# Trabajo Práctico


Barraza, Veronica y Maldonado, Kevin


Junio 2022



# Markov Chain Monte Carlo: el método de Metrópolis-Hastings

## Planteo del problema

Vamos a plantear dos problemas de estimación Bayesiana:

1. Se quiere saber cuál es la proporción $p$ de fumadores en la Capital Federal. Se asume una distribuión a priori $Beta(3,5)$ para $p$. Se encuestan a 300 personas, y resulta que exactamente 50 de ellas fuman. Con esta información, se quiere saber el estimador Bayes para $p$.
2. Supongamos que para un determinado genotipo, la probabilidad del alelo $a$ es $p$ y la del alelo $A$ es $1-p$. Si se asume que la población se reproduce aleatoriamente, esto implica que la probabilidad de los genotipos $aa$, $aA$ y $AA$ son $p2, 2p(1-p)y(1-p)^2$ respectivamente. Se asume una distribución a priori $U(0,1)$ para $p$. Se observan 13 personas con el genotipo $aa$, 210 con el genotipo $aA$ y 240 con el genotipo $AA$. Con esta información, se quiere saber el estimador Bayes para $p$.

# Primer Problema

a- Encontrar analíticamente la distribución a posteriori para el problema 1.

Sea $X = x_1, \dots, x_N$ una muestra aleatoria, con $X_i$ variable con distribución bernoulli y una distribución beta a priori para el parámetro.

$$x \sim Bern(\theta),$$

$$\theta \sim beta(\alpha, \beta),$$

$$\theta \in [0,1].$$

**Posterior**

$$P(\theta|X=x) \propto \prod_{i=1}^{N} p(x_n|\theta)p(\theta)$$

$$P(\theta|X=x) \propto \left(\prod_{i=1}^{N} \theta^{x_n}(1-\theta)^{(1-x_n)}\right) \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{(\beta-1)}\mathbb{1}_{(0,1)}(\theta)$$

$$P(\theta|X=x) \propto \theta^{\sum_n x_n+\alpha-1}(1-\theta)^{N-\sum_n x_n+\beta-1}$$

Recordando la Distribución **beta**:

$$\mathrm{B}(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

Por lo tanto vemos que el posterior es proporional a la distribución beta:

$$\theta|X \sim B(\alpha_N, \beta_N)$$

$$\alpha_N = \sum_{n=1}^{N} x_n + \alpha$$

$$\beta_N = N - \sum_{n=1}^{N} x_n + \beta$$

Por lo tanto, podemos identificar esto como la distribución Beta con nuevos parámetros. Es decir, la distribución posterior vuelve a ser una distribución Beta (igual que la anterior). Esto significa que Beta prior es un Conjugate Prior para $\theta$ en el modelo de muestreo bernoulli.

A continuación vamos a gráficar la distribución a priori:

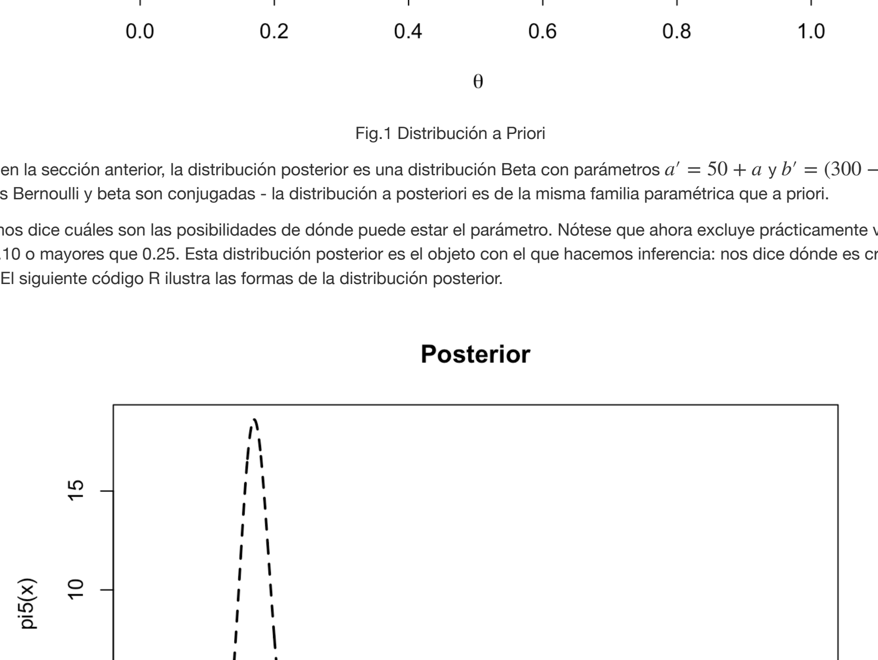


Fig.1 Distribución a Priori

Como vimos en la sección anterior, la distribución posterior es una distribución Beta con parámetros $a' = 50 + a$ y $b' = (300-50) + b$. Distribuciones Bernoulli y beta son conjugadas - la distribución a posteriori es de la misma familia paramétrica que a priori.

La posterior nos dice cuáles son las posibilidades de dónde puede estar el parámetro. Nótese que ahora excluye prácticamente valores más chicos que 0.10 o mayores que 0.25. Esta distribución posterior es el objeto con el que hacemos inferencia: nos dice dónde es creíble que esté el parámetro.El siguiente código R ilustra las formas de la distribución posterior.

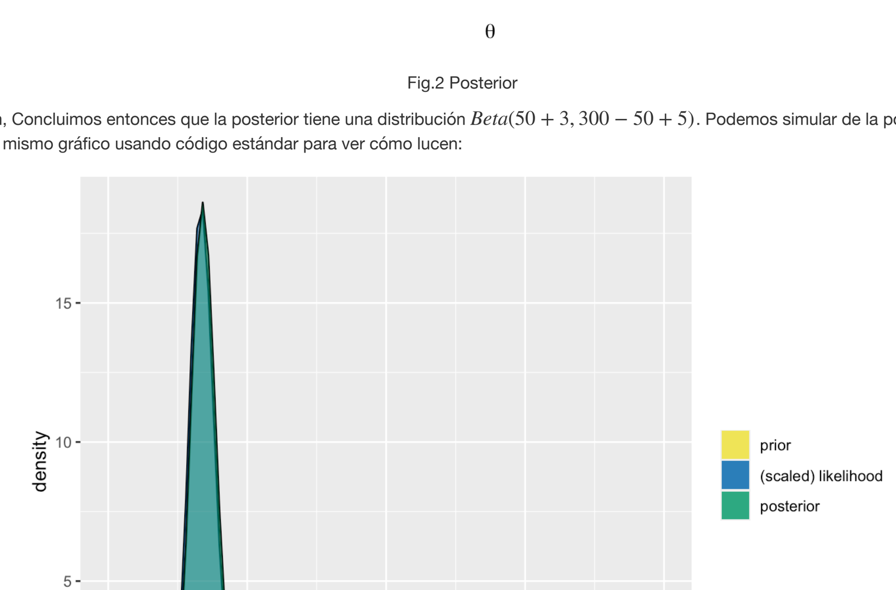


Fig.2 Posterior

En resumen, Concluimos entonces que la posterior tiene una distribución $Beta(50+3, 300-50+5)$. Podemos simular de la posterior y a priori en un mismo gráfico usando código estándar para ver cómo lucen:

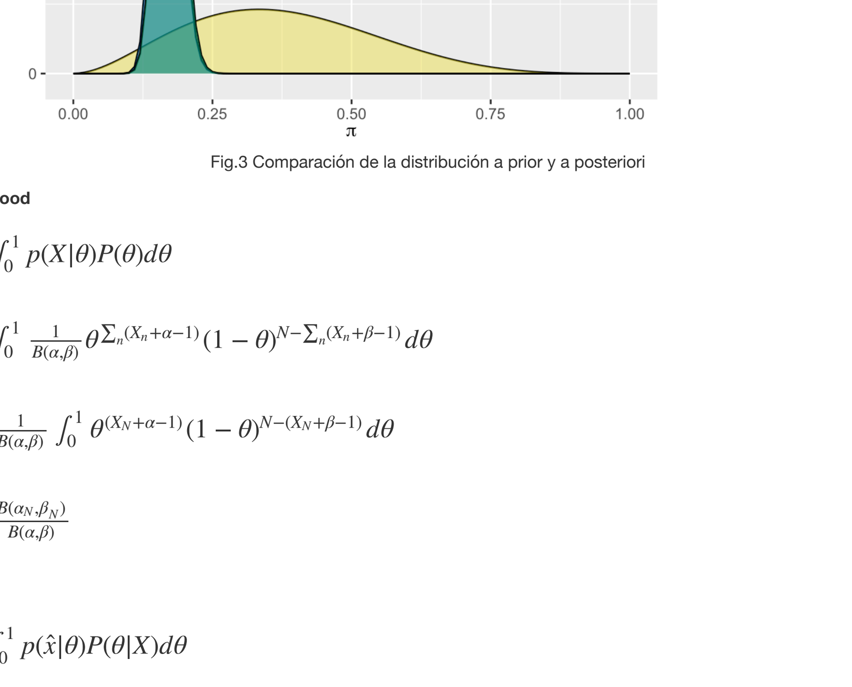


Fig.3 Comparación de la distribución a prior y a posteriori

**Marginal likelihood**

$$p(X) = \int_0^1 p(X|\theta)P(\theta)d\theta$$

$$p(X) = \int_0^1 \frac{1}{B(\alpha,\beta)}\theta^{\sum_n(X_n+\alpha-1)}(1-\theta)^{N-\sum_n(X_n+\beta-1)}d\theta$$

$$p(X) = \frac{1}{B(\alpha,\beta)}\int_0^1 \theta^{(X_N+\alpha-1)}(1-\theta)^{N-(X_N+\beta-1)}d\theta$$

$$p(X) = \frac{B(\alpha_N,\beta_N)}{B(\alpha,\beta)}$$

**Posterior**

$$p(\hat{x}) = \int_0^1 p(\hat{x}|\theta)P(\theta|X)d\theta$$

$$p(\hat{x}) = \frac{1}{B(\alpha_N,\beta_N)}\int_0^1 \theta^{(\hat{x}_N+\alpha-1)}(1-\theta)^{N-(\hat{x}_N+\beta-1)}d\theta$$

$$p(\hat{x}) = \frac{B(\hat{x}+\alpha_N, 1-\hat{x}+\beta_N)}{B(\alpha_N,\beta_N)}$$

Como tenemos la forma analítica de la posterior, es posible hacer los cálculos de la media posterior, por ejemplo, integrando la densidad posterior a mano. Esto generalmente no es factible, y en este ejemplo preferimos hacer una aproximación numérica. En este caso particular es posible usando cálculo, y sabemos que la media de una $Beta(\alpha,\beta)$ es $\alpha/(\alpha+\beta)$ , de modo que nuestra media posterior es $0.17$

| model | alpha | beta | mean | mode | var | sd |
|---|---|---|---|---|---|---|
| <chr> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> |
| prior | 3 | 5 | 0.3750000 | 0.3333333 | 0.0260416667 | 0.16137431 |
| posterior | 53 | 255 | 0.1720779 | 0.1699346 | 0.0004610586 | 0.02147228 |

2 rows

b- Aproximar la distribución a posteriori implementando el método de Metropolis-Hastings para este problema (elegir $p_{init} = 0.5$ y $\sigma^2 = 0.01$, $N = 50000$ y el M que parezca adecuado,mirando cuando se estabiliza la secuencia.)

A continuación se muestra la implementación del método Metropolis-Hastings para este problema:

```
likBinom<-function(p,s,n){
  lik<-dbinom(s,size=n,prob = p)
  return(lik)
}
priorBeta<-function(p,a,b){
  prob<-dbeta(p,shape1=a,shape2 = b)
  return(prob)
}
proDist<-function(currentP,sd){
  p<-rnorm(n=1,mean=currentP,sd=sd)
  if(p>0 & p<1) return(p)
  proDist(currentP,sd)
}
nIter<-50000
mySD<-sqrt(0.01);
a<-3;b<-5
p<-vector()
p[1]<-0.5
for(i in 1:nIter){
  proposedP<-proDist(p[i],mySD)
  r<-min(1, ( likBinom(proposedP,50,300)*priorBeta(proposedP,a,b) ) / ( likBinom(p[i],50,300) *priorBeta(p[i],a,b
) ) )
  if(runif(1)<r) {
    p[i+1]<-proposedP
  } else {
    p[i+1]<-p[i]
  }
}
```

```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

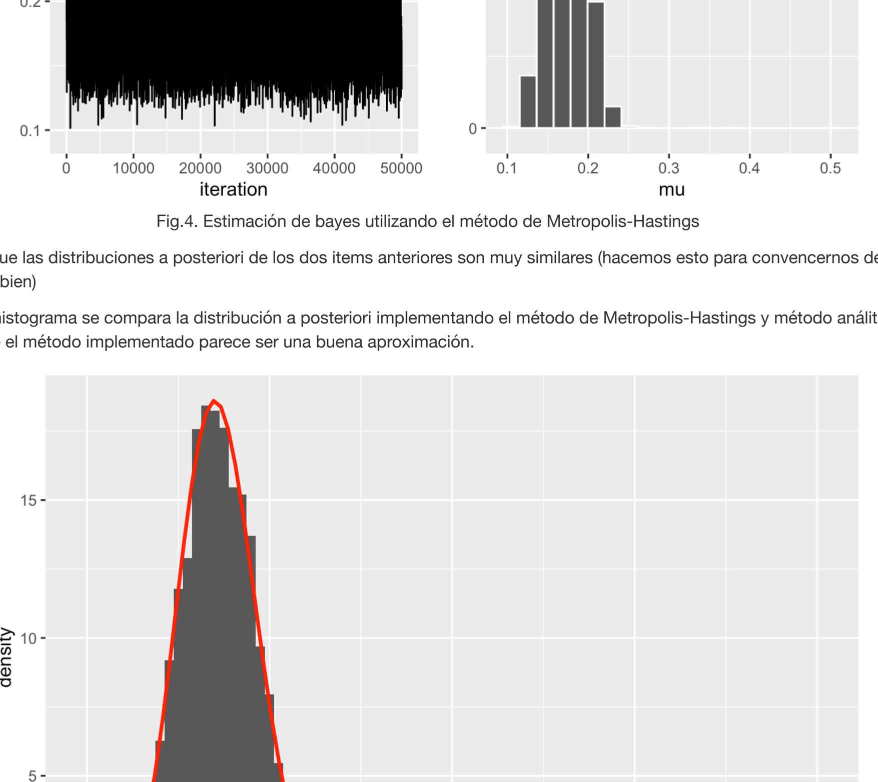


Fig.4. Estimación de bayes utilizando el método de Metropolis-Hastings

c- Comprobar que las distribuciones a posteriori de los dos items anteriores son muy similares (hacemos esto para convencernos de que este m ́etodo funciona bien)

En el siguiente histograma se compara la distribución a posteriori implementando el método de Metropolis-Hastings y método análitico, observando que el método implementado parece ser una buena aproximación.

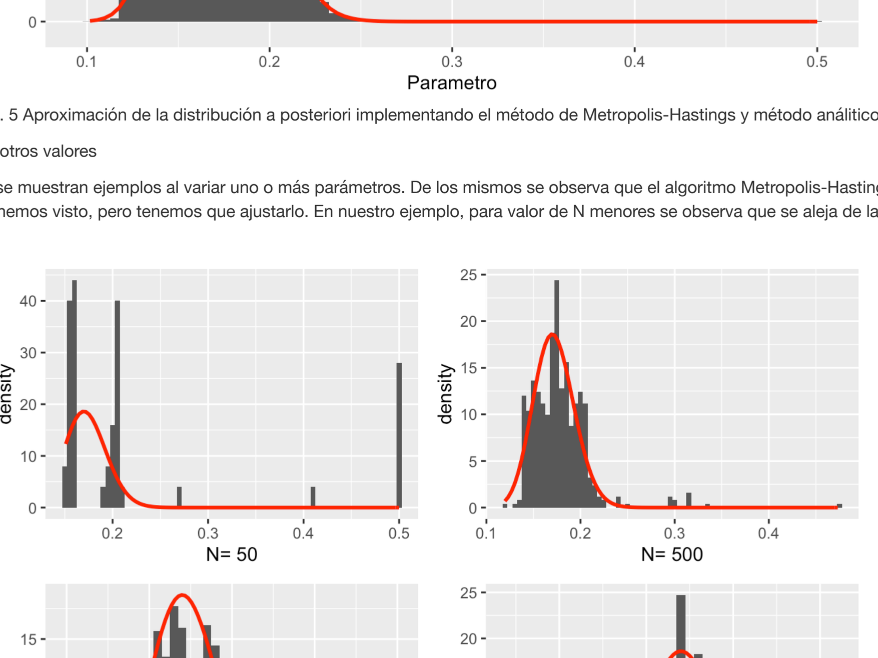


Fig. 5 Aproximación de la distribución a posteriori implementando el método de Metropolis-Hastings y método análitico

d- Explorar con otros valores

A continuación se muestran ejemplos al variar uno o más parámetros. De los mismos se observa que el algoritmo Metropolis-Hastings puede funcionar, ya lo hemos visto, pero tenemos que ajustarlo. En nuestro ejemplo, para valor de N menores se observa que se aleja de la distribución teorica.

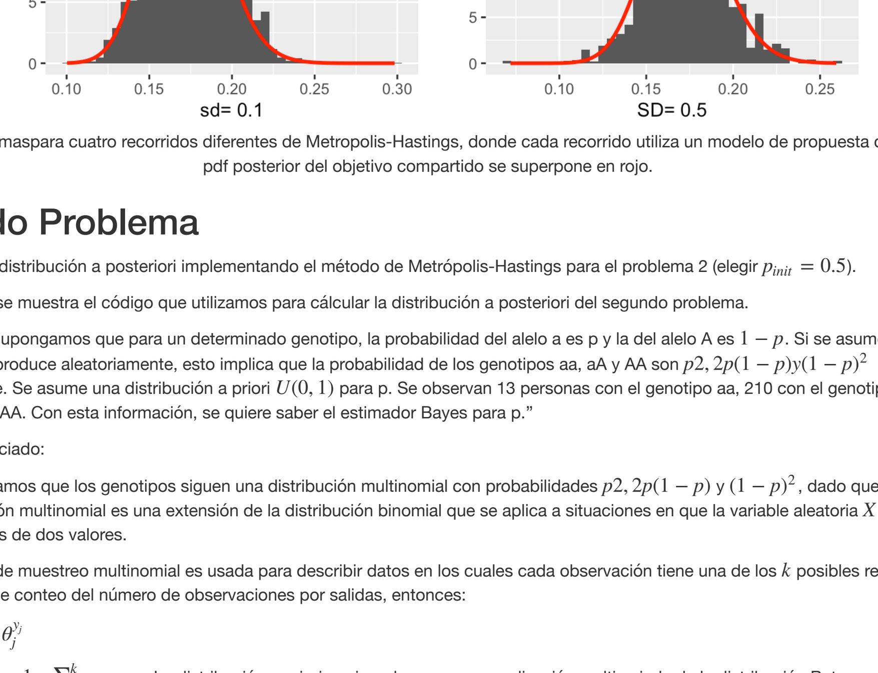


Fig.6 Histogramaspara cuatro recorridos diferentes de Metropolis-Hastings, donde cada recorrido utiliza un modelo de propuesta diferente. El pdf posterior del objetivo compartido se superpone en rojo.

# Segundo Problema

2- Aproximar la distribución a posteriori implementando el método de Metrópolis-Hastings para el problema 2 (elegir $p_{init} = 0.5$).

A continuación se muestra el código que utilizamos para cálcular la distribución a posteriori del segundo problema.

Recordando: ‘’Supongamos que para un determinado genotipo, la probabilidad del alelo a es p y la del alelo A es $1-p$. Si se asume que la población se reproduce aleatoriamente, esto implica que la probabilidad de los genotipos aa, aA y AA son $p2, 2p(1-p)y(1-p)^2$ respectivamente. Se asume una distribución a priori $U(0,1)$ para p. Se observan 13 personas con el genotipo aa, 210 con el genotipo aA y 240 con el genotipo AA. Con esta información, se quiere saber el estimador Bayes para p.”

En base al enunciado:

- Consideramos que los genotipos siguen una distribución multinomial con probabilidades $p2, 2p(1-p)$ y $(1-p)^2$, dado que La distribución multinomial es una extensión de la distribución binomial que se aplica a situaciones en que la variable aleatoria $X$ puede tomar más de dos valores.

La distribución de muestreo multinomial es usada para describir datos en los cuales cada observación tiene una de los $k$ posibles resultados. Si $y$ es un vector de conteo del número de observaciones por salidas, entonces:

$P(y|\theta) \propto \prod_{j=1}^{k} \theta_j^{y_j}$

donde: $\sum_{j=1}^{k} \theta_j = 1$ y $\sum_{j=1}^{k} y_j = n$ La distribución a priori conjugada es una generalización multivariada de la distribución Beta conocida como la distribución de **Dirichlet**, dada por:

$P(\theta|\alpha) = \prod_{j=1}^{k} \theta_j^{\alpha_j-1}$

donde la distribución es restringida a $\theta_j$ no negativos con $\sum_{j=1}^{k} \theta_j = 1$. La distribución a posteriori para los $\theta_j$ es nuevamente un proceso de Dirichlet con parámetros $\alpha_j + y_j$. Entonces la distribución a posterior es:

$\theta|y \sim Dirichlet(\alpha_j + y_j)$

```
currentValue = 0.5
N = 50000
stepSd = sqrt(0.01)
eps = 1e-9
results = rep(0, N)
for (rep in 1:N) {
  step = rnorm(n=1, mean=0, sd=stepSd)
  candidate = currentValue + step
  if (candidate < eps || candidate > 1-eps) {
    acceptance = 0
  } else {
    g_candidate = punif(candidate)
    g_current = punif(currentValue)
    likelihoodCandidate = dmultinom(c(13, 210, 240), prob=c(candidate*candidate,
                                                       2*candidate*(1-candidate),
                                                       (1-candidate)^2))
    likelihoodCurrent = dmultinom(c(13, 210, 240), prob=c(currentValue*currentValue,
                                                       2*currentValue*(1-currentValue),
                                                       (1-currentValue)^2))
    acceptance = min(1, (g_candidate*likelihoodCandidate)/(g_current*likelihoodCurrent))
  }
  coin = rbinom(n=1, size=1, prob=acceptance)
  if (coin > 0.5) {
    currentValue = candidate
  }
  results[rep] = currentValue
}
```

Este problema muestra que solo necesitamos MCMC para aproximar un posterior bayesiano cuando ese posterior es demasiado complicado de especificar. YAquí es donde entra en juego el algoritmo MCMC más general de Metropolis-Hastings. Metropolis-Hastings se basa en el hecho de que, incluso si no conocemos el modelo posterior, sí sabemos que la función de probabilidad posterior es proporcional al producto del anterior conocido pdf y funcion de verosimilitud.

b- Aproximar el estimador Bayes en este problema.

A continuación se muestra la distribución del parámetro estimado utilizando el método Metrópolis-Hastings. El mismo tiene un valor medio de 0.25.

```
## [1] 0.2564306
```

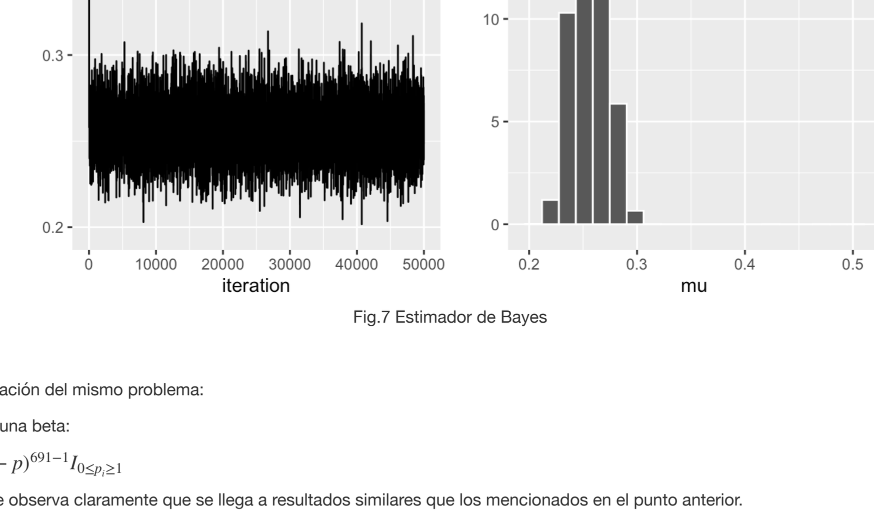


Fig.7 Estimador de Bayes

**Extra**

Otra aproximación del mismo problema:

Posteriori es una beta:

$p^{231-1} * (1-p)^{691-1} I_{0 \leq p_i \geq 1}$

Del gráfico se observa claramente que se llega a resultados similares que los mencionados en el punto anterior.

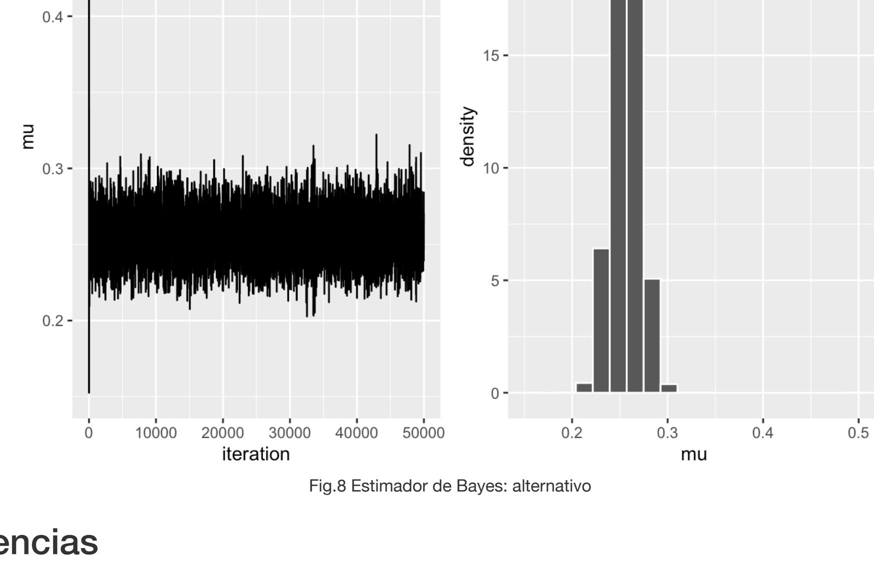


Fig.8 Estimador de Bayes: alternativo

# Referencias


An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics) libro

https://psirusteam.github.io/bookdownBayesiano/modelo-multinomial.html

Brooks, Steve, Andrew Gelman, Galin Jones, and Xiao-Li Meng. 2011. Handbook of Markov Chain Monte Carlo. CRC Press.

McElreath, Richard. 2019. “Statistical Rethinking Winter 2019 Lecture 12.” Youtube; https://www.youtube.com/watch?v=hRJtKCIDTwc.