notación	Significado
$\alpha \geq 0$	Tasa de crecimiento de las presas en ausencia de depredadores.
$\gamma \geq 0$	Tasa de mortalidad de los depredadores en ausencia de presas.
$\beta v > 0$	Tasa de mortalidad de la población de presas debida a la interacción con depredadores.
$\delta u > 0$	Tasa de crecimiento de la población de depredadores debida a la interacción presas.

2.2. El papel de la estadística bayesiana

El modelo presentado en la ecuación (2) es **determinista** lo que significa que, si los parámetros y los tamaños poblacionales fuesen perfectamente conocidos, sería posible predecir exactamente todos los estados futuros. Para lograr hacer predicciones se hará inferencia del valor de los parámetros basados en los datos observados.

Después de lo anterior, el lector curioso podría preguntarse: **¿En qué punto de la solución están involucradas las soluciones del modelo LV?** Asumiremos que el número de pieles recolectado está asociado con el tamaño de las poblaciones de presas y depredadores. Como primer propuesta asumimos que, en promedio, el tamaño de las poblaciones en cada momento del tiempo y el número de pieles recolectado para cada especie son iguales y que hay posible variación por error de medición y simplificación del modelo.

De manera simple y adoptando un esquema similar al de regresión lineal simple, donde el predictor lineal se sustituye por las soluciones del sistema LV, la descripción anterior se traduce en contemplar el modelo

$$\begin{aligned}\text{número de pieles observado en presas} &= \text{número de presas} + \epsilon_1 \\ \text{número de pieles observado en depredadores} &= \text{número de depredadores} + \epsilon_2\end{aligned}$$

con una ligera modificación: dado que las cantidades involucradas (número de pieles y tamaño de las poblaciones) son positivas, tenemos que restringir de alguna manera a que los términos de error sean no negativos. La manera más simple de hacerlo es considerar al número de pieles y tamaño de las poblaciones en escala logarítmica y en la siguiente sección damos más detalle sobre esto.

2.2.1. Especificación del modelo bayesiano.

Con el modelo bayesiano que utilizamos, podemos cuantificar la incertidumbre en los parámetros mediante la asignación de una distribución previa. Una de las ventajas de este enfoque respecto de otros (como métodos de optimización), es que permite incorporar conocimiento previo acerca de estos y actualizarlo con los datos observados.

En esta sección, detallaremos la especificación del modelo estadístico que utilizamos y explicaremos cómo asignamos la distribución previa a los parámetros. Esto nos permitirá avanzar en nuestra tarea de modelado con una comprensión profunda del enfoque bayesiano que estamos utilizando y cómo esto nos permite cuantificar la incertidumbre en los resultados de nuestro modelo.

2.2.2. Verosimilitud

Ahora, se especificará un modelo probabilístico que describa el proceso que generó los datos. En él, relacionaremos el número de pieles observado y el tamaño poblacional de estas dos especies. Para ello, tomaremos en cuenta que los datos del número de pieles son una aproximación del tamaño de las poblaciones.

Para cada momento del tiempo $n \geq 1$, sea $y_n = (u_n, v_n)$ en donde

- u_n denota el número de pieles observado para las presas.
- v_n denota el número de pieles observado para los depredadores.

y para cada $t = 1, \dots, M$, sean

$$\mathbf{z}_t = \begin{bmatrix} u(t) \\ v(t) \end{bmatrix} \quad (3)$$

las soluciones del modelo LV al tiempo t , dados los parámetros $\theta = (\alpha, \beta, \gamma, \delta)$ y la condición inicial $\mathbf{x}_{init} = [u(t=0), v(t=0)]$.

Nota

Cada \mathbf{z}_j es aleatoria pues es función de los parámetros \mathbf{x}_{init} y $\theta = (\alpha, \beta, \gamma, \delta)$.

Dado que las variables que incluiremos en el modelo son positivas, pues son tamaños de poblaciones, lo que haremos es relacionar $\log(\text{cantidad de pieles}) > 0$ y $\log(\text{tamaño de población}) > 0$. Esta transformación tiene la ventaja de poder emplear el enfoque de regresión lineal normal clásico y por lo tanto los errores no están restringidos a ser positivos.

Más precisamente,

Modelo 2

$$\begin{aligned}\log(u_n) &= \log(u(n; \mathbf{x}_{init}, \theta)) + \epsilon_1 \\ \log(v_n) &= \log(v(n; \mathbf{x}_{init}, \theta)) + \epsilon_2\end{aligned}\tag{4}$$

con $\epsilon_i \sim \mathbf{N}(0, \sigma_i)$, $i = 1, 2$ o bien

$$\begin{aligned}\log(u_n) \mid \mathbf{x}_{init}, \theta &\sim \mathbf{N}(\log(x(n)), \sigma_1) \\ \log(v_n) \mid \mathbf{x}_{init}, \theta &\sim \mathbf{N}(\log(y(n)), \sigma_2)\end{aligned}$$

En donde los términos de error representan:

- Falta de certeza en que las mediciones se generaron cumpliendo los supuestos del modelo LV, . Así que, aún si conocieramos el verdadero valor de los parámetros que dieron lugar a éstas y usáramos el modelo LV para calcular los tamaños poblacionales en cada momento del tiempo, entonces los resultados no coincidirían exactamente con los datos observados.
- Existen variables como el clima, enfermedades infecciosas y otros factores de largo plazo que provocan variación en el tamaño de las poblaciones y que no incluiremos en nuestro proceso de modelado (Carpenter 2018).

Por otro lado, otra forma de expresar el modelo (4) es la siguiente.

$$\begin{aligned}u_n &= \exp(\log(u(n; \mathbf{x}_{init}, \theta)) + \epsilon_1) \\ &= u(n; \mathbf{x}_{init}, \theta) \exp(\epsilon_1)\end{aligned}\tag{5}$$

y como

$$\log(u(n; \mathbf{x}_{init}, \theta)) + \epsilon_1 := Z \sim \mathbf{N}(\log(u(n; \mathbf{x}_{init}, \theta)), \sigma_1)$$

entonces $U_n \sim \mathbf{Log-normal}(0, \sigma_1)$ y concluiríamos similarmente para V_n . Eso nos lleva a expresar el modelo en (4) como sigue:

Modelo 3

$$\begin{aligned}u_n \mid \mathbf{x}_{init}, \theta &\sim \mathbf{Log-normal}(\log(u(n)), \sigma_1) \\ v_n \mid \mathbf{x}_{init}, \theta &\sim \mathbf{Log-normal}(\log(v(n)), \sigma_2)\end{aligned}$$

Nota

La intención de expresar de otra manera nuestro modelo es para notar que, en promedio, el número de pieles recolectadas es proporcional al tamaño de la población, con un factor de corrección que tiene en cuenta algún porcentaje del total.

En efecto, sin pérdida de generalidad considere únicamente la relación descrita para las presas según la ecuación (5), entonces

$$\begin{aligned}\frac{u_n}{u(n; \mathbf{x}_{init}, \theta)} &= \mathbb{E}[\exp(\epsilon_1) \mid u_n, u(n; \mathbf{x}_{init}, \theta)] \\ &= \exp(0 + \sigma_1^2/2)\end{aligned}$$

siempre que $u(n; \mathbf{x}_{init}, \theta) \neq 0$ y donde para la última igualdad hemos hecho uso del Teorema 1 del Apéndice.

2.2.3. Previas

El último paso para completar el modelo bayesiano corresponde a la definición de nuestro conocimiento previo mediante distribuciones, las cuales nos permitan delimitar las regiones de valores que puede tomar cada uno de los parámetros en cuestión. De esta manera, procedemos a incluir nuestro conocimiento previo para los parámetros $\theta = (\alpha, \beta, \gamma, \delta)$ y \mathbf{x}_{init} , considerando el significado de cada parámetro y la forma en que fueron codificados los datos (en miles). Sabemos de los datos que las mediciones del número de pieles para ambas especies son del orden de 10^1 , por lo que, de acuerdo con (Carpenter 2018), estableceremos las siguientes previas:

$$\begin{aligned}\alpha, \gamma &\sim \mathbf{N}(1, 0.5) \\ \beta, \delta &\sim \mathbf{N}(0.05, 0.05)\end{aligned}$$

Si bien estas distribuciones son un poco vagas en el sentido de tener mucha varianza, nos permiten delimitar los posibles valores que pueden tomar los parámetros considerando la escala de los datos y el contexto del problema. Posteriormente en la subsección **Chequeos predictivos previos** evaluaremos el efecto de nuestros supuestos distribucionales para los parámetros en cantidades observables.

3. Análisis exploratorio

Como en toda tarea de inferencia, los datos son uno de los ingredientes principales. En esta ocasión, se cuenta con dos series de tiempo, cada una compuesta por la cantidad de pieles de lince o de liebre (en miles) que fueron recolectadas entre 1900-1920 por la compañía Bahía de Hudson.

Es importante aclarar que el número de pieles fue redondeado al número más cercano de 100, lo cual implica un margen de error de medición del orden del 1 %.

Cuadro 1: Cantidad de pieles recogidas por la Compañía de la Bahía de Hudson en los años 1900-1920

Año	Lince	liebre	Año	Lince	liebre	Año	Lince	liebre	Año	Lince	liebre
1900	4.0	30.0	1905	41.7	20.6	1910	7.4	27.1	1916	29.7	11.2
1901	6.1	47.2	1906	19.0	18.1	1911	8.0	40.3	1917	15.8	7.6
1902	9.8	70.2	1907	13.0	21.4	1912	12.3	57.0	1918	9.7	14.6
1903	35.2	77.4	1908	8.3	22.0	1913	19.5	76.6	1919	10.1	16.2
1904	59.4	36.3	1909	9.1	25.4	1914	45.7	52.3	1920	8.6	24.7

La siguiente gráfica facilita el entendimiento de de ambas series de tiempo.

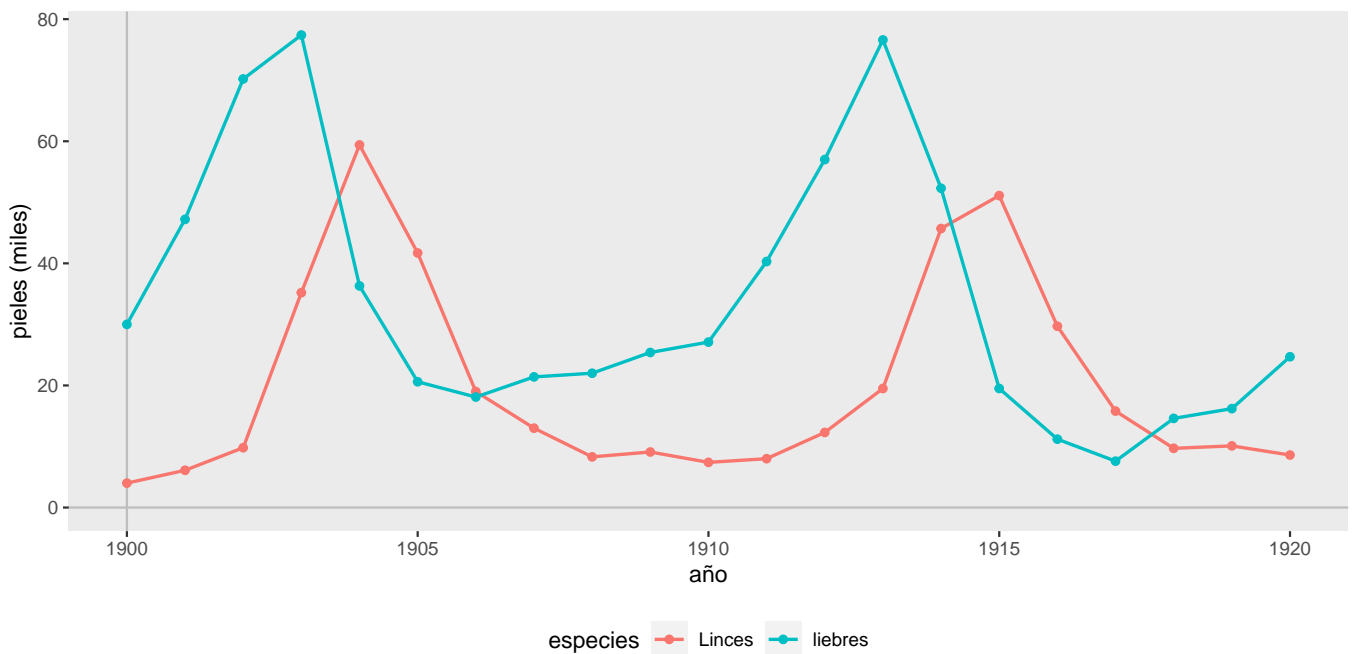


Figura 2: Número de pieles de lince y liebre (en miles) recolectadas entre 1900 y 1920.

Podemos observar un patrón de oscilación para ambas poblaciones. En particular, se alcanza un punto máximo en la cantidad de pieles recolectadas, para luego descender hasta un punto mínimo y nuevamente, incrementar hasta alcanzar otro punto máximo; hecho que es consistente con los supuestos de nuestro modelo.

Notemos que no hay un dominio de la cantidad de pieles recolectadas para alguna especie en particular, es decir que para ciertos momentos del tiempo la cantidad de pieles de lince fue mayor que la cantidad de pieles de liebre, y que en periodos sucesivos del tiempo este comportamiento se invierte.

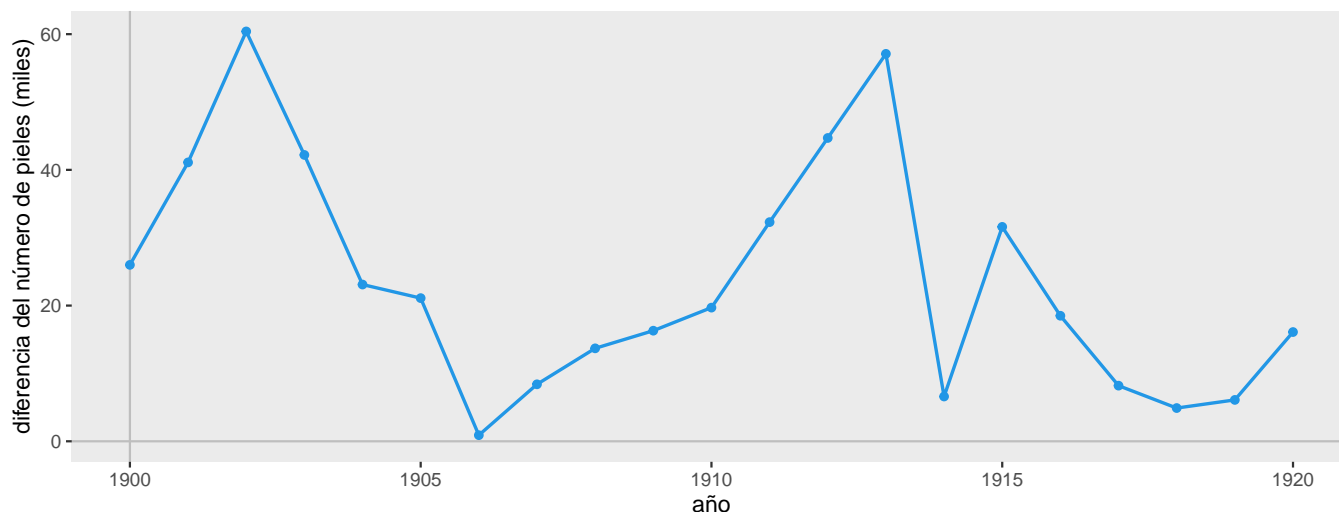


Figura 3: Diferencia absoluta del número de pieles de linces y liebres (en miles) recolectadas entre 1900 y 1920.

Por otro lado, la diferencia absoluta del número de pieles recolectadas entre ambas especies está en el rango de 0 a 60. Esta observación es importante pues hace notar la consistencia en el espacio de valores que puede tomar el número de pieles observadas para cada especie y consideramos que es un estadístico útil para evaluar el comportamiento de los datos.

Por último, podemos graficar la cantidad de pieles de linces contra la cantidad de pieles de liebres con la intención de observar algún patrón en la información, además de entender la interacción que hay ambas especies a través de los datos.

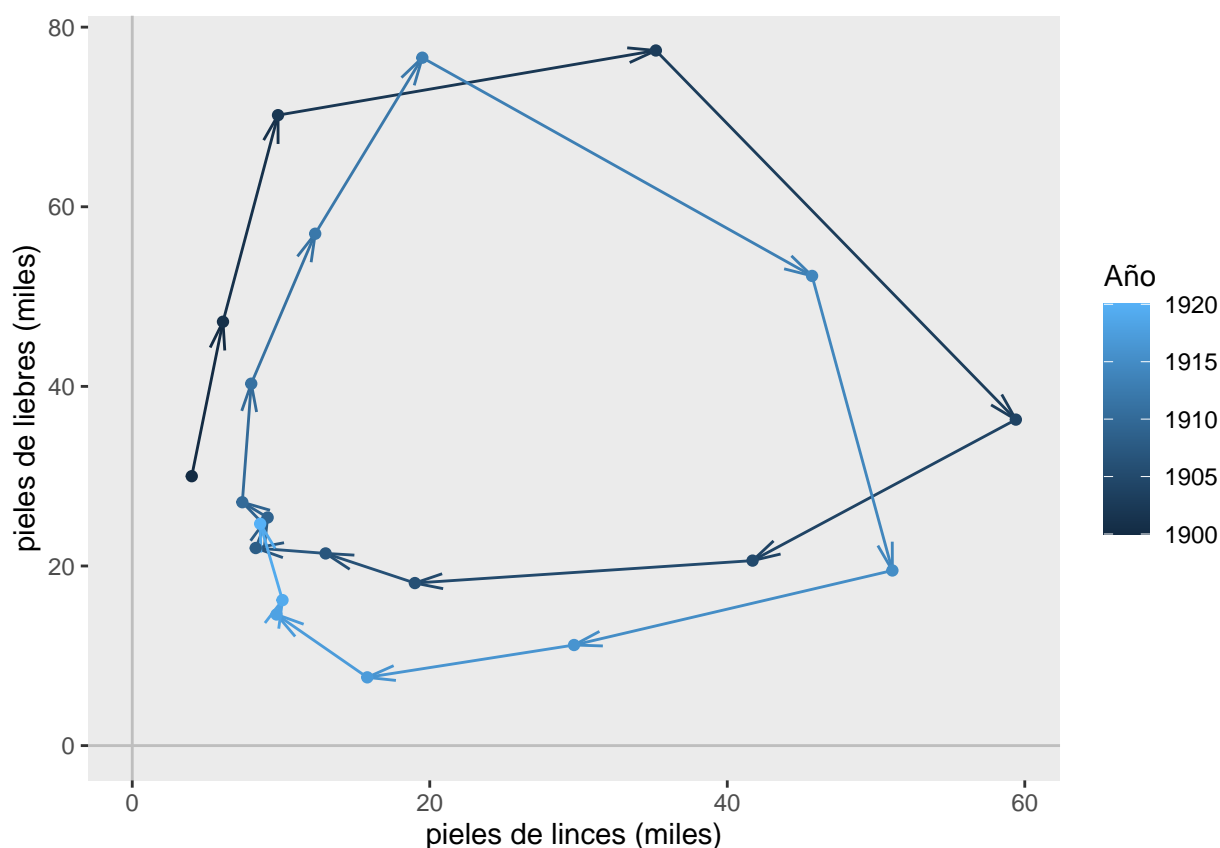


Figura 4: Número de pieles de linces contra número de pieles de liebres recolectadas de 1900 a 1920.

Como era de esperarse, en la gráfica anterior se observa un patrón de oscilación similar al que se pudo observar al inicio de esta sección, con la diferencia de que en esta ocasión existe una clara estabilidad. Dicho de otra manera, la gráfica anterior nos muestra que el número de pieles recolectadas tendrá un comportamiento similar a lo largo del tiempo, lo que facilita el modelado de este problema mediante un enfoque dinámico.

4. Resultados

En esta sección mostramos los resultados de la implementación del modelo bayesiano en **Stan**. A su vez, como parte de un flujo de trabajo bayesiano estudiamos el modelo generativo propuesto en la subsección Especificación del modelo bayesiano. antes y después de ajustar nuestro modelo para evaluar y criticar los supuestos distribucionales tanto para el modelo de verosimilitud como las previas de los parámetros.

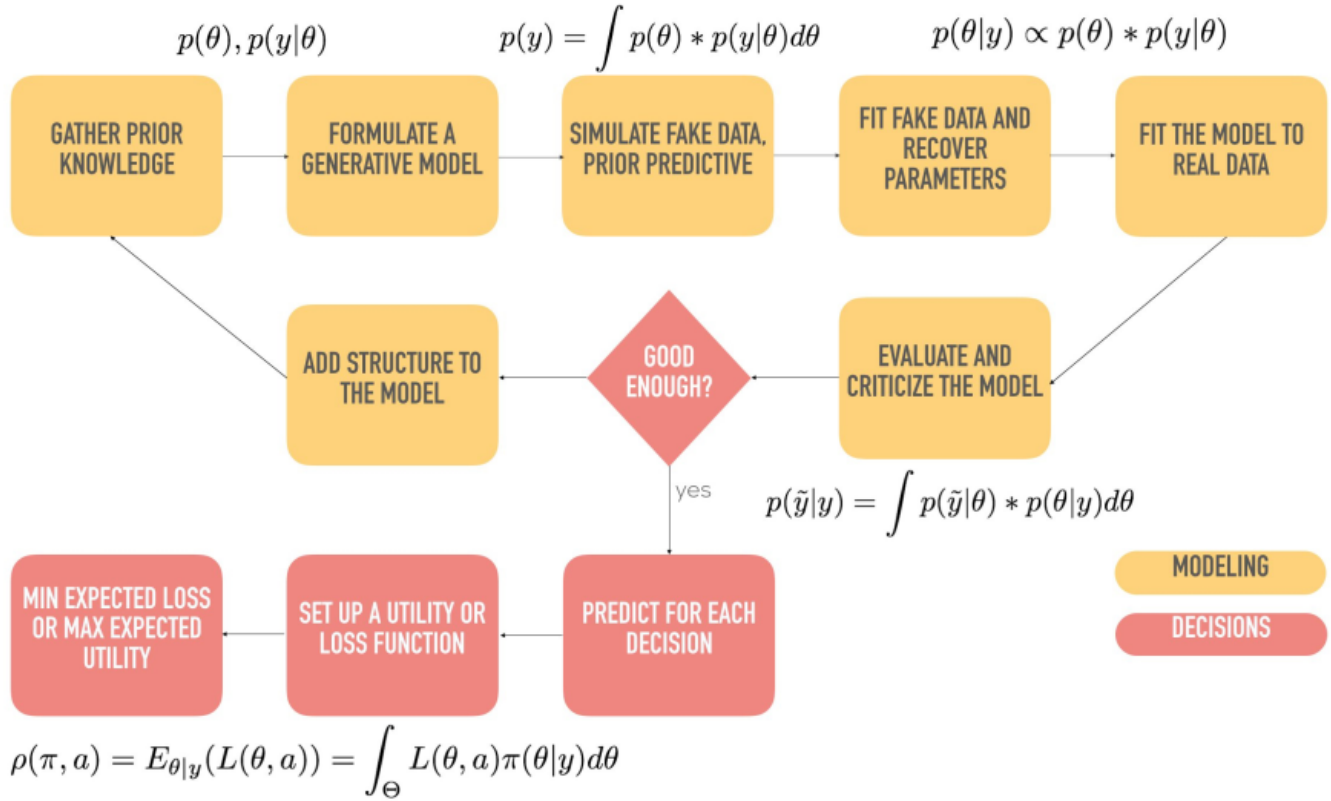


Figura 5: Flujo de trabajo bayesiano. Tomada de Garbuno-Iñigo (2023b)

4.1. Chequeos predictivos previos.

La intención de esta etapa en el flujo de trabajo bayesiano es verificar la consistencia entre la verosimilitud y la distribución previa en los parámetros. La distribución predictiva previa es una herramienta útil en este proceso ya que permite evaluar el efecto de nuestros supuestos distribucionales en cantidades observables.

El objetivo es asegurarnos que nuestra distribución inicial no asigna regiones de alta probabilidad a valores que no tienen sentido en el contexto del problema que estamos modelando. (Garbuno-Iñigo 2023b)

$$\pi(\tilde{y}) = \int \pi(\tilde{y} | \theta) \pi(\theta) d\theta$$

Esta distribución es previa en el sentido de que no se basa en ninguna observación y , sin embargo, tampoco se pretende replicar los datos a partir de ésta distribución. Tomando lo anterior en cuenta, resulta pertinente realizar una evaluación sobre nuestros supuestos respecto al problema, los cuales fueron planteados mediante las distribuciones previas descritas en la subsección **Previas**.

Recordando brevemente las distribuciones previas propuestas:

$$\begin{aligned} \alpha, \gamma &\sim \mathbf{N}(1, 0.5) \\ \beta, \delta &\sim \mathbf{N}(0.05, 0.05) \end{aligned}$$

estas distribuciones implican que, con 95 % de probabilidad

- La tasa de crecimiento de presas en ausencia de depredadores, así como la tasa de mortalidad de depredadores en ausencia de presas, están entre los valores de 0 y 2.²
- La tasa de mortalidad de presas por interacción depredadores, así como la tasa de crecimiento de depredadores por interacción con presas, están entre los valores -0.05 y 0.15.

En el caso de los parámetros asociados a la interacción entre presas y depredadores, podemos observar que el intervalo contiene valores tanto positivos como negativos, lo que nos permite entender que la interacción pudiera ser tanto positiva como negativa a la hora de estimar el número de pieles para ambas especies, favoreciendo el comportamiento oscilatorio del número de pieles a lo largo del tiempo. Por otro lado, podemos tomar en cuenta el rango del número de pieles para ambas especies con el fin de considerar los valores que pudieran tomar las escalas σ_i asociadas a los errores de medición. En este caso, sabemos que el rango del número de pieles se encuentra entre 4 y 77³, de modo que esperaríamos que las escalas fueran consistentes con dicho rango. Por lo tanto, podemos definir que los valores que puedan tomar las escalas sean valores entre 0.05 y 3, lo que nos lleva a considerar

$$\sigma_i \sim \text{Log-normal}(-1, 1) \quad (6)$$

que es consistente con el hecho de que $\sigma > 0$ y que, considerando la escala logarítmica de los datos, $\exp(\sigma_i) \in [e^{-1-2}, e^{-1+2}]$.

Finalmente, para la población inicial (medida en miles)

$$u(t=0), v(t=0) \sim \text{Log-normal}(\log(10), 1)$$

Nota

Una justificación del por qué establecer dichas previas es el siguiente: el modelo de LV está bien estudiado, así como el comportamiento entre lince y liebres, de modo que se tiene una noción sobre el rango en el cual pueden estar los valores del tamaño de las poblaciones. Todas las previas asignadas nos permiten acotar los posibles valores que pudieran tomar los parámetros del modelo y ser consistentes con la información disponible. Así mismo, estas iniciales son de gran utilidad para establecer la escala de la respuesta (número de pieles en escala logarítmica). Por último, la definición de estas previas nos permiten agilizar el cómputo en **Stan**, de modo que no se tengan que explorar grandes regiones para los posibles de valores que puedan tomar los parámetros. (Carpenter, 2018.)

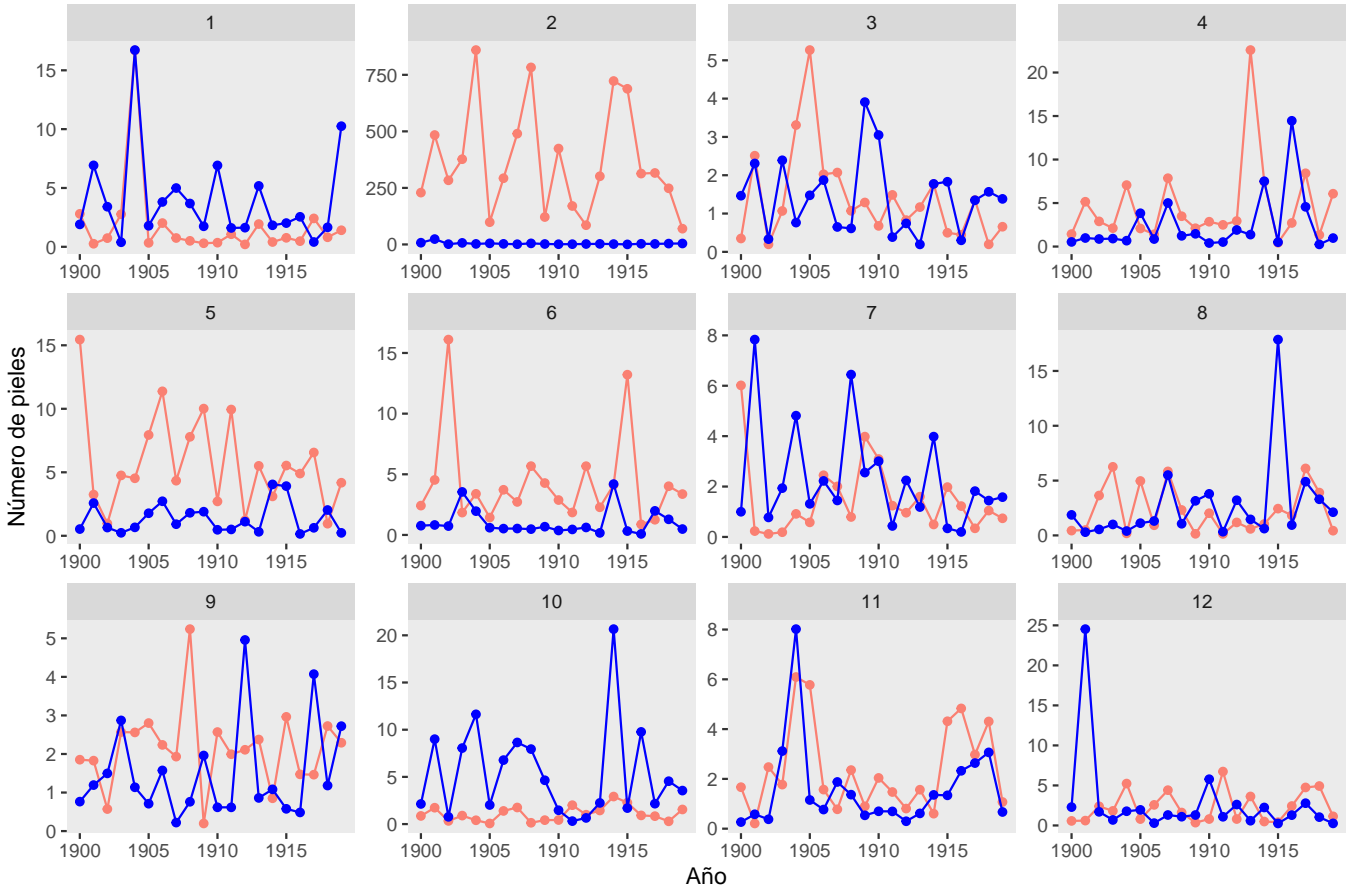
Nota

¿Cómo es que **Stan** está calculando las soluciones del sistema LV? Para cada muestra de los parámetros del modelo LV $\theta^{(j)} = (\alpha^{(j)}, \beta^{(j)}, \gamma^{(j)}, \delta^{(j)})$ y condiciones iniciales $u^{(j)}(0), v^{(j)}(0)$, **Stan** resuelve el sistema de ecuaciones diferenciales (2). En particular, se resuelven tantos sistemas de ecuaciones diferenciales como muestras, de la previa o posterior, haya.

Explotaremos esta observación para la generación de datos replicados bajo la previa y, posteriormente, bajo la posterior.

²un intervalo del 95 % para una cantidad $\theta \sim \mathbf{N}(\mu, \sigma)$ está dado por $\mu \pm 2\sigma$

³véase tabla 1



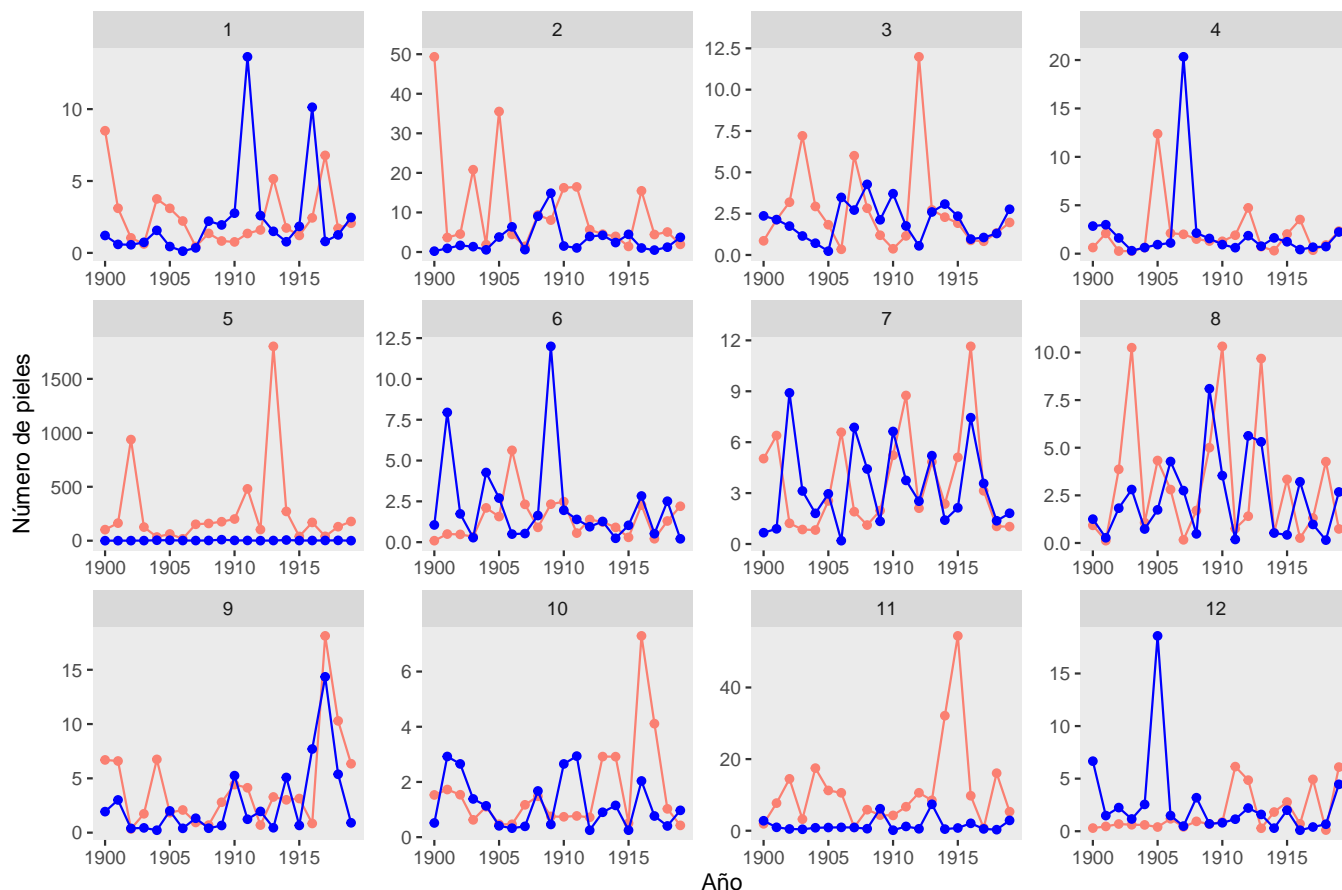
La figura anterior presenta la dinámica del número de pieles a partir de la simulación de $N = 12$ conjuntos de datos replicados con la predictiva previa. Como puede observarse, para cada conjunto de datos se logra capturar la tendencia cíclica del comportamiento de las especies, además de que se logra acotar correctamente el rango que pueden tomar los valores de las pieles recolectadas. Sin embargo, al lector le puede parecer inconsistente el uso de las distribuciones previas explicadas anteriormente, ya que son distribuciones que pueden generar valores negativos y no se ajustan a los supuestos y restricciones del problema, por lo que una solución viable es proponer el uso de distribuciones previas alternativas que consideren todos estos inconvenientes.

Se proponen las siguientes distribuciones previas alternativas

$$\begin{aligned}\alpha &\sim \text{HalfNormal}(0.5, 0.1) \\ \beta &\sim \text{HalfNormal}(0.05, 0.05) \\ \gamma &\sim \text{HalfNormal}(0.025, 0.05) \\ \delta &\sim \text{HalfNormal}(0.5, 0.1)\end{aligned}$$

Estas distribuciones consideran los siguientes supuestos: - Para α , se redujo la tasa de crecimiento de liebres en ausencia de linces con la intención de hacerlo más restrictivo. - Para β , la tasa de mortalidad de liebres por interacción entre liebres y linces se mantiene igual. - Para γ , la tasa de mortalidad de la población de linces se mantiene más restrictiva. - Para δ , la tasa de crecimiento de linces por interacción con liebres permite tomar valores más grandes, es decir que el consumo de liebres puede ser mayor en comparación con las previas anteriores. - Cada una de las distribuciones muestrea valores estrictamente positivos, cumpliendo con la restricción del modelo LV respecto a sus parámetros.

Considerando las previas alternativas, a continuación se presenta la dinámica del número de pieles a partir de la simulación de $N = 12$ conjuntos de datos replicados con la predictiva previa.



Como puede observarse, para cada conjunto de datos también se logra capturar la tendencia cíclica del comportamiento de las especies, además de que se logra acotar con mayor precisión el rango que pueden tomar los valores de las pieles recolectadas. Sin embargo, el desempeño de estas distribuciones previas no muestra una mejora significativa en la generación de nuevos datos en comparación con las distribuciones previas iniciales, por lo que no existe una razón suficiente para modificar las distribuciones previas en nuestro problema. Además, se debe tomar en cuenta que la intención de realizar chequeos predictivos previos no es replicar los datos, por lo que la elección de las distribuciones previas no debe realizarse tomando en cuenta si los resultados son semejantes a los datos observados, sino por la capacidad de ajustar el modelo. Por lo tanto, el resto del flujo bayesiano incluido en este trabajo se realizará considerando las distribuciones previas iniciales.

4.2. Diagnósticos MCMC.

Parte esencial de reportar estimadores en estadística bayesiana consiste en asegurarnos que las muestras de la distribución posterior sean de **calidad**. Esto es, que las muestras extraídas de las cadenas de Markov simuladas hayan explorado bien el soporte de la distribución posterior y por lo tanto las estimaciones de momentos de dicha distribución sean confiables. Para diagnosticar ésto, diversos autores han desarrollado estadísticos de los cuales hacemos uso. Así mismo, presentamos diagnósticos visuales.

4.2.1. Diagnósticos numéricos

Después de ajustar el modelo en **Stan** son suficientes

- 3 cadenas con 10,000 iteraciones y 500 de *warmup*

Ahora, verificamos que no haya problemas tanto con el muestreador como con la convergencia de las cadenas simuladas.

Si inspeccionamos el resultado **global** de las simulaciones observamos el siguiente resumen estadístico

Cuadro 2: Diagnóstico de las muestras generadas

parámetro	rhat	ESS	ESS.bulk	ESS.tail
α	1.000056	7567.429	7562.728	11580.68
β	1.000066	8204.846	8176.530	12050.66
γ	1.000025	7293.087	7367.242	10859.03
δ	1.000164	7697.775	7724.638	11438.93
σ_1	1.000157	19782.703	21315.292	18477.63
σ_2	1.000199	21762.238	23754.678	18807.74
$u(t=0)$	1.000146	19380.407	19453.346	19348.58
$v(t=0)$	1.000087	15917.922	16057.044	16759.31

En particular, estamos satisfechos con el número efectivo de simulaciones que es, por mucho, mayor al mínimo valor recomendado por Vehtari et al. (2021) en p.677 (es decir $ESS=50$ para cada cadena). También el valor de \hat{R} es menor que 1.1 en cada caso por lo que, en este caso, da indicios de convergencia a la distribución objetivo.

4.2.2. Diagnósticos gráficos.

Alternativamente a los diagnósticos numéricos de las cadenas simuladas, podemos hacer diagnósticos gráficos tanto del muestreador como diagnósticos MCMC.

Podemos resumir los diagnósticos \hat{R} y ESS, de manera gráfica como sigue

El gráfico de la derecha, los puntos que representan los valores de n_{eff}/N están coloreados según si son menores que 0.1, entre 0.1 y 0.5 o mayores que 0.5. Estos valores particulares son arbitrarios en el sentido de que no tienen un significado teórico particular, pero una heurística útil es preocuparse por cualquier $n_{eff}/N < 0.1$ (Gabry and Modrák, M. 2022)

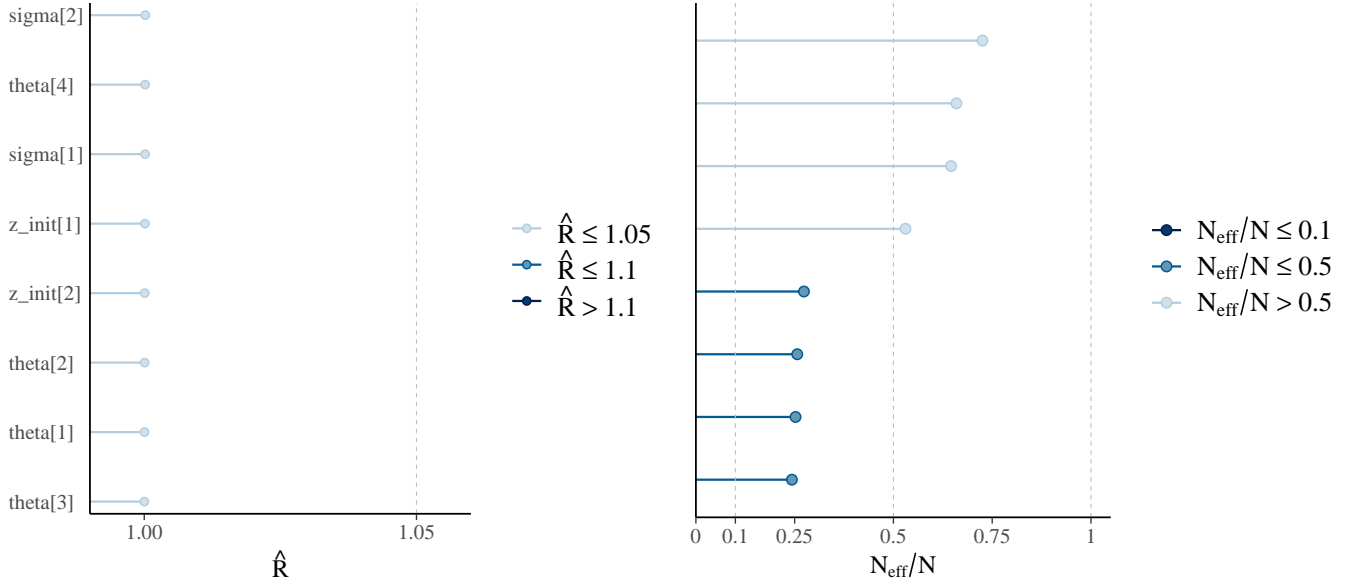


Figura 6: Diagn3sticos gr3ficos MCMC

Por otro lado, un gr3fico de coordenadas paralelas para observar de manera simult3nea ciertos patrones en los componentes asociados al sistema LV as3 como posibles trayectorias problem3ticas del hamiltoniano.

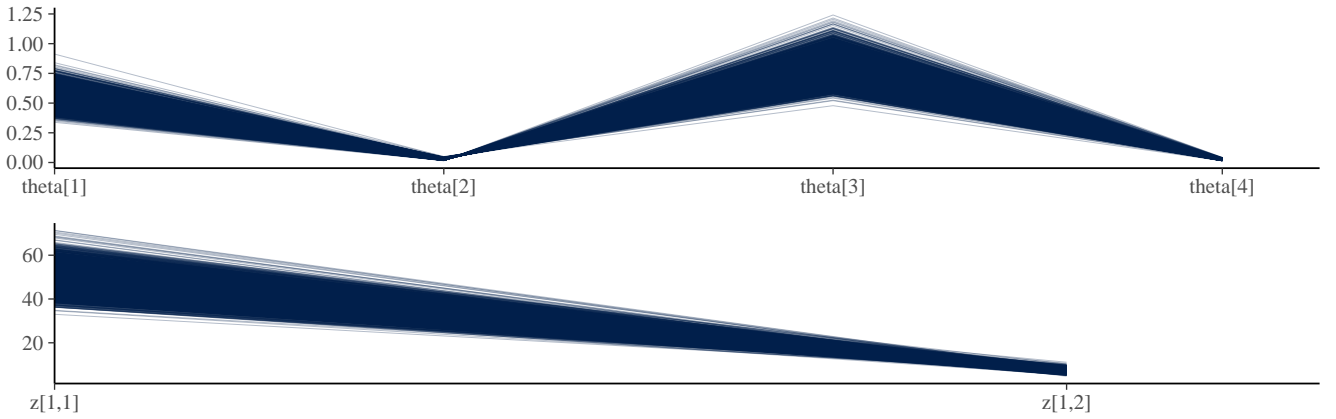


Figura 7: Gr3fico de coordenadas paralelas para los pr3metros $\alpha, \beta, \gamma, \delta$

Se verifica que ninguna trayectoria result3 problem3tica. Pues de ser as3, ser3 resaltada en color rojo.

A continuaci3n mostramos *rank plots*, que resultan ser una mejor alternativa a los gr3ficos de traza de cadenas, especialmente para cadenas de longitud larga.

Los *rank plots* son histogramas de los rangos de las simulaciones de la posterior y se grafican de manera separada para cada cadena. Si, para cada par3metro, todas las cadenas apuntan a la misma distribuci3n posterior, esperamos que los rangos en cada cadena sean aproximadamente uniformes. Si alguna cadena tiene diferente localizaci3n o par3metro de escala, esto se ver3 reflejado en la desviaci3n de la uniformidad.

Si los *rank plots* de todas las cadenas son similares, esto indica buena mezcla entre cadenas. (Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Burkner, P. C., 2021, p. 682)

Los *rank plots* de cada cadena para cada par3metro lucen uniformes lo cual nos da indicios de buena mezcla entre cadenas.

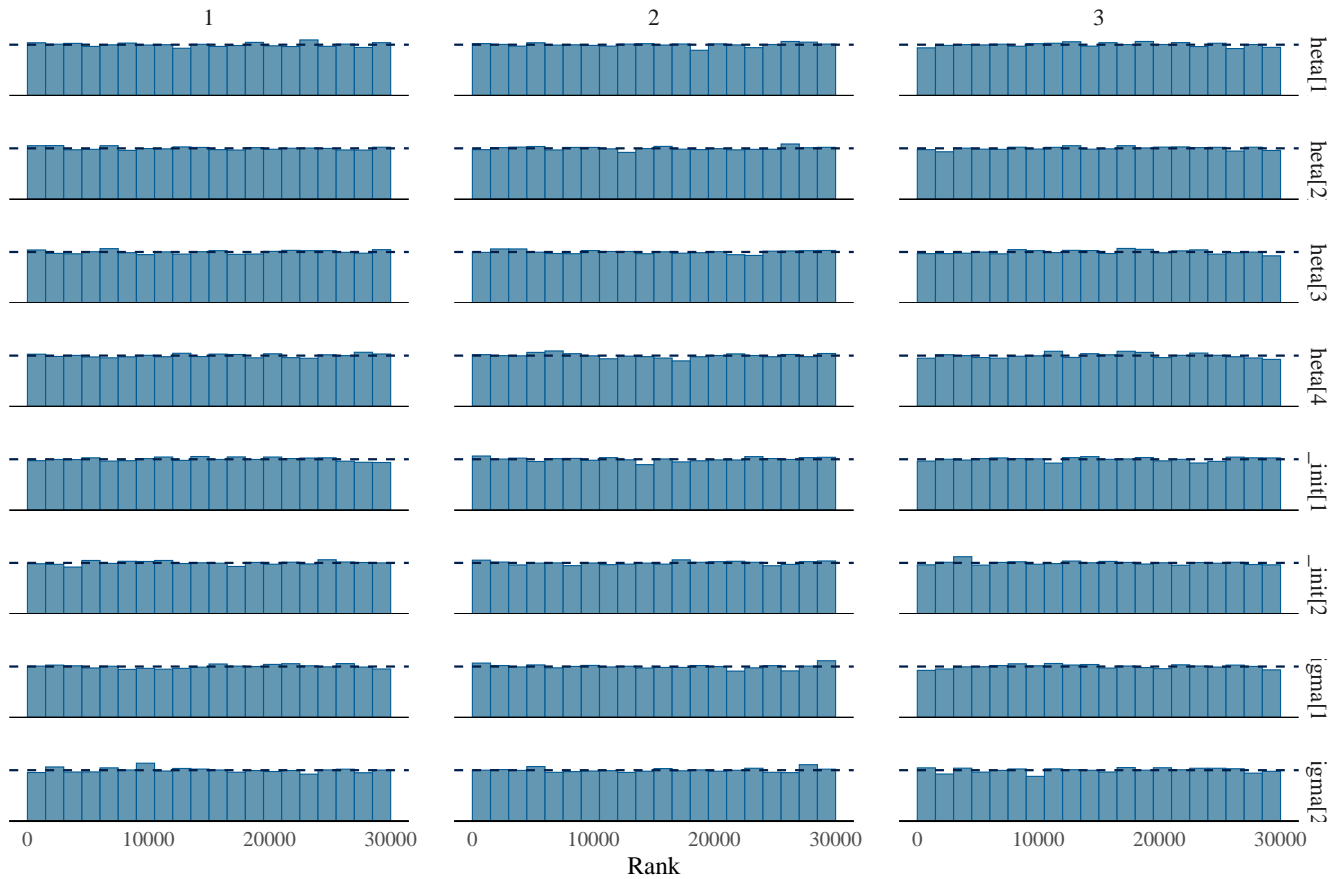


Figura 8: *Rank plots* de cada cadena para cada parámetro.

4.3. Calibración basada en simulación

En esta sección, como parte del flujo de trabajo bayesiano, empleamos simulación basada en calibración (SBC). La idea es que ajustamos nuestro modelo usando datos simulados (provenientes de la predictiva previa) con el fin de recuperar los parámetros estimados.

Aunque esta tarea debiera realizarse un número grande de veces, pues puede quedar la duda sobre si lo que observado es un artificio de una simulación (por ejemplo de fijar una semilla) o es un comportamiento generalizable [garbuno_calibration] es un paso que no terminamos de explorar por

- El gran poder de cómputo requerido: Idealmente quisiéramos ajustar muchos modelos con datos hipotéticos, lo cual requiere resolver un sistema de ecuaciones en cada iteración y después muestrear de la posterior (con sus consideraciones de adelgazamiento y calentamiento).
- Por el punto anterior, existen conjuntos de datos provenientes de la predictiva previa los cuales implican problemas de *underflow* al momento de obtener muestras de la posterior.

Sin embargo las dos consideraciones mencionadas antes, no implican un modelo mal calibrado. De hecho en la subsección Cheques predictivos posteriores de la página 19, se concluye que el modelo está bien calibrado.

Generamos, a partir de la previa, los siguientes parámetros del modelo

Cuadro 3: Parámetros simulados de la previa

parámetro	valor	parámetro	valor
$\tilde{\alpha}$	0.7471781	$\tilde{\sigma}_1$	0.1141502
$\tilde{\beta}$	0.0737454	$\tilde{\sigma}_2$	0.8068587
$\tilde{\gamma}$	0.7338785	$\tilde{u}(t=0)$	9.4899581
$\tilde{\delta}$	0.0765696	$\tilde{v}(t=0)$	0.9820811

Con ellos, usando la predictiva previa, obtenemos el conjunto de datos simulados siguiente

Cuadro 4: Datos simulados de la predictiva posterior

Año	Liebres	Línces	Año	Liebres	Línces	Año	Liebres	Línces
1900	9.950994	0.4992645	1905	338.936760	0.1437102	1910	2.063823	4.8715905
1901	18.998554	2.9296976	1906	570.349668	0.9521856	1911	2.622869	0.9689234
1902	38.086101	0.5644267	1907	310.521347	15.7123937	1912	3.491188	5.1090951
1903	63.933856	0.2445093	1908	13.000630	64.7647153	1913	7.629594	0.4000875
1904	158.703199	0.3761219	1909	3.023967	45.5782561	1914	17.376719	0.4179358
			Año	Liebres	Línces			
			1915	30.67361	0.6566568			
			1916	70.01957	0.0926112			
			1917	146.69558	0.1173297			
			1918	296.76074	0.0929496			
			1919	525.41870	0.5892563			

el se puede visualizar en la Figura 11.

Después de ajustar el modelo con los datos falsos, obtenemos la siguiente estimación de parámetros.

Cuadro 5: Estimación posterior usando datos falsos.

parámetro	media	mcse_mean	sd	q2.5	q97.5	rhat	ESS.tail
$\tilde{\alpha}$	0.7715541	0.0004240	0.0451050	0.6877516	0.8638561	1.000146	14884.95
$\tilde{\beta}$	0.0973626	0.0002547	0.0311372	0.0432776	0.1655524	1.000066	13351.40
$\tilde{\gamma}$	0.8090305	0.0006943	0.0811296	0.6581959	0.9770874	1.000102	17473.44
$\tilde{\delta}$	0.0079159	0.0000118	0.0013513	0.0055790	0.0108507	1.000166	18278.09
$\tilde{\sigma}_1$	0.3697290	0.0005766	0.0702893	0.2618358	0.5351116	1.000373	15989.70
$\tilde{\sigma}_2$	1.5042527	0.0018412	0.2551097	1.0982882	2.0930103	1.000183	18640.39
$\tilde{u}(t=0)$	6.1395480	0.0065566	0.8278862	4.6233625	7.8875895	1.000126	16551.02
$\tilde{v}(t=0)$	1.2343682	0.0030753	0.3785923	0.6774470	2.1412103	1.000028	16113.89

Y podemos comparar la media posterior, que se presenta en la segunda columna de la tabla anterior, con el valor verdadero de los parámetros que generó los datos.

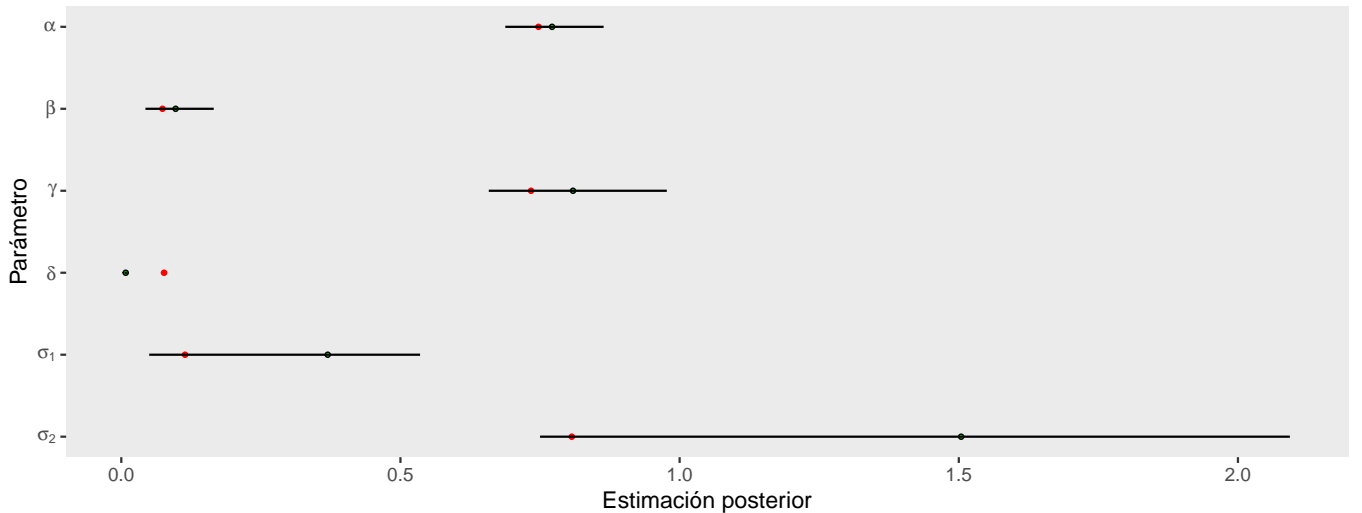


Figura 9: En rojo: valor real de parámetros $\alpha, \beta, \gamma, \delta, \sigma_1, \sigma_2$ bajo la previa; en verde: predicción puntual posterior y línea negra: intervalos al 95 %.

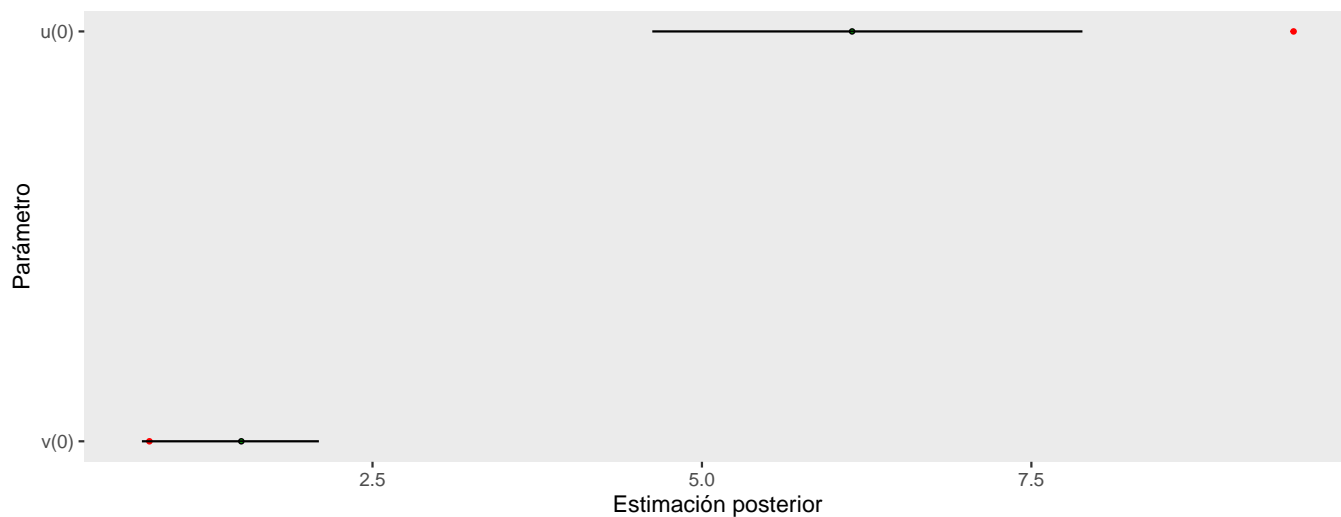


Figura 10: En rojo: valor real de condiciones iniciales $u(0), v(0)$ bajo la previa; en verde: predicción puntual posterior y línea negra: intervalos al 95 %.

Se verifica que con probabilidad 95 % se recupera el valor de casi todos los parámetros. Ya casi estamos listos para ajustar nuestro modelo con los datos observados.

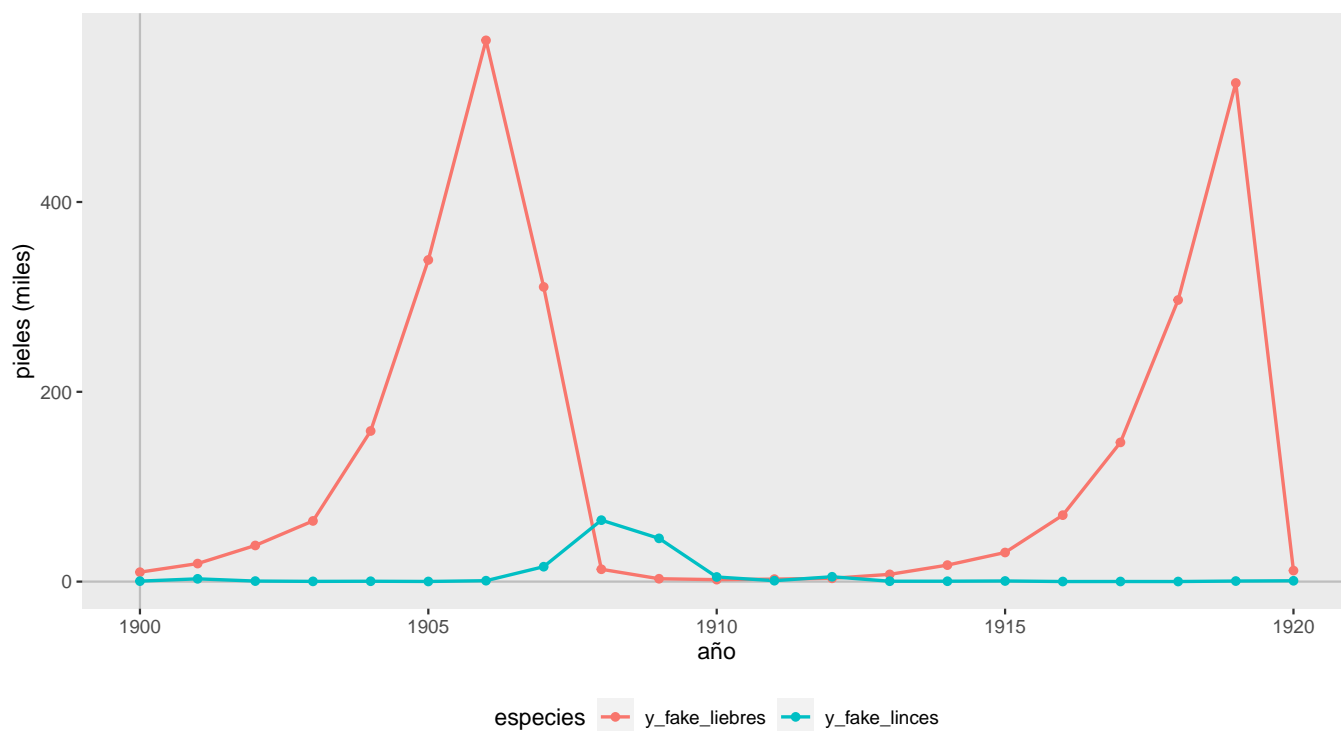


Figura 11: Datos replicados de la predictiva previa.

4.4. Estimación posterior

En la subsección anterior, analizamos el desempeño de las cadenas de Markov que utilizamos para generar muestras de la distribución posterior de nuestros parámetros. Ahora que estamos seguros de su buen comportamiento, es momento de reportar tanto las estimaciones puntuales como los intervalos de confianza correspondientes para cada uno de ellos, con el fin de tener una idea clara de los valores que podrían ser considerados como más probables y de la incertidumbre asociada a estas estimaciones.

De esta manera, podremos determinar si existen valores atípicos o si algún parámetro presenta una alta varianza en su distribución posterior.

Cuadro 6: Resumen global de las muestras generadas

parámetro	media	mcse_mean	sd	q2.5	q97.5
α	0.5477271	0.0007304	0.0635349	0.4312506	0.6813040
β	0.0277598	0.0000463	0.0041906	0.0204061	0.0367925
γ	0.7988815	0.0010552	0.0901140	0.6382273	0.9915938
δ	0.0240367	0.0000407	0.0035707	0.0178101	0.0318292
σ_1	0.2485651	0.0003050	0.0429025	0.1810300	0.3479133
σ_2	0.2515989	0.0002990	0.0441118	0.1827040	0.3546983
$u(t = 0)$	34.0899195	0.0207157	2.8839034	28.6995975	40.0277250
$v(t = 0)$	5.9436907	0.0041787	0.5272136	4.9778395	7.0696242

Esta visualización hace notar que la incertidumbre asociada al valor verdadero de los parámetros del modelo LV se reduce considerablemente. Esto es un indicio de que los datos con los que se alimenta el modelo propuesto son verdaderamente informativos.

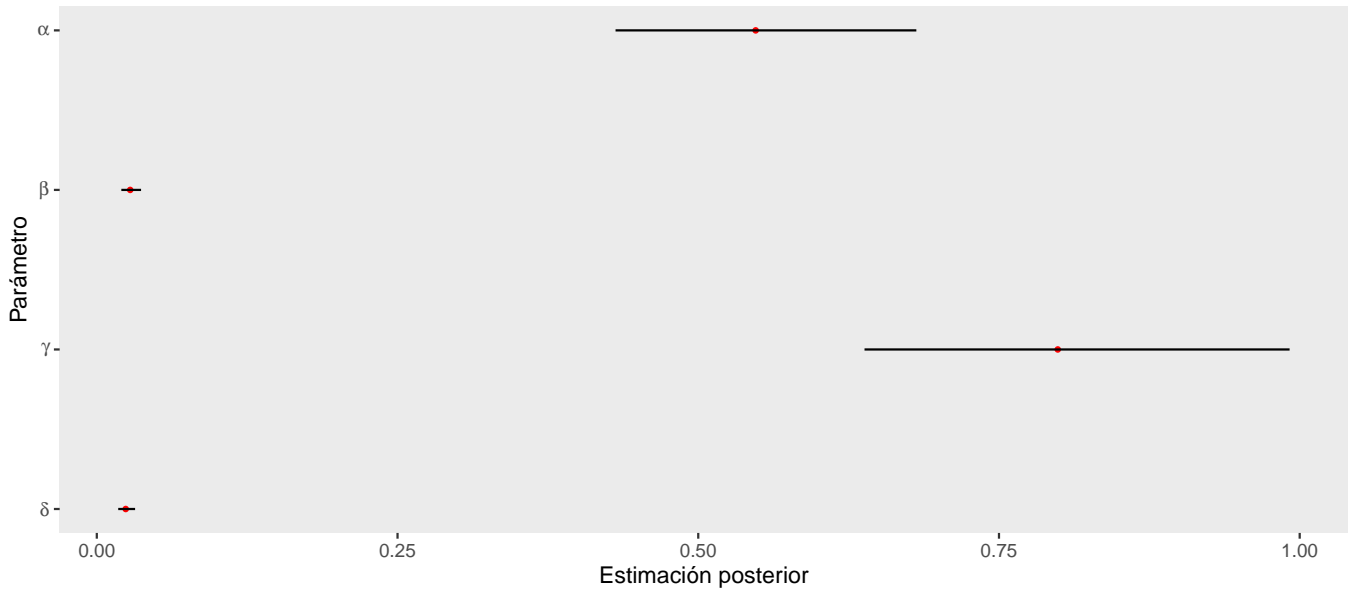


Figura 12: Predicción puntual e intervalos de confianza al 95 % para los parámetros del sistema dinámico

4.5. Contando con incertidumbre: Inferencia del tamaño de las poblaciones

El modelo estadístico que implementamos, nos permite hacer predicciones sobre el tamaño de las poblaciones de lince y liebres en diferentes momentos del tiempo, lo que nos brinda una valiosa herramienta para entender cómo evolucionan estas poblaciones.

Por supuesto, al lector y a nosotros nos interesa predecir el tamaño futuro de las poblaciones. Una forma es optar por el camino determinista, es decir **enchufar** las estimaciones posteriores que obtuvimos en la sección 4.3 en el modelo LV y con eso usar algún resolutor numérico de ecuaciones diferenciales. El lector interesado puede profundizar en este enfoque en Howard, P. (2009).

Lamentablemente, este enfoque determinista no aprovecha el potencial de la maquinaria bayesiana, en particular el poder de cuantificar incertidumbre en nuestras predicciones. Lo que haremos para predecir es aprovechar una de las herramientas más poderosas en inferencia bayesiana: la **distribución predictiva posterior**

$$\pi(y^{\text{rep}}|y) = \int \pi(y^{\text{rep}}|\theta) \pi(\theta|y) d\theta,$$

donde y^{rep} son datos hipotéticos y y los datos observados.

La distribución $\pi(\cdot|y)$, por definición, incorpora la incertidumbre asociada con los parámetros (mediante la distribución posterior) así como la incertidumbre del error de medición y variación no explicada por los datos (capturado por la verosimilitud).

4.5.1. Chequeos predictivos posteriores

En esta fase de nuestro trabajo, nos enfocamos en evaluar la precisión de nuestro modelo en relación a los datos utilizados para su estimación mediante la distribución predictiva posterior la cual nos permite **replicar** valores del número de pieles de las especies de interés y posteriormente comparar dichos valores con los observados.

En la gráfica superior de la Figura 13, los puntos representan los datos observados. Cada línea es una serie de tiempo a partir de la predictiva posterior. La variación de estas se debe al error de medición no correlacionado.

Implícitamente se verifica que los datos observados para cada especie están dentro de intervalos posteriores, de al menos 95 %. Este hecho es consistente con que el modelo está bien calibrado.

Por otro lado, en la gráfica de abajo representamos diferentes configuraciones para las poblaciones con estimación posterior, bajo pérdida cuadrática, de los parámetros.

A diferencia de las predicciones de las pieles, estas son trayectorias suaves sin error de medición pues, se obtienen a partir del modelo determinista de Lotka Volterra.

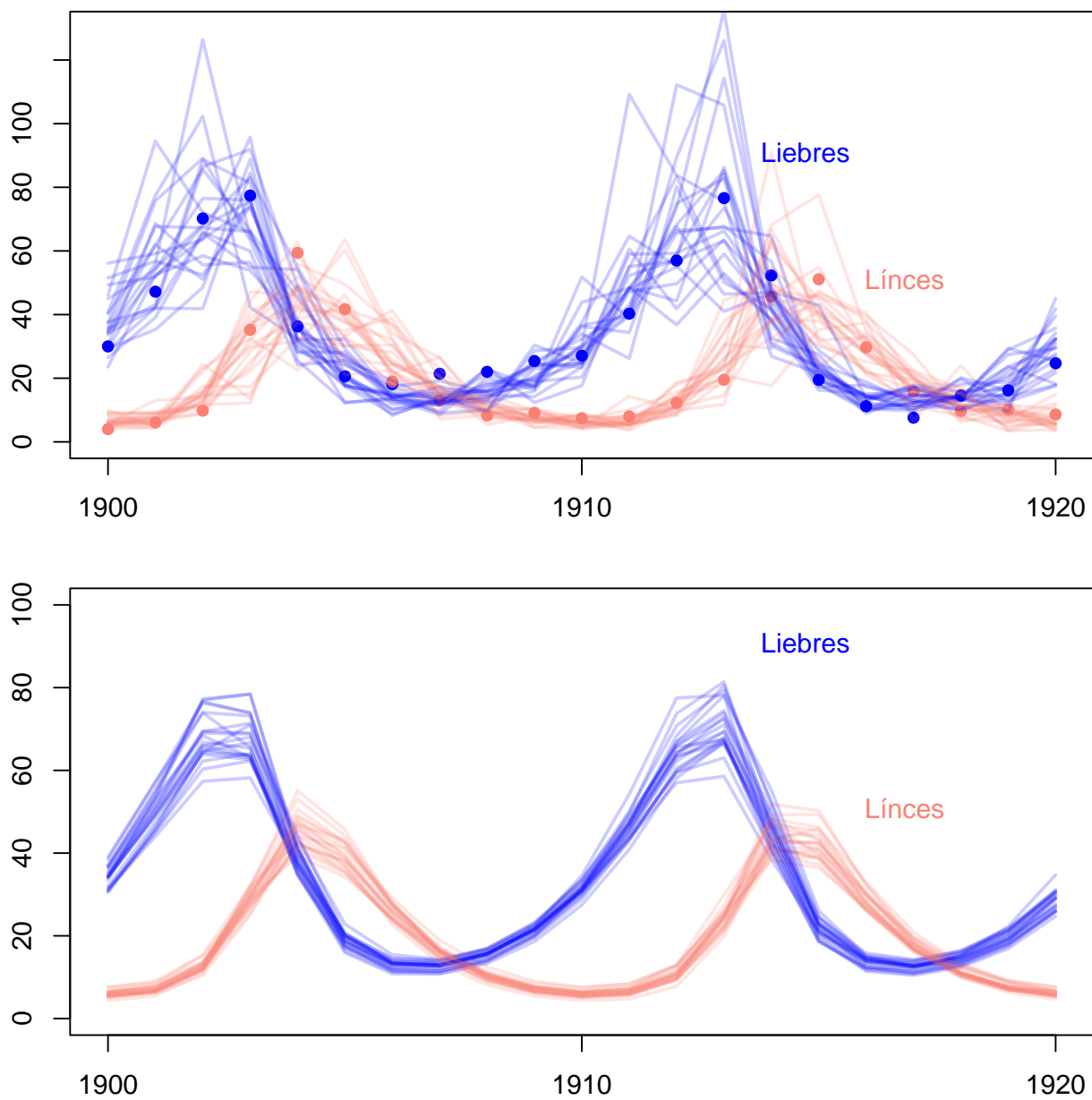


Figura 13: Chequeo predictivo posterior con datos replicados para el número de pieles y tamaño poblacional.

Si al lector no le convenció que nuestro modelo ajusta bien a los datos, presentamos una prueba *lineup* para comparar la dinámica del número de pieles (en miles) generada por el modelo con datos observados.

La figura presenta la dinámica del número de pieles a partir de la simulación de $N = 19$ conjuntos de datos replicados a partir de la predictiva posterior en **Stan**. Cada gráfica representa la media de una muestra posterior del modelo.

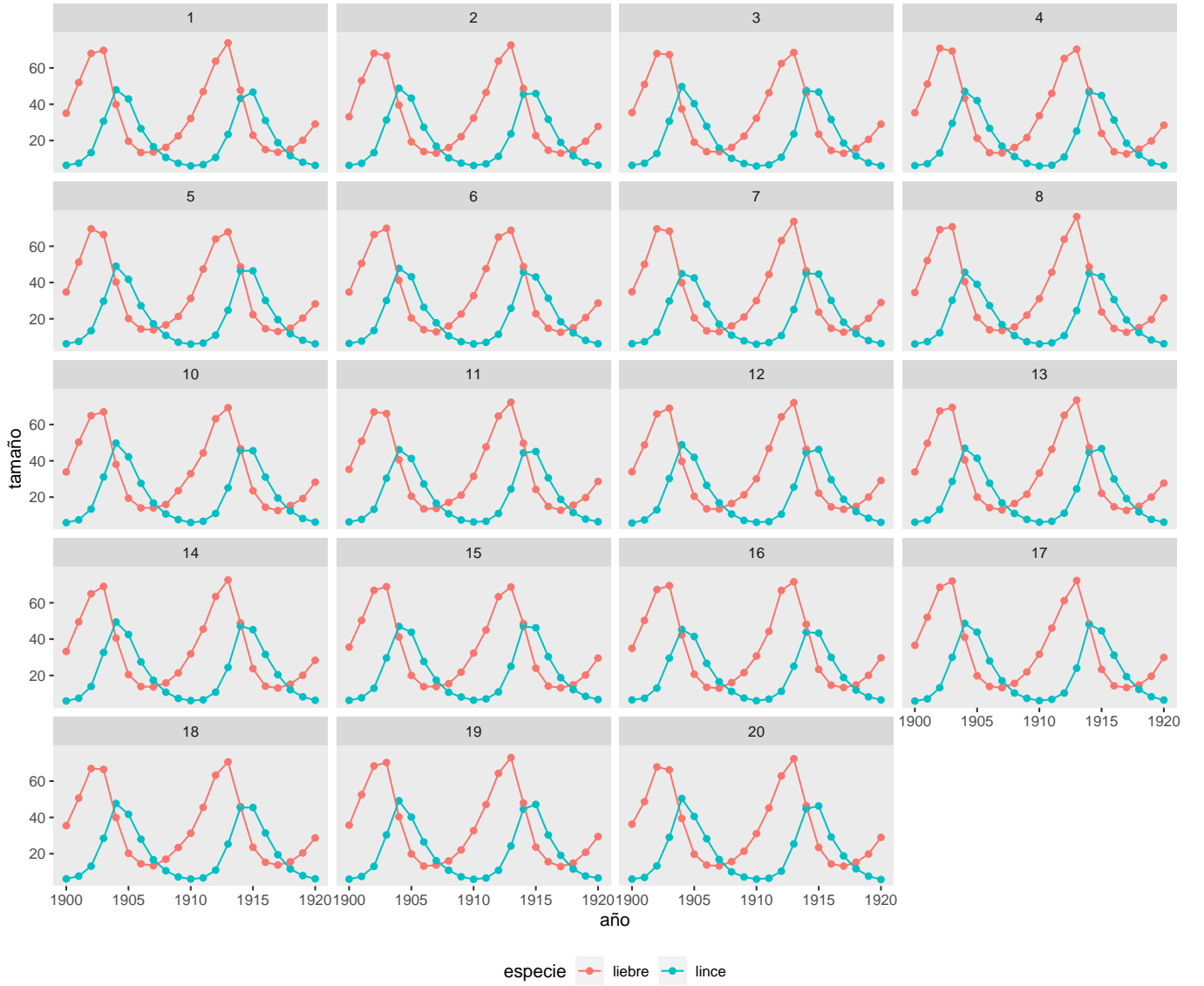


Figura 14: Prueba *lineup* de datos replicados para la evolución de las pieles recolectadas entre 1900 y 1920.

En el mismo espíritu de chequeos predictivos posteriores, presentamos dos gráficas con cien diagramas fase, es decir cien trayectorias de la dinámica del número de pieles recolectadas para cada especie. En conjunto estas gráficas proporcionan un bosquejo de la incertidumbre posterior para la dinámica de población esperada entre línices y liebres.

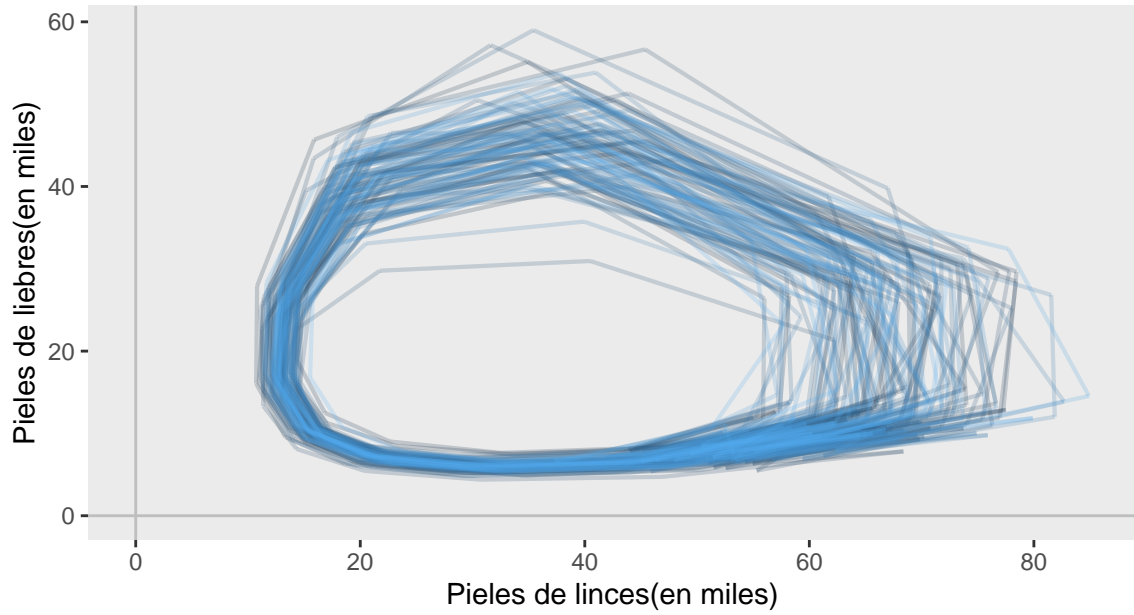


Figura 15: Dinámica del número de pieles (en miles) a partir del *solver* de Stan considerando 100 muestras posteriores de θ .

Este gráfico se generó a partir de resolver 100 sistemas dinámicos, cada uno determinado por una muestra $\theta^{(j)} \sim \pi(\cdot | y)$, mediante el *solver* de Stan. De este modo representamos, de manera aislada, la incertidumbre referente a la estimación de los parámetros en nuestra distribución posterior.

Por otro lado, la siguiente visualización se genera obteniendo muestras de la distribución predictiva posterior de manera que, nos permite presentar de manera conjunta la incertidumbre en el verdadero valor de los parámetros y en el proceso generador de datos.

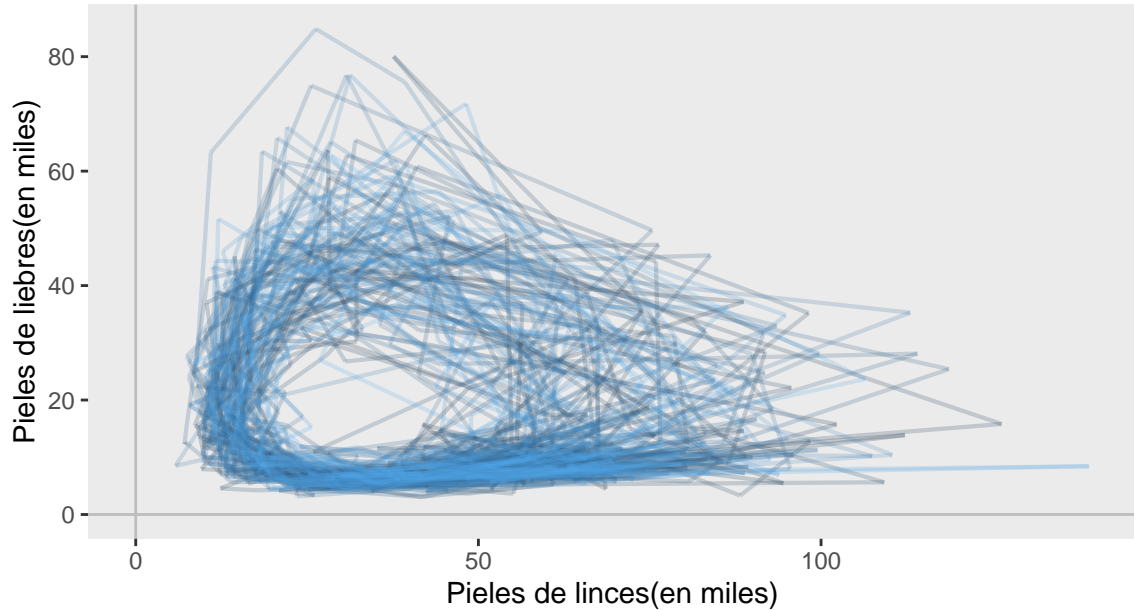


Figura 16: Número de pieles de línices y liebres (en miles) esperadas, considerando 100 muestras de los parámetros obtenidos de la distribución posterior.

4.5.2. *Forecasting y backcasting.*

Extenderemos las predicciones 20 años en el futuro y 20 en el pasado con el fin de observar si existe algún comportamiento extraño en la incertidumbre medida a largo plazo.

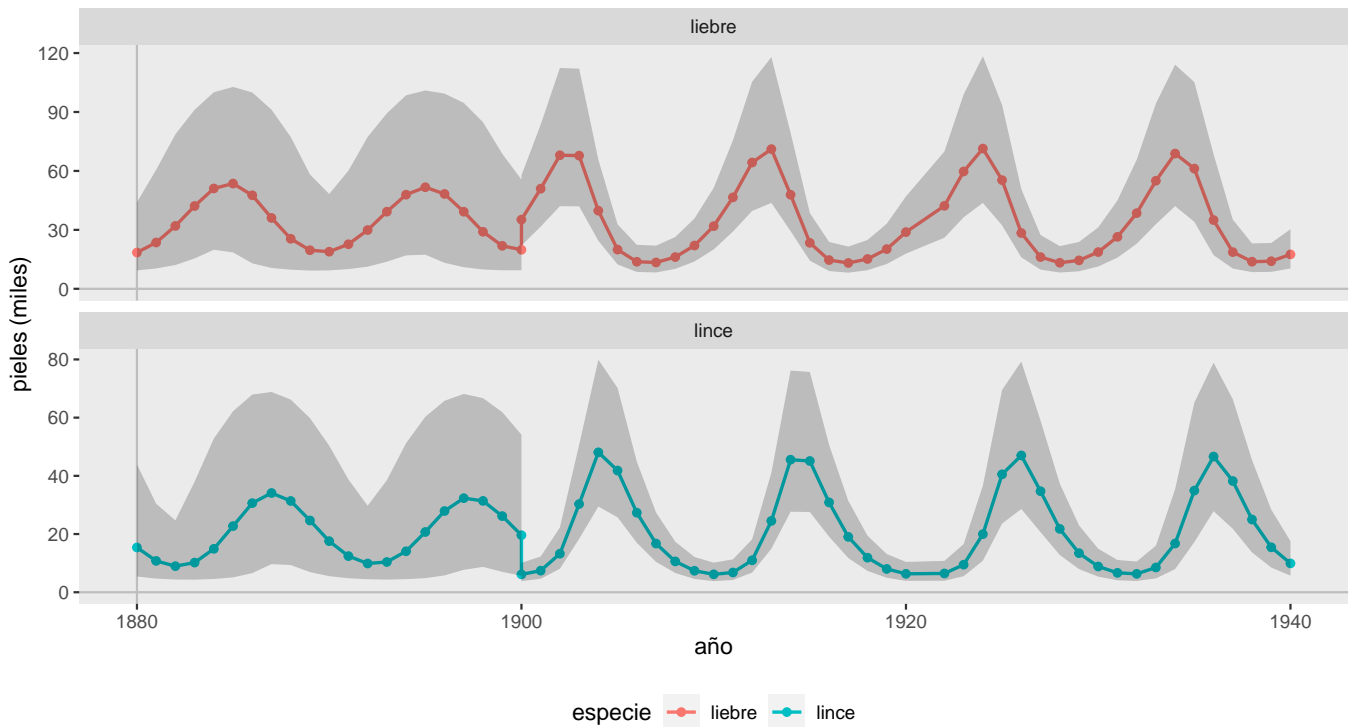


Figura 17: Predicción media de liebres y linces 50 años atrás y adelante a partir de la predictiva posterior.

Y de las gráficas anteriores se puede notar que, en el caso de *backcasting*, existe bastante incertidumbre para todos los años predichos. Posiblemente este comportamiento se deba a que no existen observaciones para años anteriores a 1900, por lo que al predecir para años donde no se cuenta con información la incertidumbre aumenta. Otro factor del porqué este comportamiento se presenta es que las poblaciones iniciales corresponden al mismo año de inicio de observación, es decir 1900, por lo que, al hacer predicciones para años pasados, la incertidumbre sobre lo que no se tiene información incrementa.

Al lector le puede parecer interesante que la predicción a futuro no siga este mismo comportamiento, sin embargo hay que recordar que este modelo es determinista, es decir que una vez que se tienen estimados los parámetros, las predicciones futuras para todos los estados pueden conocerse. De esta manera, podemos notar que la incertidumbre es menor gracias a la tendencia existente durante los años 1900 a 1920, además de que los parámetros ya fueron estimados y contamos con poblaciones iniciales para el año 1900.

4.6. Número efectivo de parámetros

Como parte de la evaluación posterior de nuestro modelo, incluimos el diagnóstico p_{WAIC} una de las métricas del criterio de información Watanabe-Akaike (WAIC).

$$p_{\text{WAIC}} = \sum_{i=1}^n \mathbb{V}_{\theta|y_n} (\log \pi(y_i|\theta)) ,$$

el cual es una medida del número efectivo de parámetros en el modelo.

Nota

Esta medida tiene que emplearse por separado para lince y liebres.

Presentamos a continuación las métricas del WAIC

Cuadro 7: Métricas WAIC para liebres (izquierda) y lince (derecha)

Métrica	Estimate	SE	Métrica	Estimate	SE
elpd_waic	-72.200002	3.401374	elpd_waic	-72.200002	3.401374
p_waic	3.426135	1.190349	p_waic	3.426135	1.190349
waic	144.400004	6.802748	waic	144.400004	6.802748

donde

- **elpd_waic**: valor esperado de la log-densidad predictiva puntual, penalizada por p_{WAIC}
- **p_waic** : número efectivo de parámetros.

En esta situación muy particular donde la estructura de los datos es una serie de tiempo no tiene sentido evaluar las métricas con **LOO-CV** (*leave one out cross validation*) pues, emplea una técnica que no respeta la estructura de serie de tiempo. En particular se viola el principio fundamental de la validación:

Hacemos validación apegándonos lo más posible a la forma en que nuestro modelo, ya en producción, va a consumir datos y hacer predicciones.

Sin embargo, es importante destacar que Vehtari y sus coautores han desarrollado técnicas para aplicar el enfoque de validación cruzada *leave-future-out* en modelos Bayesianos de series temporales (Jonah, Gabry, J., and Vehtari, A. 2023). A pesar de que por limitaciones de tiempo no pudimos aplicar estas técnicas en nuestro proyecto, consideramos que su implementación representa una línea de trabajo valiosa para futuras iteraciones y mejoras del mismo.

5. Conclusiones

“All models are wrong, but some are useful.”(George Edward Pelham Box)

Este trabajo se desarrolló apegándonos a un flujo de trabajo bayesiano con el propósito de realizar la estimación de los parámetros del modelo Lotka-Volterra utilizando un enfoque bayesiano. En la primera parte de este nos enfocamos en la definición de las distribuciones previas y de máxima verosimilitud. En una segunda etapa nos enfocamos en el diagnóstico, validación y calibración del modelo a través de chequeos predictivos previos, chequeos predictivos posteriores, diagnósticos de MCMC y la calibración por simulación. Finalmente, en la última etapa con un modelo calibrado se realizaron predicciones hacia el futuro (*forecasting*) y el pasado (*backcasting*).

Un punto relevante en la definición del modelo fue la de incorporar dos parámetros independientes (uno para cada especie) referentes a variaciones por errores de medición y la simplificación del modelo con los supuestos utilizados por Lotka-Volterra. Posteriormente, con la estimación puntual de los parámetros observamos que los valores de estos parámetros resultaron ser similares en un alto grado. Este hallazgo sugiere que existe una posibilidad de simplificar el modelo considerando únicamente un parámetro para representar las variaciones en las poblaciones de ambas especies.

Las cadenas MCMC fueron confiables y exploraban bien el soporte de la distribución posterior. Confirmamos con diagnósticos gráficos y numéricos que todos los parámetros presentaron valores de \hat{R} inferiores a 1.01, todos tuvieron muestras efectivas mayores al 20 % del total de simulaciones y de igual forma confirmamos que no había problemas en las colas de las cadenas. Las estimaciones puntuales de los parámetros no presentaron valores atípicos, todos con un error estándar Monte Carlo menor al 0.02, lo cual es relevante ya que indica una mayor precisión y confiabilidad en las estimaciones del modelo, proporcionando una medida más precisa de la incertidumbre asociada a los resultados obtenidos.

Adicionalmente, con el chequeo predictivo posterior nos confirmamos la precisión de nuestro modelo con diagnósticos gráficos como la prueba de *lineup* donde no es posible distinguir los datos observados de los datos replicados con la posterior. También pudimos observar como los datos observados están en el intervalo del 95 % de los datos replicados.

En general podemos decir que con el modelo generado se obtuvo una estimación posterior de los parámetros muy aceptables, que permitieron realizar estimaciones a largo plazo *forecasting* y *backcasting* donde observamos mayor incertidumbre en las proyecciones.

A pesar de la limitante del tiempo, este proyecto tuvo un alcance mayor al planteado en el artículo de Bob Carpenter. Se cubrieron dos de los tres ejercicios adicionales propuestos por el autor (Calibración por simulación, *Forecasting* y *Backcasting*), adicionalmente se plantea una aplicación del modelo Lotka-Volterra en el contexto mexicano abordando aspectos específicos de la vaquita marina en el Golfo de California.

5.1. Trabajo futuro

Como extensión del trabajo actual, se exploraron escenarios en los que se podría implementar un modelo Lotka-Volterra con un proceso de inferencia bayesiana para la estimación de parámetros. Estos escenarios implican la aplicación de este modelo en diferentes contextos, lo que podría ser de gran interés para futuras investigaciones en el campo de los escenarios. Sin embargo, debido a limitaciones de tiempo para este proyecto en particular, solo se presenta una descripción conceptual de estos modelos. El desarrollo detallado de los modelos quedará como un trabajo a futuro.

Caso: Vaquitas marinas en México La vaquita marina es una especie endémica del norte del Golfo de California en México y se encuentra en grave peligro de extinción. De acuerdo a cifras de la PROFEPA su población ha disminuido de una manera alarmante y preocupante. En 1997 se estimaron alrededor de 570 individuos; en 2015 menos de 60 y en 2016 sólo 30. La vaquita marina se ve amenazada indirectamente por la pesca ilegal de la totoaba. La totoaba es un pez que se vende en mercados ilegales de Asia a precios exorbitantes debido a sus supuestas propiedades curativas. La pesca ilegal de la totoaba se lleva a cabo con gigantes redes de enmalle, las cuales atrapan y matan a las vaquitas marinas.

Esta especie juega un papel crucial en el ecosistema marino del Golfo de California. La extinción de la vaquita marina en México tendría consecuencias ecológicas y ambientales de gran alcance. La vaquita marina desempeña un papel crucial en el ecosistema marino del Golfo de California al regular las poblaciones de peces, calamares y crustáceos, que son su principal fuente de alimento. Sin la presencia de la vaquita marina, se produciría un desequilibrio en la cadena alimentaria, lo que podría llevar a un aumento descontrolado de las poblaciones de sus presas y un agotamiento de los recursos disponibles. Esto afectaría a toda la red alimentaria interdependiente en la que se encuentra la vaquita marina, y podría tener un impacto negativo en otras especies marinas.

A pesar de la importancia de esta especie, de acuerdo con la PROFEPA, la vaquita marina es extremadamente difícil de estudiar debido a su comportamiento discreto, su tamaño pequeño, sus tiempos de inmersión relativamente largos y la turbidez del agua en la que habita.

El modelo de Lotka-Volterra, utilizado en la dinámica de poblaciones, puede aplicarse al caso de la vaquita marina y los pescadores para comprender mejor la interacción entre estas dos especies. El modelo se basa en la premisa de que la población de vaquitas marinas y la de pescadores están interrelacionadas y se afectan mutuamente.

El modelo de la dinámica de población de vaquitas marinas puede plantearse como:

Modelo 4(Lotka-Volterra)

$$\begin{aligned}\frac{d}{dt}v &= v(\alpha - \beta p) \\ \frac{d}{dt}p &= p(-\gamma + \delta v)\end{aligned}\tag{7}$$

Donde: $\alpha \geq 0$ es la tasa de crecimiento de las vaquitas marinas en ausencia de pescadores.

$\gamma \geq 0$ es la tasa de disminución de los pescadores en ausencia de las vaquitas marinas.

$\beta p > 0$ es la tasa de mortalidad de la población de vaquitas debida a la interacción con pescadores.

$\delta v > 0$ es la tasa de crecimiento de la población de pescadores debida a la interacción con las presas.

La utilización del modelo Lotka-Volterra en este contexto permite analizar cómo los cambios en la población de vaquitas marinas afectan a los pescadores y cómo las actividades pesqueras impactan en la supervivencia de la vaquita marina.

El enfoque bayesiano en la estimación de los parámetros del modelo Lotka-Volterra permite incorporar información previa y tener en cuenta la incertidumbre asociada a los datos limitados disponibles. De esta manera, se busca obtener una estimación más realista de los parámetros y comprender mejor la dinámica de la población de vaquitas marinas y pescadores. Los resultados de este modelo pueden proporcionar información valiosa para el diseño de estrategias de conservación y manejo sostenible, con el objetivo de preservar la vaquita marina y garantizar un equilibrio en el ecosistema marino del Golfo de California.

6. Apéndice

En esta sección presentamos con formalidad definiciones y resultados matemáticos usados a lo largo de nuestro trabajo así como una sección de análisis del modelo LV desde un enfoque de sistemas dinámicos.

6.1. Definiciones y resultados.

Definición 1 (Sistema dinámico) *Un sistema dinámico consiste en el conjunto de posibles estados junto con una regla que determina el estado presente en términos de estados pasados. (Alligood, Sauer, T. D., and Yorke, J. A., K. 1996)*

Definición 2 (Distribución log-normal) *Una variable aleatoria Y sigue una distribución **log-normal** con parámetros μ, σ si existe $X \sim \mathbf{N}(\mu, \sigma)$ tal que $Y = \exp(X)$.*

Teorema 1 (Media y varianza lognormal) *Sea $Y \sim \mathbf{Log-normal}(\mu, \sigma)$ entonces*

$$\mathbb{E}(Y) = \exp\left(\mu + \frac{\sigma^2}{2}\right) \qquad \mathbb{V}(Y) = (\exp(\sigma^2) - 1) \exp(2\mu + \sigma^2)$$

6.2. Análisis dinámico

Una pregunta que es de especial interés desde un enfoque de sistemas dinámicos es ¿cómo es el comportamiento de las soluciones del sistema dinámico de Lotka-Volterra?

Podríamos responder de inmediato nuestra pregunta si pudiéramos resolver el sistema de manera analítica. No obstante optaremos por una manera más elegante de hacerlo: si eliminamos el tiempo t y nos centramos en la relación entre u y v , conseguimos el objetivo.

En términos geométricos estudiaremos el plano (u, v) que comúnmente se llama **espacio fase**. Aquí observamos geoméricamente las **órbitas** o **trayectorias** de las soluciones, es decir las curvas que representan la relación funcional entre u, v parametrizadas por el tiempo t .

Para eliminar t de la ecuación (2), considere

$$\frac{dv/dt}{du/dt} = \frac{d}{du}v = \frac{v(-\gamma + \delta u)}{u(\alpha - \beta v)}$$

la cual es una ecuación diferencial de variables separables cuya solución satisface

$$\int \frac{-\gamma + \delta u}{u} du = \int \frac{\alpha - \beta v}{v} dv$$

de donde se sigue

$$-\gamma \log(u) - \alpha \log(v) + \delta u + \beta v = c$$

con c una constante de integración. Así pues, definiendo

$$V(u, v) = -\gamma \log(u) - \alpha \log(v) + \delta u + \beta v$$

toda órbita del sistema está dada implícitamente por la ecuación $V(u, v) = c$

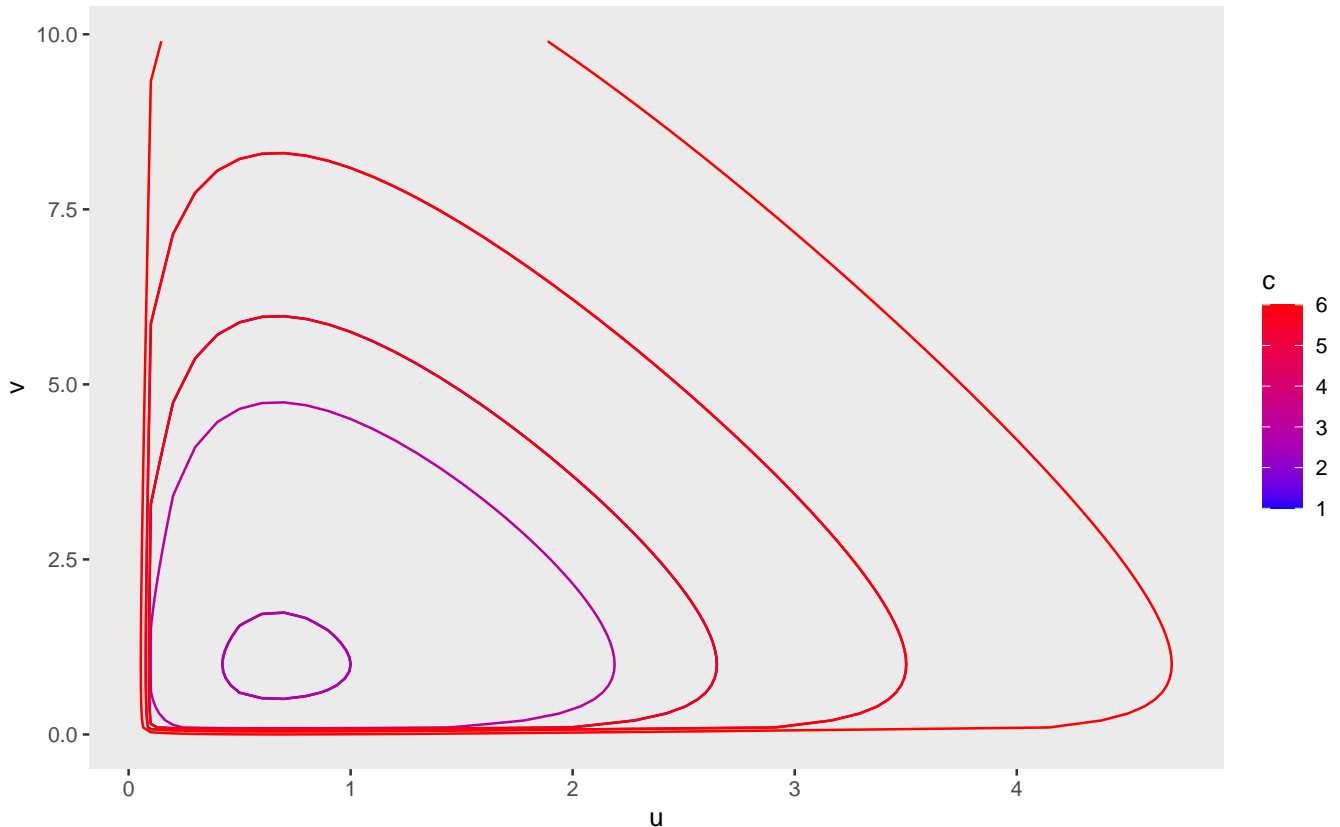


Figura 18: Espacio fase para diferentes soluciones del modelo Lotka-Volterra.

Y, aunque este no es un argumento riguroso, se puede verificar que las soluciones del modelo Lotka Volterra oscilan para cierta configuración de sus parámetros.

Por otro lado, note que el sistema dinámico (2) es equivalente, asumiendo $u, v \neq 0$, a

$$\begin{aligned}\frac{d}{dt} \log(u) &= \alpha - \beta v \\ \frac{d}{dt} \log(v) &= -\gamma + \delta u\end{aligned}$$

de donde se sigue, integrando sobre un solo periodo de oscilación T de modo que $u(0) = u(T), v(0) = v(T)$, que:

$$\frac{1}{T} \int_0^T u(t) dt = \frac{\alpha}{\beta} \quad (8)$$

$$\frac{1}{T} \int_0^T v(t) dt = \frac{\gamma}{\delta} \quad (9)$$

lo cual significa que en cada periodo de oscilación, el promedio de ambas especies es α/β y γ/δ , respectivamente y éste promedio no depende de las condiciones iniciales.

Además, entre más grande sea la tasa de crecimiento α de las presas y más pequeña sea la tasa de extinción γ de los depredadores, de acuerdo con (8), el promedio de presas u disminuye y el promedio de v aumenta, es decir la proporción $v/(u + v)$ de depredadores aumenta.

7. Bibliografía

- Alligood, K., Sauer, T. D., and Yorke, J. A., K. 1996. *CHAOS: An Introduction to Dynamical Systems*. Springer.
- Bacaër, N. 2011. *A Short Story of Mathematical Population*. Springer.
- Brauer, F., and Castillo-Chavez, C. 2012. *Mathematical Models in Population Biology and Epidemiology*. Vol. 40. Texts in Applied Mathematics. New York, NY: Springer New York. <https://doi.org/10.1007/978-1-4614-1686-9>.
- Carpenter, B. 2018. “Predator-Prey Population Dynamics: The Lotka-Volterra Model in Stan.” <https://mc-stan.org/users/documentation/case-studies/lotka-volterra-predator-prey.html>.
- Czajkowski, C., and Schillak, P. 1990. “Iterative Solutions of the Lotka-Volterra Equations.” *Il Nuovo Cimento D*, 1990.
- Gabry, J., and Modrák, M. 2022. “Visual MCMC Diagnostics Using the Bayesplot Package.” <https://mc-stan.org/bayesplot/articles/visual-mcmc-diagnostics.html>.
- Garbuno-Iñigo, A. 2023a. “Calibración Basada En Simulación.” <https://github.com/agarbuno/modelacion-bayesiana/blob/spring-2023/docs/08-calibracion.pdf>.
- . 2023b. “Muestreo Predictivo.” <https://github.com/agarbuno/modelacion-bayesiana/blob/spring-2023/docs/07-muestreo-predictivo.pdf>.
- . 2023c. “Workflow.” <https://github.com/agarbuno/modelacion-bayesiana/blob/spring-2023/docs/06-workflow.pdf>.
- Howard, P. 2009. “Modeling Basics. Lecture Notes for Math 442, Texas A&M University.”
- Iannelli, M., and Pugliese, A. n.d. *An Introduction to Mathematical Population Dynamics*. Vol. 79. Springer.
- Jonah, P., Gabry, J., and Vehtari, A. 2023. “Approximate Leave-Future-Out Cross-Validation for Bayesian Time Series Models.” <https://mc-stan.org/loo/articles/loo2-lfo.html>.
- Murty, K. N., Srinivas, M. A. S., and Prasad, K. R. 1990. “Approximate Analytical Solutions to the Three-Species Ecological System.” *Journal of Mathematical Analysis and Applications*, 1990. <https://www.sciencedirect.com/science/article/pii/0022247X9090433G>.
- STAN Development Team. n.d. “Stan User’s Guide Version 2.31.” https://mc-stan.org/docs/2_31/stan-users-guide-2_31.pdf.
- Vehtari, A., Gelman, A., Simpson, D., and Carpenter, B. 2021. “Rank-Normalization, Folding, and Localization: An Improved \hat{R} for Assessing Convergence of MCMC (with Discussion).” *Bayesian Analysis*, 2021. <https://www.sciencedirect.com/science/article/pii/0022247X9090433G>.
- Volterra, V. 1926. “Variazioni e Fluttazioni Del Numero d’individui in Specie Animali Conviventi.” *Mem. Acad. Sci. Lincei* 2: 31–113.