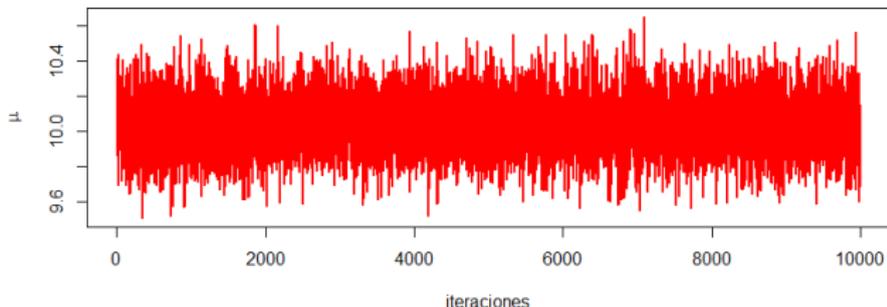


# El muestreo de Gibbs

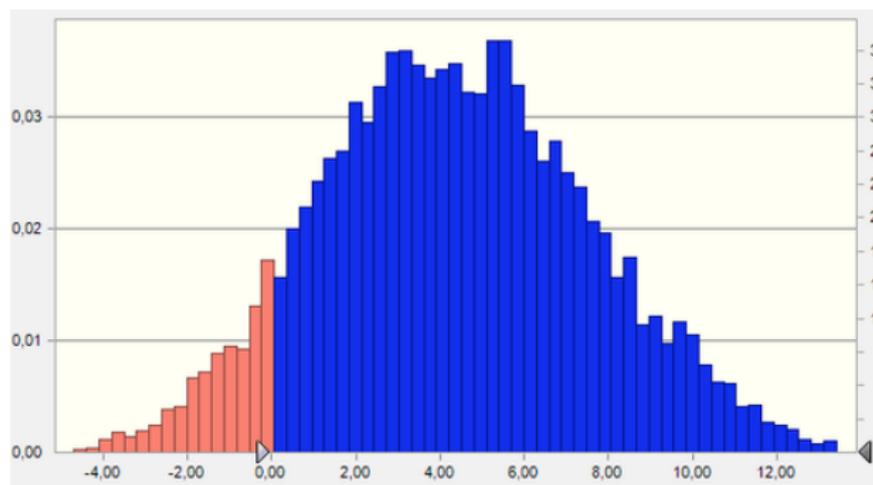


Mike Wiper

Departamento de Estadística  
Universidad Carlos III de Madrid

Grado en Estadística y Empresa

# Objetivo



Introducir el muestreo de Gibbs.

# Inferencia para una muestra normal

Hasta ahora, en el modelo  $Y \sim \text{Normal}\left(\mu, \frac{1}{\phi}\right)$ , hemos utilizado una distribución a priori (normal-gamma) conjugada:

$$\begin{aligned}\mu|\phi &\sim \text{Normal}\left(m, \frac{1}{c\phi}\right) \\ \phi &\sim \text{Gamma}\left(\frac{a}{2}, \frac{b}{2}\right).\end{aligned}$$

¿Hay algún rasgo extraño de esta distribución a priori?

# Inferencia para una muestra normal

Hasta ahora, en el modelo  $Y \sim \text{Normal}\left(\mu, \frac{1}{\phi}\right)$ , hemos utilizado una distribución a priori (normal-gamma) conjugada:

$$\begin{aligned}\mu|\phi &\sim \text{Normal}\left(m, \frac{1}{c\phi}\right) \\ \phi &\sim \text{Gamma}\left(\frac{a}{2}, \frac{b}{2}\right).\end{aligned}$$

¿Hay algún rasgo extraño de esta distribución a priori?

No es muy razonable suponer que las creencias sobre la media  $\mu$  dependen del valor de la precisión  $\phi$ .

# Una a priori más natural

Sería más razonable utilizar a prioris independientes para  $\mu$ ,  $\phi$ , por ejemplo:

$$\begin{aligned}\mu &\sim \text{Normal}\left(m, \frac{1}{d}\right) \\ \phi &\sim \text{Gamma}\left(\frac{a}{2}, \frac{b}{2}\right).\end{aligned}$$

En este caso, las creencias sobre la media no dependen de la precisión.

¿Podemos calcular la distribución a posteriori de manera analítica?

# ¿Podemos calcular la distribución a posteriori de manera analítica?

En este caso, no. Tenemos:

$$f(\mu, \phi | \text{datos}) \propto \phi^{\frac{a+n}{2}-1} \exp\left(-\frac{1}{2} [\phi (b + (n-1)s^2 + n(\mu - \bar{y})^2) + d(\mu - m)^2]\right)$$

y no podemos calcular la constante de integración o las distribuciones marginales fácilmente sin integración numérica.

¿Podemos calcular las distribuciones condicionales a posteriori?

# ¿Podemos calcular las distribuciones condicionales a posteriori?

Calcular las distribuciones condicionales es más fácil:

$$\begin{aligned}f(\phi|\mu, \text{datos}) &= \frac{f(\mu, \phi|\text{datos})}{f(\mu|\text{datos})} \\&\propto f(\mu, \phi|\text{datos}) \\&\propto \phi^{\frac{a+n}{2}-1} \exp\left(-\frac{1}{2} [\phi (b + (n-1)s^2 + n(\mu - \bar{y})^2) + d(\mu - m)^2]\right) \\&\propto \phi^{\frac{a+n}{2}-1} \exp\left(-\frac{\phi}{2} [b + (n-1)s^2 + n(\mu - \bar{y})^2]\right)\end{aligned}$$

que es una distribución gamma:  $\phi|\mu, \text{datos} \sim \text{Gamma}\left(\frac{a+n}{2}, \frac{b+(n-1)s^2+n(\mu-\bar{y})^2}{2}\right)$ .

# ¿Podemos calcular las distribuciones condicionales a posteriori?

De manera parecida,

$$\begin{aligned}f(\mu|\phi, \text{datos}) &\propto \exp\left(-\frac{1}{2} [\phi n(\mu - \bar{y})^2 + d(\mu - m)^2]\right) \\ &\propto \exp\left(-\frac{1}{2} [(\phi n + d)\mu^2 - 2(\phi n\bar{y} + dm)\mu]\right) \\ &\propto \exp\left(-\frac{\phi n + d}{2} \left[\mu - \frac{\phi n\bar{y} + dm}{\phi n + d}\right]^2\right)\end{aligned}$$

y entonces,  $\mu|\phi, \text{datos} \sim \text{Normal}\left(\frac{\phi n\bar{y} + dm}{\phi n + d}, \frac{1}{\phi n + d}\right)$ .

# Un algoritmo para generar una muestra (aproximada) de la distribución a posteriori

- 1 Sea  $t = 0$  y fijar un valor inicial para  $\mu$ , por ejemplo  $\mu^{(0)} = \bar{y}$ .
- 2 Generar  $\phi^{(t+1)} \sim \text{Gamma} \left( \frac{a+n}{2}, \frac{b+(n-1)s^2+n(\mu^{(t)}-\bar{y})^2}{2} \right)$
- 3 Generar  $\mu^{(t+1)} \sim \text{Normal} \left( \frac{\phi^{(t+1)}n\bar{y}+dm}{\phi^{(t+1)}n+d}, \frac{1}{\phi^{(t+1)}n+d} \right)$
- 4  $t = t + 1$ . Ir a 2.

A lo largo, las pares  $\mu^{(t)}, \phi^{(t)}$  se acercan a una muestra de la verdadera distribución conjunta a posteriori.

## Ejemplo

Volvemos a los datos de Illingworth que miramos hace dos semanas.

Esta vez, usamos distribuciones a priori independientes:  $\mu \sim \text{Normal}(0, 1000)$  y  $\phi \sim \text{Gamma}(0,25, 0,25)$ .

Corremos el algoritmo 10000 iteraciones.

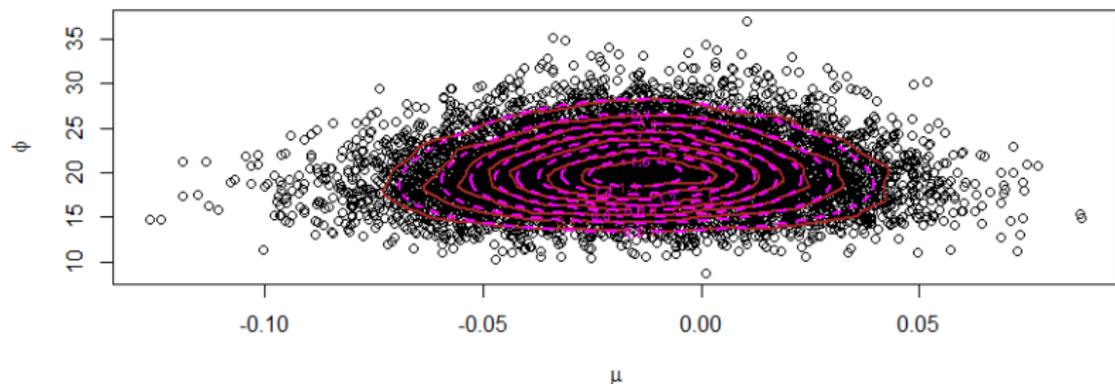
## Ejemplo

Volvemos a los datos de Illingworth que miramos hace dos semanas.

Esta vez, usamos distribuciones a priori independientes:  $\mu \sim \text{Normal}(0, 1000)$  y  $\phi \sim \text{Gamma}(0,25, 0,25)$ .

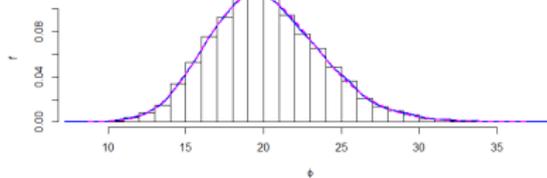
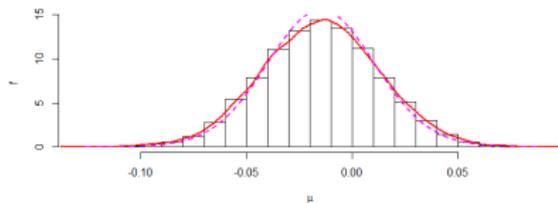
Corremos el algoritmo 10000 iteraciones.

Vemos los puntos generados y contornos (en marrón) de la distribución conjunta (con los contornos estimados por integración numérica en magenta).

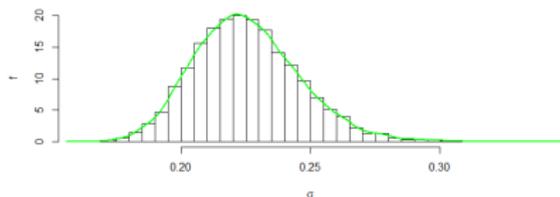


# Ejemplo

Las distribuciones marginales también se aproximan a las estimadas por integración numérica.



Además, es posible estimar la densidad de  $\sigma$  a través de los valores simulados:  
 $\sigma(t) = 1/\sqrt{\phi(t)}$ .

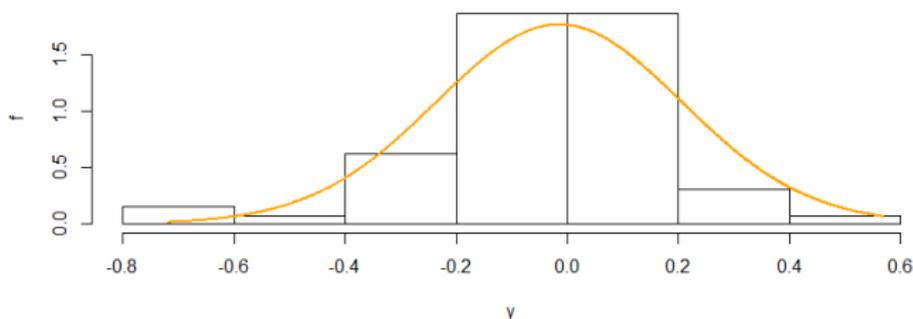


# Ejemplo

Podemos también estimar la distribución marginal de  $Y$  recordando que:

$$f(y|\text{datos}) = \int \int f(y|\mu, \phi) f(\mu, \phi|\text{datos}) \approx \frac{1}{N} \sum_{i=1}^N f\left(y|\mu^{(t)}, \frac{1}{\phi^{(t)}}\right).$$

Histograma de los datos de llingworth con la densidad ajustada



# ¿El algoritmo no es el Monte Carlo: verdad?

En el algoritmo Monte Carlo, todos los valores son generados de manera independiente.

En contraste, en este algoritmo:

- Cada valor generado depende del valor anterior.

# ¿El algoritmo no es el Monte Carlo: verdad?

En el algoritmo Monte Carlo, todos los valores son generados de manera independiente.

En contraste, en este algoritmo:

- Cada valor generado depende del valor anterior.
- Estamos generando de una *cadena de Markov*.

# ¿El algoritmo no es el Monte Carlo: verdad?

En el algoritmo Monte Carlo, todos los valores son generados de manera independiente.

En contraste, en este algoritmo:

- Cada valor generado depende del valor anterior.
- Estamos generando de una *cadena de Markov*.
- Se puede demostrar que la distribución del equilibrio de la cadena es la verdadera distribución a posteriori que buscamos.

# ¿El algoritmo no es el Monte Carlo: verdad?

En el algoritmo Monte Carlo, todos los valores son generados de manera independiente.

En contraste, en este algoritmo:

- Cada valor generado depende del valor anterior.
- Estamos generando de una *cadena de Markov*.
- Se puede demostrar que la distribución del equilibrio de la cadena es la verdadera distribución a posteriori que buscamos.
- El algoritmo es un ejemplo del *muestreo de Gibbs*.

# Cadenas de Markov

- Una cadena de Markov es una secuencia de variables  $X_1, \dots, X_k, \dots$  con la propiedad de Markov:

$$P(X_{k+1} = x | x_1, \dots, x_k) = P(X_{k+1} = x | x_k).$$

- La cadena es homogénea en el tiempo si

$$P(X_{k+1} = x | X_k = y) = P(X_k = x | X_{k-1} = y) \quad \forall x, y, k.$$

En este caso, la matriz de transición es  $P = p_{ij}$  donde los estados de la cadena son  $x_1, x_2, \dots$  y  $p_{ij} = P(X_{k+1} = x_j | X_k = x_i)$ .

- Un estado  $x$  de la cadena es periódico con periodo  $p$  si sólo es posible de volver a  $x$  en múltiplos de  $p$  pasos.
- Un estado  $x$  es transitorio si, empezando en  $x$ , existe una probabilidad positiva de que nunca se vuelva a  $x$ . Si no es el caso, el estado es recurrente.

# Cadenas de Markov

- Un estado recurrente es positivo recurrente si el tiempo medio que tarda en llegar al estado otra vez es finito.
- La cadena es irreducible si es posible llegar a cualquier estado dado un estado inicial cualquiera.
- Para una cadena irreducible con todos sus estados positivo recurrentes y aperiodicos, existe una distribución del equilibrio,  $\pi$  satisfaciendo

$$\pi = \pi P.$$

- En cadenas con estado de espacios infinitos existe una definición parecida: la cadena tiene una distribución de equilibrio si es una *cadena de Harris*.

# Muestreo de Gibbs

- Queremos muestrear de la distribución  $f(Y)$  donde  $Y = Y_1, \dots, Y_k$ .
- Supongamos que todas las distribuciones condicionales  $f(Y_i|Y_{-i})$  son fáciles de muestrear.
- El algoritmo Gibbs fija un valor inicial  $y^{(0)}$  y después muestra valores sucesivamente de las distribuciones condicionales.
- Es una cadena de Markov y tiene distribución del equilibrio igual a  $f(Y)$ .

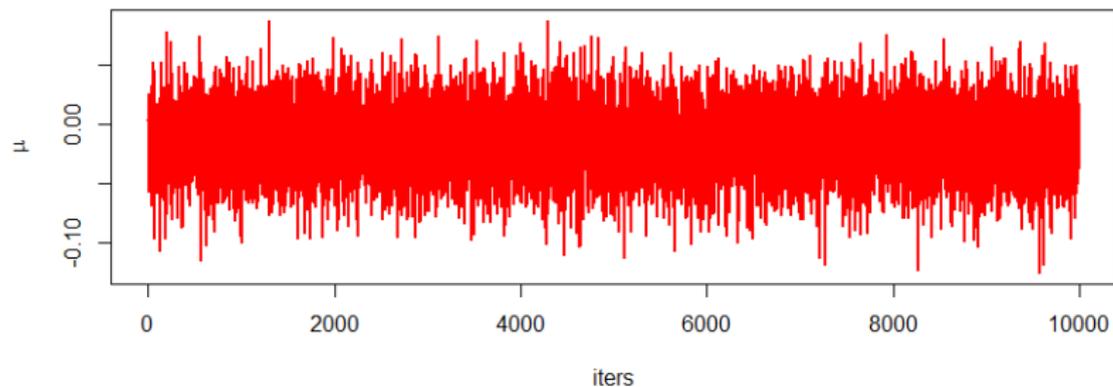
# ¿Cómo sabemos si estamos en el equilibrio?

Los valores generados dependerán del valor inicial utilizado y de cómo de rápido converge. Usamos varios métodos para mirar la convergencia:

- Correr múltiples cadenas con valores iniciales distintos diferentes.
- Mirar los valores generados para ver que se mezcla bien la cadena. [▶ Ejemplo](#)
- Calcular medias de los valores generadas y mirar su convergencia de manera visual. [▶ Ejemplo](#)
- Mirar las autocorrelaciones de los valores generados. [▶ Ejemplo](#)
- Pruebas formales de convergencia.

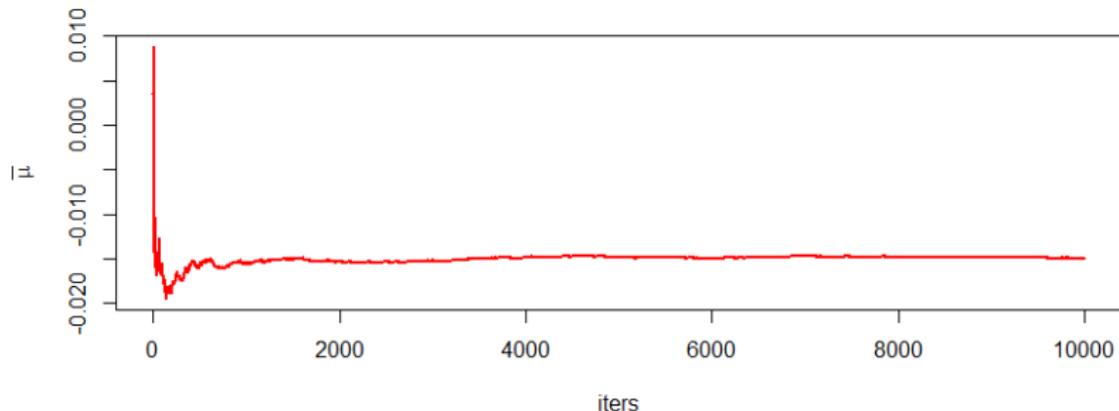
# Ejemplo

Vemos que en este caso, la cadena mezcla bien y no se queda atrapada.



# Ejemplo

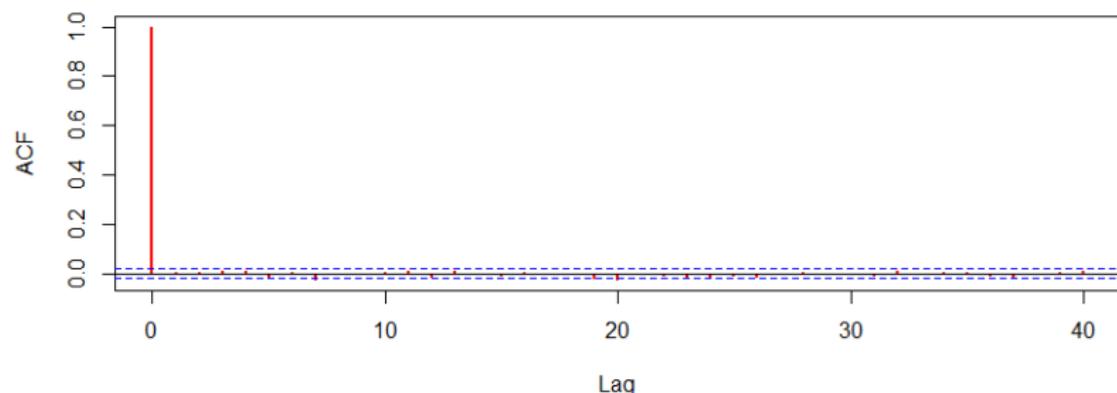
La estimación de la media converge tras aproximadamente 1000 iteraciones.



Sugiere que podemos tirar los primeros mil iteraciones y usar los siguientes para la inferencia.

# Ejemplo

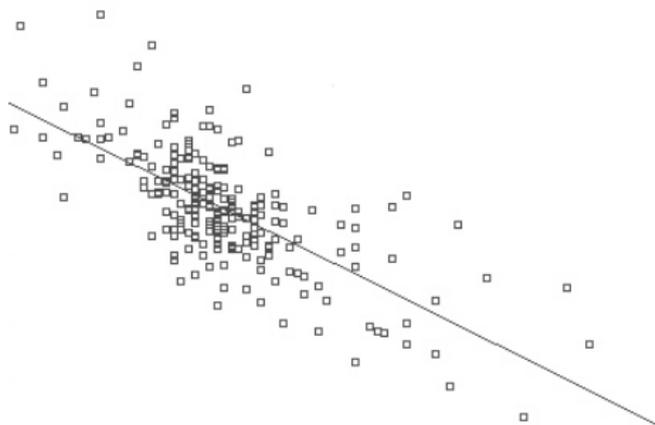
En este caso, casi no existe autocorrelación.



Si los datos son autocorrelados, podemos utilizar “thinning” para tener una muestra aproximadamente independiente.

# Resumen y siguiente sesión

En esta sesión, hemos examinado como utilizar el muestreo de Gibbs para implementar la inferencia cuando se utiliza una distribución a priori semi-conjugada en el modelo normal.



En la siguiente sesión, empezamos a explorar el enfoque bayesiano a problemas de regresión y otros modelos lineales.