



| **UNR** Universidad  
Nacional de Rosario

UNIVERSIDAD NACIONAL DE ROSARIO

FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA

## **TRABAJO PRÁCTICO: "El bayesiano interestelar"**

ESTADÍSTICA BAYESIANA

**Autores:** Priscila Fenoy - María Victoria Fernandez - Francisco Ovando

**Docentes:** Tomás Capretto - Ignacio Evangelista

**Fecha:** 9 de Junio de 2025

# Índice

<b>Introducción</b>	<b>1</b>
<b>Experimento en el planeta X</b>	<b>1</b>
<b>Relación entre el modelo cinemático y el modelo lineal normal</b>	<b>2</b>
<b>Planteo del modelo lineal</b>	<b>3</b>
Modelo 1: Regresión cuadrática con distribuciones <b>a priori</b> uniformes . .	3
Especificación del modelo . . . . .	4
Estructura del modelo expresada en Stan . . . . .	4
<b>Análisis gráfico del modelo bayesiano</b>	<b>7</b>
<b>Construcción de distribuciones a priori informativas a partir de datos de la Tierra y Marte</b>	<b>8</b>
<b>ITEM 7</b>	<b>9</b>

## Introducción

En un contexto donde la exploración espacial se vuelve cada vez más relevante, contar con herramientas para comprender y estimar características fundamentales de otros planetas se vuelve clave. En esta oportunidad hemos sido convocados por la Agencia de Investigación Espacial para participar en un experimento crucial: medir la aceleración gravitatoria en un planeta recién descubierto, el Planeta X. La estrategia elegida para tal fin, aunque sencilla en apariencia, exige precisión: lanzar una pelota de tenis verticalmente hacia arriba y registrar detalladamente su trayectoria. Con la ayuda de una nave espacial de última generación y dispositivos tecnológicos capaces de captar con exactitud el movimiento del objeto

Antes de llegar al destino final, el entrenamiento incluyó ensayos previos en la Tierra y una parada intermedia en Marte. Estos pasos no solo sirvieron para familiarizarnos con los instrumentos y condiciones del experimento, sino también para recolectar información valiosa que luego podrá ser utilizada para enriquecer el análisis estadístico.

A lo largo del informe se presentan visualizaciones exploratorias, especificaciones de modelos, implementaciones en **Stan** y análisis de los resultados obtenidos. El objetivo central es obtener una estimación confiable de la gravedad en el Planeta X, así como evaluar el comportamiento predictivo del modelo en distintos escenarios de lanzamiento, contribuyendo así a resolver un interrogante que mantiene en vilo tanto a la comunidad científica como a la población general.

## Experimento en el planeta X

Con el objetivo de comenzar a explorar los datos recolectados durante el experimento realizado en el Planeta X, a continuación se presenta la trayectoria registrada de la pelota de tenis en función del tiempo. Esta visualización permite una primera evaluación visual del comportamiento del sistema y del funcionamiento del dispositivo de registro utilizado durante la misión.

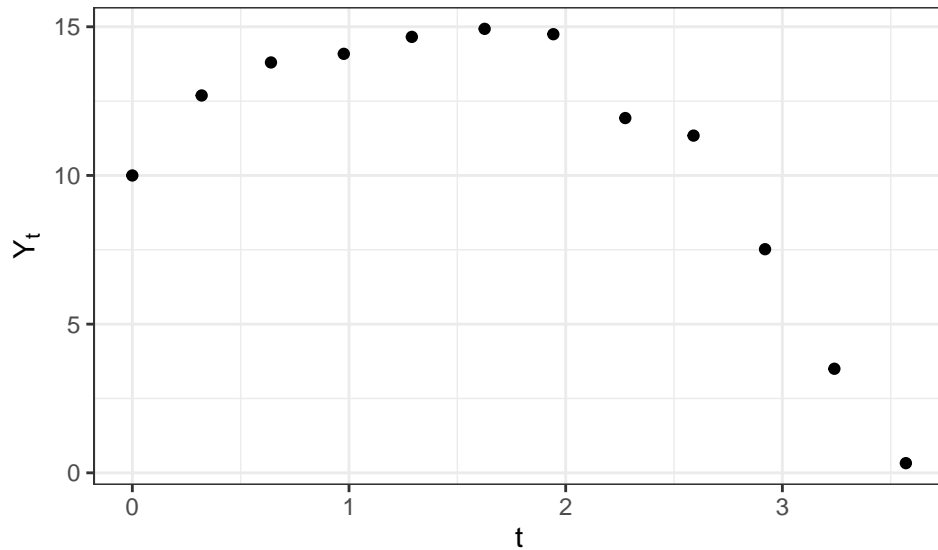


Figura 1: Posición de la pelota en el planeta X durante el experimento

La Figura 1 muestra una secuencia de mediciones que sigue, en términos generales, la forma esperada para un lanzamiento vertical bajo aceleración constante. La pelota fue lanzada desde una altura de 10 metros y aunque la trayectoria no se ajusta perfectamente a una curva suave, las pequeñas desviaciones observadas pueden atribuirse a errores de medición, propios del instrumento o del entorno experimental. Los datos parecen haber sido recolectados de manera adecuada, por lo que se consideran válidos para avanzar hacia una modelización estadística más precisa.

## Relación entre el modelo cinemático y el modelo lineal normal

En este trabajo se utilizará como base una ecuación fundamental de la cinemática que describe el movimiento de un objeto bajo aceleración constante. Esta, vincula la posición del objeto en función del tiempo con parámetros clave del movimiento: la aceleración, la velocidad inicial y la posición de partida.

Se trabajará con un modelo lineal normal, donde los coeficientes del mismo se pueden expresar en función de los parámetros de la ecuación cinemática:

$$\begin{cases} \beta_0 = x_0 \\ \beta_1 = v_0 \\ \beta_2 = \frac{1}{2}a \end{cases}$$

Además, se puede establecer un vínculo entre las variables de ambos modelos que se puede plantear de la siguiente manera:

$$\begin{cases} X_1 = t \\ X_2 = t^2 \end{cases}$$

A partir de esta relación, se propondrá el modelo estadístico que servirá como base para el análisis bayesiano posterior, sin incorporar aún información a priori sobre los parámetros.

## Planteo del modelo lineal

A continuación, se presenta el modelo lineal que será utilizado para analizar los datos obtenidos en el experimento realizado en el Planeta X:

$$Y_{t_i} \mid \mu_{t_i}, \sigma \sim \text{Normal}(\mu_{t_i}; \sigma^2) \quad i = 1, \dots, 12$$

$$\mu_{t_i} = \beta_0 + \beta_1 t_i + \beta_2 t_i^2$$

## Modelo 1: Regresión cuadrática con distribuciones a priori uniformes

Este modelo tiene como objetivo describir la relación entre tiempo transcurrido desde el inicio de la observación (en segundos) denotada como  $t$  y posición vertical final de la pelota después de un tiempo  $t$  (en metros) denotada como  $y$ , a través de una regresión cuadrática con errores normales. La estimación se realiza bajo el enfoque bayesiano utilizando el lenguaje probabilístico **Stan**, a través del muestreo por cadenas de Markov (MCMC).

## Especificación del modelo

El modelo asume que la relación entre  $y$  y  $t$  puede ser representada mediante una función cuadrática, es decir:

$$y_i \sim \mathcal{N}(\mu_i, \sigma), \quad \text{donde} \quad \mu_i = \beta_0 + \beta_1 t_i + \beta_2 t_i^2$$

donde:

- $y_i$ : valor observado de la posición vertical de la pelota en el tiempo  $t_i$  (en metros).
- $\beta_0$ : posición vertical inicial de la pelota en  $t_i = 0$  (en metros).
- $\beta_1$ : velocidad inicial de la pelota al comienzo del intervalo de tiempo  $t_i$  (en metros por segundo).
- $\beta_2$ : el doble de la aceleración constante que experimenta la pelota (en metros por segundo al cuadrado).
- $\sigma$ : desviación estándar del error.
- $t_i$ : tiempo transcurrido desde el inicio de la observación (en segundos).

## Estructura del modelo expresada en Stan

```
modelo1 <- "  
data {  
  int<lower=0> N; // Cantidad de observaciones  
  vector[N] t;   // Valores de la variable predictora  
  vector[N] y;   // Valores de la variable respuesta  
}  
parameters {  
  real beta0;      // Intercepto  
  real beta1;      // Pendiente  
  real beta2;      // Pendiente  
  real<lower=0> sigma; // Desvio estándar del error  
}  
model {  
  y ~ normal(beta0 + beta1 * t + beta2 * t^2, sigma);  
}  
generated quantities {  
  vector[N] mu;  
  vector[N] y_rep;  
}
```

```
// Calcular 'mu'
mu = beta0 + beta1 * t + beta2 * t^2;

for (i in 1:N) {
  // Obtención de muestras de la distribución predictiva a posteriori
  y_rep[i] = normal_rng(mu[i], sigma);
}
}"
```

En este modelo se utilizaron distribuciones *a priori* no informativas, permitiendo que los datos tengan un rol principal en la actualización de la distribución a posteriori. Por lo que, las distribuciones *a priori* utilizadas son:

- Para  $\beta_0$ ,  $\beta_1$  y  $\beta_2$ : distribuciones *a priori* uniformes, el plano se encuentra sobre los reales.
- Y para  $\sigma$ : un *prior* uniforme, el plano se encuentra sobre los reales positivos.

Luego de ajustar el modelo bayesiano a los datos obtenidos en el experimento del Planeta X, se generaron muestras de la distribución predictiva a posteriori de la respuesta ( $y_{rep}$ ) a partir de los valores ajustados  $\mu$  y el desvío estimado  $\sigma$ , lo cual permite realizar chequeos de ajuste (posterior predictive checks) para evaluar si el modelo reproduce adecuadamente los patrones observados en los datos, es decir se analizaron las distribuciones marginales a posteriori de los parámetros del modelo:

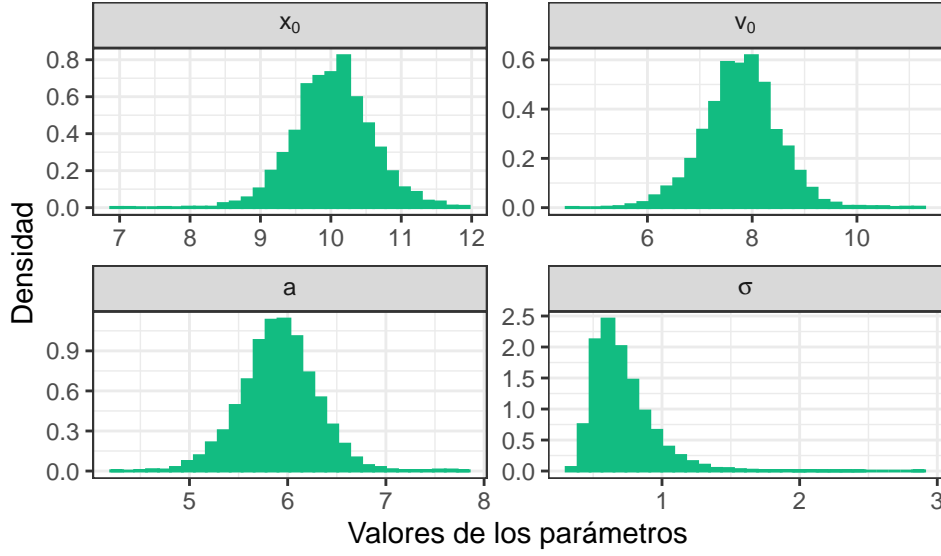


Figura 2: Distribución marginal de los parámetros estimados obtenidos a partir de las muestras generadas

En la Figura 2 se muestran las distribuciones a posteriori de los parámetros estimados por el modelo. En todos los casos, se observa que las cadenas convergen a distribuciones unimodales, con densidades concentradas, lo que indica buena estabilidad del muestreo y una adecuada capacidad del modelo para aprender sobre los parámetros. En particular, se observa que:

- El posterior de  $a$  presenta una forma aproximadamente simétrica, concentrada en torno a un valor cercano a  $6 \text{ m/s}^2$ . Esto sugiere que la intensidad de la gravedad en el Planeta X es inferior a la terrestre ( $9.8 \text{ m/s}^2$ ), pero mayor a la de Marte ( $3.7 \text{ m/s}^2$ ).
- El posterior de la velocidad inicial estimada  $v_0$  se concentra alrededor de  $8 \text{ m/s}$ , en línea con lo esperado según el entrenamiento previo del astronauta. La forma de la distribución es simétrica y angosta, indicando baja incertidumbre.
- El posterior del valor de  $x_0$  se sitúa en torno a los 10 metros, que coincide con la altura de la nave espacial The Bayesian desde donde se realizó el lanzamiento. Esta estimación es coherente con las condiciones iniciales conocidas del experimento.
- A diferencia de las anteriores, el posterior de  $\sigma$  es asimétrico y presenta una cola pesada hacia la derecha. Se observa que la mayor concentración cerca



de valores bajos sugiere que el modelo logra ajustar bastante bien los datos observados.

## Análisis gráfico del modelo bayesiano

En este ítem se muestra la curva de regresión estimada para la distribución condicional de  $\mu_i$  acompañada de sus bandas de credibilidad del 95%, que representan la incertidumbre asociada a la predicción del modelo ajustado.

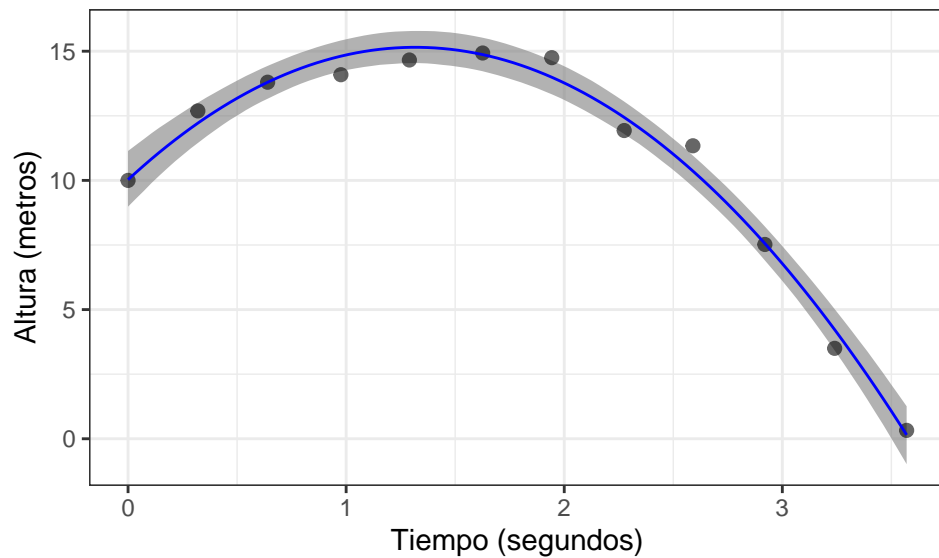


Figura 3: Curva de regresión estimada para  $\mu[i]$  en el modelo 1

La curva visualizada en la Figura representa la posición esperada de la pelota a lo largo del tiempo, según el modelo ajustado con los datos del Planeta X.

La curva sigue una trayectoria parabólica, lo cual es coherente con el movimiento de un objeto lanzado hacia arriba bajo la influencia de la gravedad. Se observa que la mayoría de los puntos quedan contenidos dentro de las bandas de credibilidad, lo que indica un buen ajuste del modelo a los datos, aunque con un prior mas informativo esta estimación podría mejorar.

Además, se incluye un gráfico de la distribución predictiva a posteriori para distintos valores de tiempo, lo que permite evaluar el comportamiento esperado de la pelota en futuros lanzamientos.

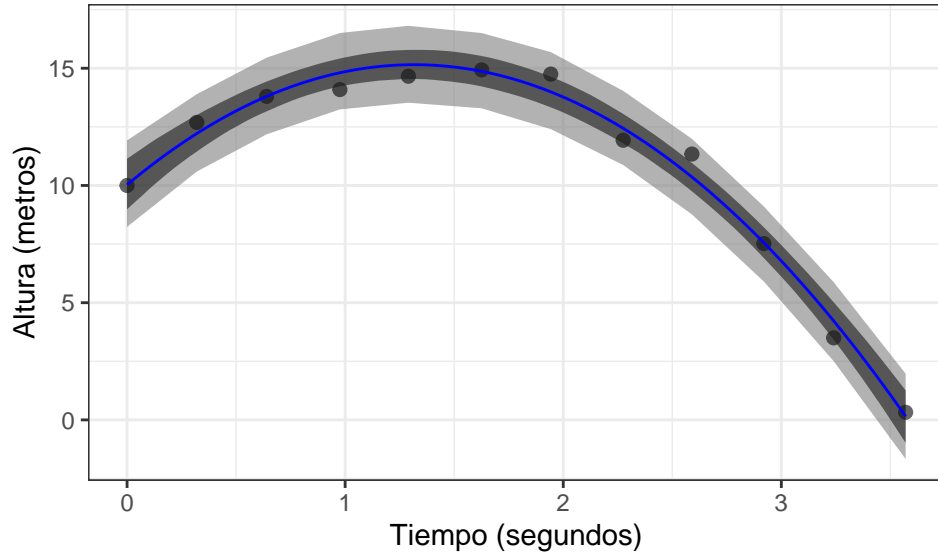


Figura 4: Distribución predictiva a posteriori en el modelo 2

En la Figura 4 se observa que las bandas de credibilidad asociadas a la distribución predictiva a posteriori son bastante angostas. Esto refleja que el modelo tiene una buena capacidad para predecir futuras observaciones.

## Construcción de distribuciones a priori informativas a partir de datos de la Tierra y Marte

Se dispone de datos provenientes de experimentos realizados en la Tierra y Marte, cuyos valores de gravedad son conocidos. En este punto, se visualizan dichos datos junto con los del Planeta X, con el objetivo de analizar similitudes y diferencias en los comportamientos registrados. Esta comparación permite identificar patrones útiles para definir distribuciones a priori más informativas para los parámetros del modelo, mejorando así la calidad de las inferencias bayesianas sobre la gravedad en el Planeta X.

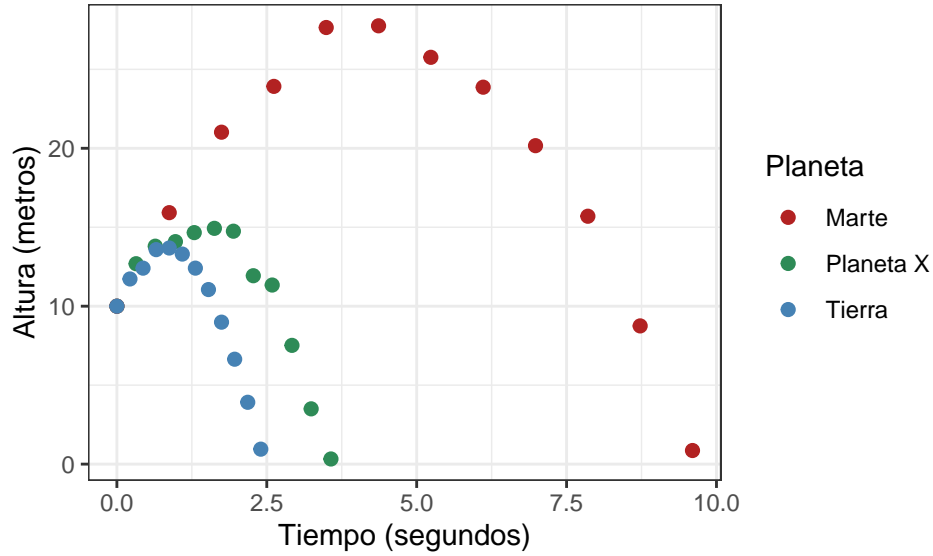


Figura 5: Posición vertical de la pelota en función del tiempo

La Figura 5 muestra la posición vertical de la pelota en función del tiempo para los tres planetas. En la Tierra, la pelota alcanza una menor altura y regresa rápidamente al suelo, lo que refleja una mayor aceleración gravitatoria. En el Planeta X, se observa una altura intermedia y un tiempo de vuelo moderado, mientras que en Marte la pelota alcanza la mayor altura y permanece más tiempo en el aire, lo que sugiere una gravedad más débil.

En conjunto, los datos sugieren que la gravedad en el Planeta X es más cercana a la de la Tierra que a la de Marte. Esta comparación proporciona una base útil para definir distribuciones a priori informativas sobre la gravedad en el planeta de interés, aprovechando el conocimiento disponible sobre los entornos conocidos.

## ITEM 7

$$\beta_0 \sim N(10, 0.5) \quad \beta_1 \sim N(8, 1) \quad \beta_2 \sim N(-4, 1.5) \quad \sigma \sim N^+(1.5)$$

grav marte 3.73 grav tierra 9.807 media 6.7685

$$-0.5*(6.7685)$$

```

modelo2 <- "
data {
  int<lower=0> N; // Cantidad de observaciones
  vector[N] t;   // Valores de la variable predictora
  vector[N] y;   // Valores de la variable respuesta
}
parameters {
  real beta0;      // Intercepto
  real beta1;      // Pendiente
  real beta2;      // Pendiente
  real<lower=0> sigma; // Desvio estándar del error
}
model {
  beta0 ~ normal(10, 0.5);
  beta1 ~ normal(8, 1);
  beta2 ~ normal(-4, 1.5);
  sigma ~ normal(0, 1.5);
  y ~ normal(beta0 + beta1 * t + beta2 * t^2, sigma);
}
generated quantities {
  vector[N] mu2;
  vector[N] y_rep2;

  // Calcular 'mu'
  mu2 = beta0 + beta1 * t + beta2 * t^2;

  for (i in 1:N) {
    // Obtención de muestras de la distribución predictiva a posteriori
    y_rep2[i] = normal_rng(mu2[i], sigma);
  }
}"

```

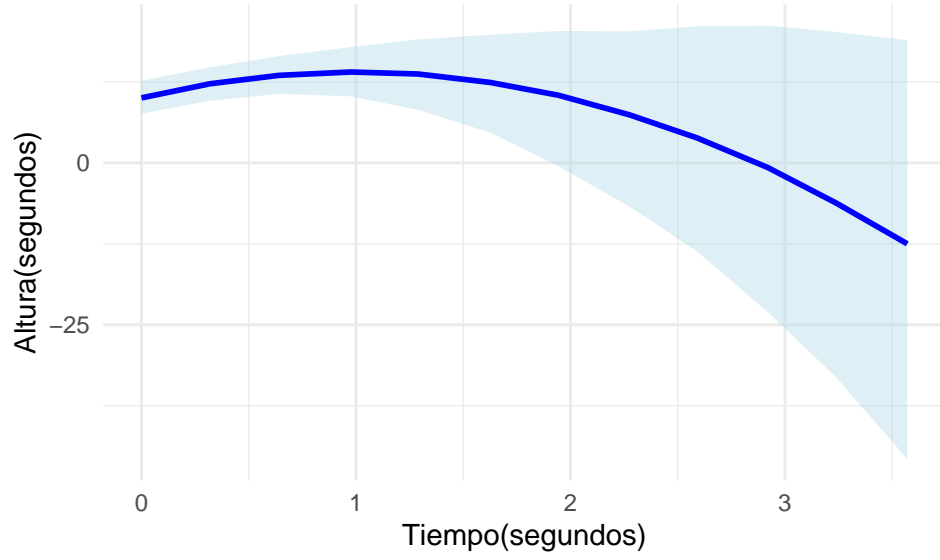


Figura 6: Distribución predictiva a priori de la posición vertical de la pelota

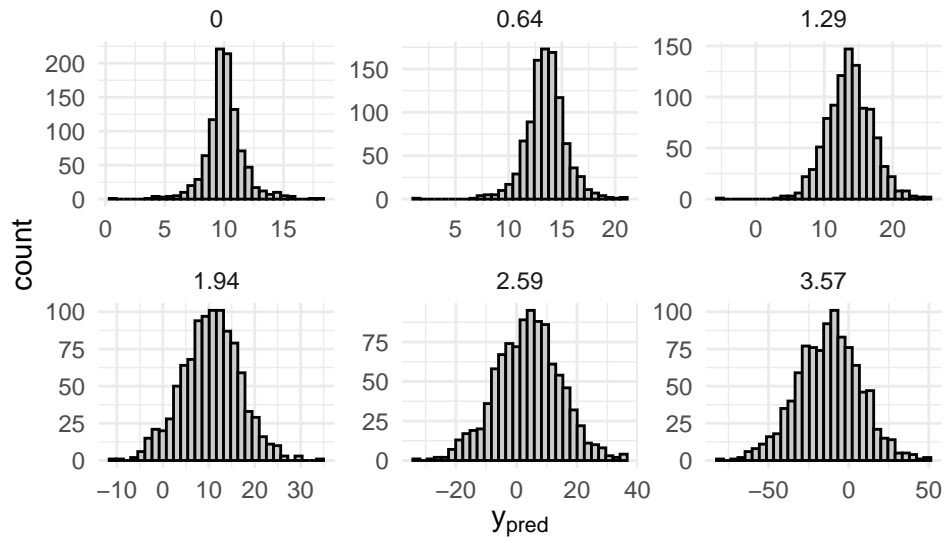


Figura 7: Distribución predictiva a priori en distintos valores del tiempo

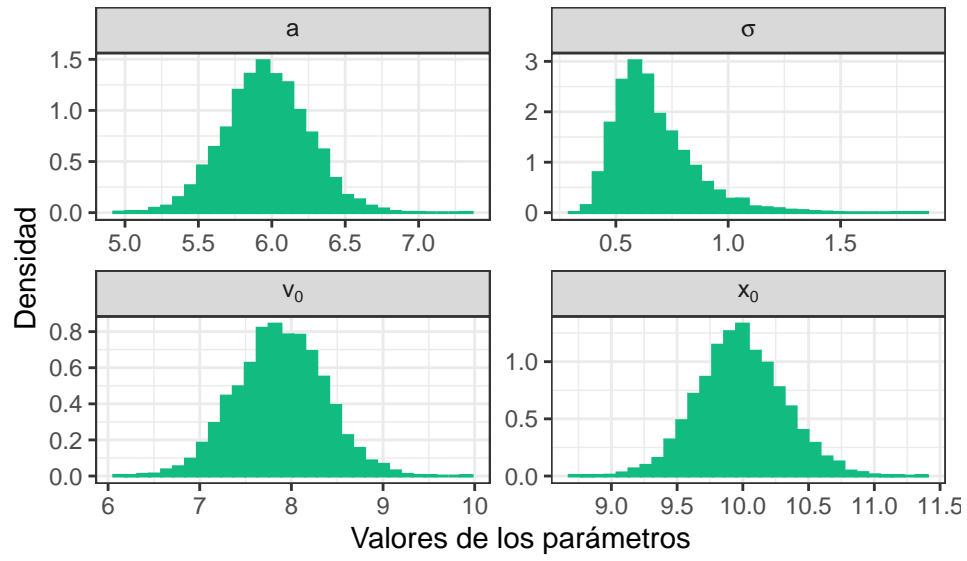


Figura 8: Distribución marginal de los parámetros estimados obtenidos a partir de las muestras generadas para el segundo modelo

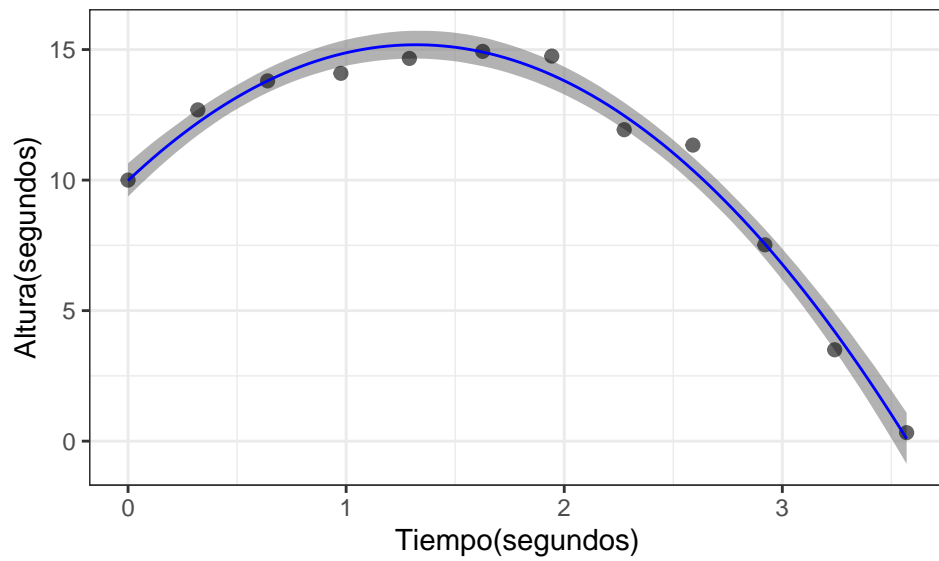


Figura 9: Curva de regresión estimada para  $\mu[i]$  en el modelo 2

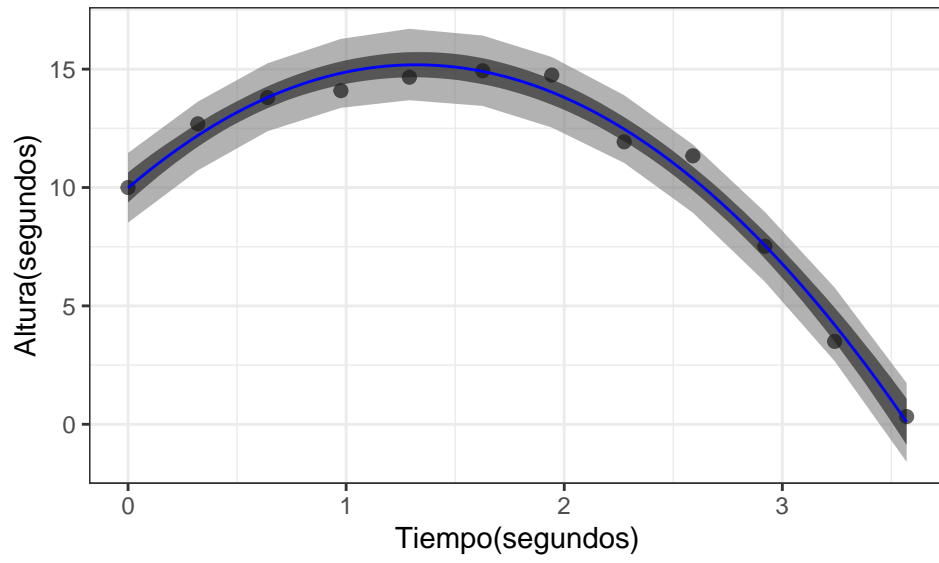


Figura 10: Distribución predictiva a posteriori en el modelo 2

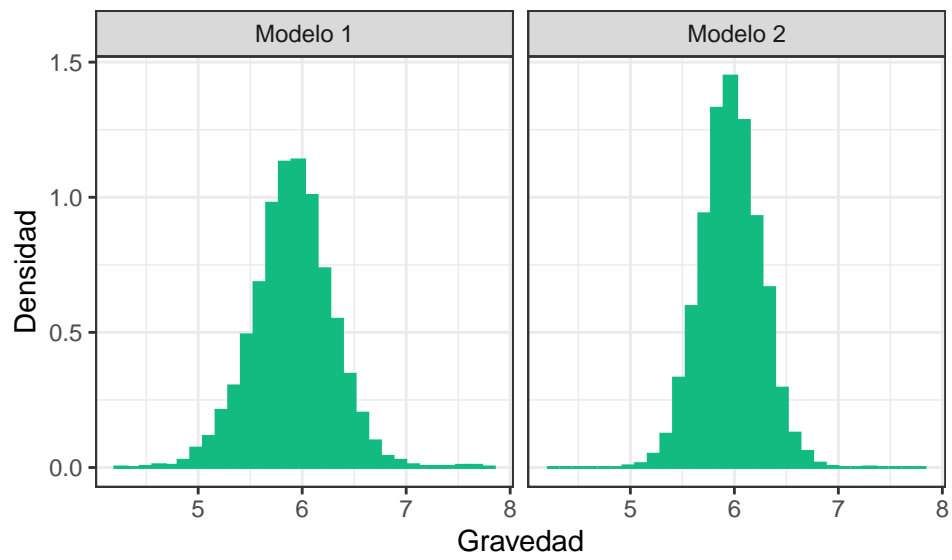


Figura 11: Distribución de la gravedad en ambos modelos

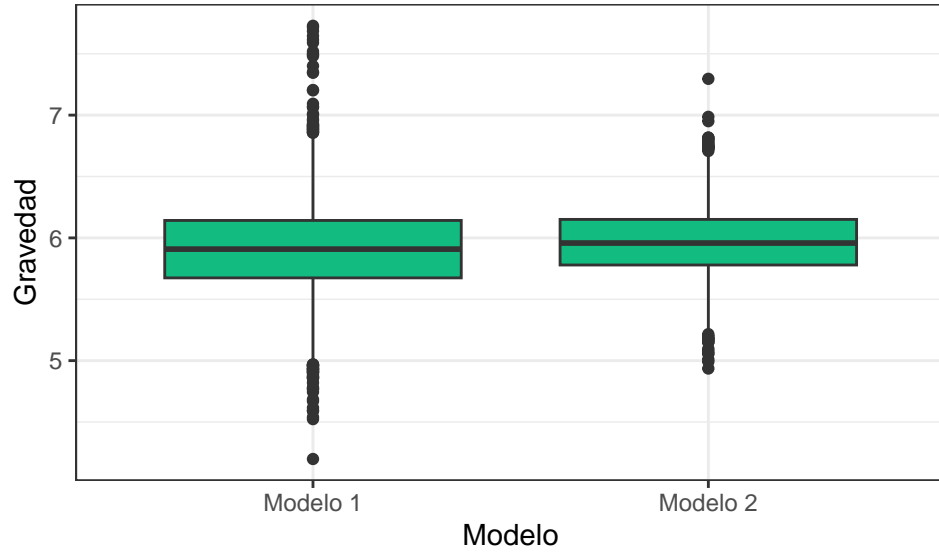


Figura 12: Distribución de la gravedad en ambos modelos

2.5%	25%	75%	97.5%
5.147896	5.674148	6.142111	6.635559

2.5%	25%	75%	97.5%
5.424308	5.779291	6.150503	6.511344

Queremos ver cuánto tarda en caer la pelota en promedio

$$\mu = 0 \rightarrow \beta_0 + \beta_1 t + \beta_2 t^2 = 0 \rightarrow t_{1,2} = \frac{-\beta_1 \pm \sqrt{\beta_1^2 - 4\beta_0\beta_2}}{2\beta_2}$$



