Tema 2: Modelos Básicos de Estadística Bayesiana

Introducción

Desde el punto de vista de la estadística bayesiana se considera que hay dos tipos de valores: *conocidos* y *desconocidos*. El objetivo es usar las cantidades o datos conocidos mediante un modelo paramétrico dado para realizar inferencias sobre las cantidades desconocidas. Por cantidades desconocidas se puede entender tanto parámetros del modelo, como observaciones *missing* o predicciones.

En un modelo básico se tiene un parámetro *de interés* θ y unos datos observados D y se considera una distribución de probabilidad conjunta para ambos que recoge cómo se relacionan: $p(\theta, D)$.

Aplicando la definición de probabilidad condicionada, se tiene que

$$p(\theta, D) = p(\theta) \cdot p(D|\theta)$$

$p(\theta)$ se denomina distribución *a priori* de θ. El segundo término se denomina *función de verosimilitud*.

En realidad, estamos interesados en determinar $p(\theta|D)$ que es la distribución de los parámetros desconocidos dadas las observaciones conocidas. Como $p(\theta, D) = p(D, \theta)$, entonces

$$p(\theta) \cdot p(D|\theta) = p(D) \cdot p(\theta|D)$$

de modo que

$$p(\theta|D) = \frac{p(\theta) \cdot p(D|\theta)}{p(D)}.$$

Esto se puede expresar como

$$\pi(\theta|D) = \frac{p(\theta) \cdot L(\theta|D)}{\int_\theta p(\theta) \cdot L(\theta|D) d\theta}$$

donde $L(\theta|D)$ denota la verosimilitud y

$$p(D) = \int_\theta p(\theta) \cdot L(\theta|D) d\theta$$

es la constante normalizadora o la *distribución predictiva a priori*.

De manera alternativa, como la integral anterior no depende de θ, se puede expresar como

$$\pi(\theta|D) \propto p(\theta) \cdot L(\theta|D)$$

es decir,

Probabilidad a posteriori $\propto$ Probabilidad a priori $\times$ Verosimilitud.

El problema de inferencia

Se parte una muestra de datos, $\mathbf{x} = (x_1, \ldots, x_n)$, de modo que la variable aleatoria X que los genera se asume que depende de unos parámetros θ y que tiene como función de probabilidad f :

$$X|\theta \sim f(\cdot)$$

Se trata de realizar inferencias sobre θ, es decir, estimar su valor.

La **Inferencia Clásica** presenta las siguientes características:

- El concepto de probabilidad está limitado a aquellos sucesos en los que se pueden definir frecuencias relativas.
- θ es un valor fijo (pero desconocido).
- Se pueden usar estimadores de máxima verosimilitud (que se justifican asintóticamente) o se pueden usar estimadores insesgados.
- Se usa el concepto de *intervalo de confianza*:

  Ejemplo: Si $(1, 3)$ es un intervalo de confianza al 95 %, significa que si repetimos el procedimiento muchas veces y calculamos un intervalo de confianza cada vez, el 95 % de los intervalos incluirán θ que es el verdadero valor del parámetro. Esto **no** significa que la probabilidad de que θ esté dentro del intervalo sea igual a 0.95.

  La definición es una consecuencia del punto de vista de la probabilidad como frecuencia relativa.
- El método de muestreo es muy importante.

La **Inferencia Bayesiana** se caracteriza por:

- Cada persona tiene sus propias creencias subjetivas previas (*a priori*) para cualquier suceso: $P(\text{cruz})$, $P(\text{lloverá mañana})$,...

  Nuestras probabilidades pueden ser diferentes porque son nuestras propias medidas de incertidumbre. La única restricción es que han de ser *coherentes* (que cumplan los axiomas de la probabilidad de Kolmogorov).

- θ es un variable y tiene una distribución $f(\theta)$. Modificamos nuestras creencias sobre θ usando el teorema de Bayes.

- La estimación es un problema de decisión. En situaciones distintas se eligen estimadores diferentes: se usa la *teoría de utilidad* para elegirlos.

- Un intervalo de credibilidad del 95 % para θ es un intervalo tal que la **probabilidad** de que contenga a θ igual a 0,95.

Comentarios y críticas respecto a los métodos bayesianos

- *Crítica*: θ no está claro que tenga que ser siempre una variable.

  Argumento: en la práctica, la distribución $f(\theta)$ muestra los conocimientos acerca de θ. El conocimiento cambia y evoluciona con la recogida de los datos.

- *Crítica*: Falta de objetividad. ¿Cómo se puede elegir la distribución a priori cuando no se tiene información?

  Argumento: se pueden establecer métodos *objetivos* para calcular la distribución a priori. A menudo, un análisis clásico equivale a un análisis bayesiano con una distribución a priori no informativa.

- Los aspectos subjetivos son explícitos en un análisis bayesiano, mientras que en los métodos clasicos *existen* aunque no se muestran claramente.

- Es imprescindible hacer un *análisis de sensibilidad*. Si los resultados cambian cuando se cambia la distribución inicial, entonces la verosimilitud no da mucha información y la elección de la distribución inicial es fundamental.

Principios fundamentales en Inferencia

El Principio de Verosimilitud

Se trata de determinar estimadores de θ. Si se observan los datos $\mathbf{x}$, toda la información necesaria para la inferencia está contenida en la función de verosimilitud $L(\theta|\mathbf{x})$. Además, dos funciones de verosimilitud tienen la misma información sobre θ si son proporcionales entre sí.

Los métodos bayesianos cumplen el principio de verosimilitud: Si $L_1(\theta|\mathbf{x}) \propto L_2(\theta|\mathbf{x})$ entonces, dada una distribución a priori $\pi(\theta)$,

$$f_1(\theta|\mathbf{x}) \propto \pi(\theta)L_1(\theta|\mathbf{x}) \propto \pi(\theta)L_2(\theta|\mathbf{x}) \propto f_2(\theta|\mathbf{x})$$

No obstante, se puede demostrar que los contrastes clásicos con significación a nivel fijo (por ejemplo $\alpha = 0{,}05$) y los intervalos de confianza no cumplen con este principio.

El Principio de Suficiencia

Un estadístico $\mathbf{t}$ es suficiente para θ si

$$L(\theta|\mathbf{x}) = h(\mathbf{t}, \theta)g(\mathbf{x}).$$

que recibe el nombre de teorema de *factorización de Neyman.*

El *Principio de Suficiencia* afirma que si existe un estadístico suficiente, $\mathbf{t}$, dadas dos muestras, $\mathbf{x}_1$, $\mathbf{x}_2$, que cumplen $\mathbf{t}(\mathbf{x}_1) = \mathbf{t}(\mathbf{x}_2)$, las conclusiones basadas en $\mathbf{x}_1$ y $\mathbf{x}_2$ deben ser iguales.

Todos los métodos estándar de inferencia estadística cumplen el principio de suficiencia.

El Principio de Condicionalidad

Suponiendo que se pueden hacer dos posibles experimentos E_1 y E_2 en relación a θ, y se elige uno con probabilidad 0,5, entonces la inferencia sobre θ depende sólo del resultado del experimento seleccionado. La inferencia bayesiana cumple el principio, pero la clásica no.

El principio de verosimilitud equivale al principio de suficiencia más el principio de condicionalidad (ver demostración en Lee (2012)).

Modelo Beta-Binomial

Este esquema se va a aplicar a problemas relativos a *proporciones*. Nos interesarán, por tanto, experimentos con dos posibles resultados, uno de los cuales se designará *éxito* y otro *fracaso*. Existen muchos ejemplos de este tipo de experimentos, por ejemplo:

- Estamos diseñando un sistema de control de una línea de producción y consideramos que los productos pueden ser defectuosos o no.
- Al diseñar un sistema de decisión inteligente, por ejemplo en Medicina, podemos considerar que los pacientes pueden tener o no una enfermedad.
- Para diseñar una política de mantenimiento de un mecanismo de control registramos si el mecanismo indica situación de peligro o no.
- Estamos interesados en evaluar la efectividad de un programa de educación para la infancia, de modo que consideramos la proporción de niños y niñas que han mejorado su rendimiento.

Nos interesa obtener información acerca de una proporción determinada, por ejemplo, sus valores típicos o si es menor que un valor prefijado. También se puede considerar el problema de comparar dos proporciones: si cierta proporción es mayor que otra, o si la diferencia entre dos proporciones es menor que 0.2, por ejemplo.

La situación que consideramos es la siguiente: se desea recoger y proporcionar información sobre p, la proporción de casos en que se produce cierto fenómeno, pudiéndose dar sólo dos resultados. Disponemos de unas creencias iniciales sobre p, que puede tomar valores entre 0 y 1.

Se tiene que determinar una distribución suficientemente flexible que modelice estas creencias sobre proporciones. Una posible distribución es la distribución beta.

NOTA:

Una v.a. tiene distribución beta de parámetros α, β en $[0,1]$ (y se representa como $X \sim Be(\alpha, \beta)$), con $\alpha, \beta > 0$ si su función de densidad es

$$f(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} I_{0<x<1}$$

Recuerda que la *función gamma* se define como

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

y sus propiedades básicas son

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$$
$$\Gamma(n+1) = n!$$

para $n \in \mathbb{N}$

Los momentos de la distribución son

$$\mu = \int_0^1 x \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\,dx =$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \int_0^1 \frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+1)\Gamma(\beta)} x^{\alpha}(1-x)^{\beta-1}\,dx =$$

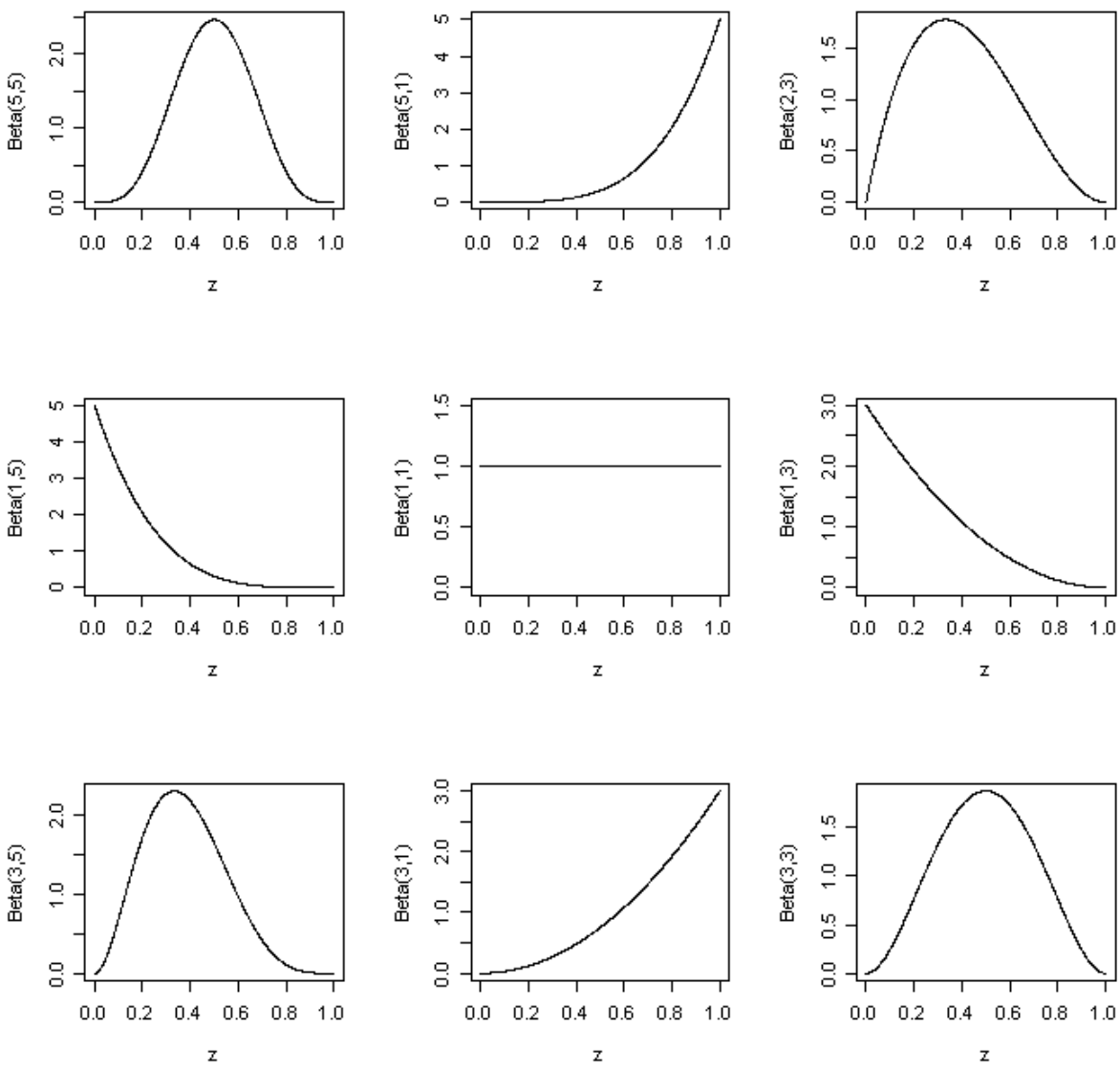
$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \cdot 1 = \frac{\alpha}{\alpha+\beta}.$$

Del mismo modo se demuestra que

$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Cuando $\alpha = \beta = 1$, se obtiene la distribución uniforme y, en general, cuando $\alpha = \beta$ la distribución es simétrica alrededor de $1/2$. Cuando $\alpha < \beta$ se concentra la probabilidad hacia la izquierda, y a la inversa cuando $\alpha > \beta$. Cuanto mayores son α y β las densidades están más concentradas en algún punto.

En la gráfica siguiente se muestran ejemplos sobre distintas formas que presenta variando sus parámetros:


Beta(5,5)
Beta(5,1)
Beta(2,3)
Beta(1,5)
Beta(1,1)
Beta(1,3)
Beta(3,5)
Beta(3,1)
Beta(3,3)
z

Las gráficas anteriores se han dibujado con R:

```
z = seq(0,1,length=1000)
par(mfrow=c(3,3))

# Parametros: alfa=5 beta=5
plot(z,dbeta(z,5,5),ylab="Beta(5,5)",type="l")

# Parametros: alfa=5 beta=1
plot(z,dbeta(z,5,1),ylab="Beta(5,1)",type="l")

# Parametros: alfa=5 beta=3
plot(z,dbeta(z,2,3),ylab="Beta(2,3)",type="l")

# Parametros: alfa=1 beta=5
plot(z,dbeta(z,1,5),ylab="Beta(1,5)",type="l")

# Parametros: alfa=1 beta=1
plot(z,dbeta(z,1,1),ylim=c(0,1.5),ylab="Beta(1,1)",type="l")

# Parametros: alfa=1 beta=3
plot(z,dbeta(z,1,3),ylab="Beta(1,3)",type="l")

# Parametros: alfa=3 beta=5
plot(z,dbeta(z,3,5),ylab="Beta(3,5)",type="l")

# Parametros: alfa=3 beta=1
plot(z,dbeta(z,3,1),ylab="Beta(3,1)",type="l")

# Parametros: alfa=3 beta=3
plot(z,dbeta(z,3,3),ylab="Beta(3,3)",type="l")
```

En nuestro caso, se tiene que

$$f(p|\alpha,\beta)=\begin{cases}\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1} & \text{para} \quad 0\leq p\leq 1\\ \\ 0 & \text{resto}\end{cases}$$

$$\begin{aligned}E(p)&=\frac{\alpha}{\alpha+\beta},\\ Var(p)&=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)},\\ Moda(p)&=\frac{\alpha-1}{\alpha+\beta-2}.\end{aligned}$$

Por ejemplo, para $\alpha=\beta=1$ se tiene la distribución uniforme que modeliza la ignorancia acerca de p.

Planteamiento del modelo beta-binomial

El problema que nos planteamos es el siguiente. Suponemos un experimento que consiste en observar n casos independientes, registrándose el número de casos favorables que se presentan. La verosimilitud (o el modelo) es, en este caso, binomial, teniéndose

$$P\left(X=x|p\right)=\binom{n}{x}p^{x}\left(1-p\right)^{n-x},$$

para $x=0,1,\ldots n$.

Se realiza, por tanto, el experimento y supongamos que se producen x éxitos. Nuestro interés se centra en estimar la proporción p de éxitos dada la muestra.

Desde el punto de vista bayesiano asumimos que tenemos una información previa sobre p que se modeliza mediante la distribución a priori de p que hemos supuesto que se recoge mediante una densidad beta. Actualizamos, entonces, nuestras creencias sobre p aplicando el teorema de Bayes. Se tiene

$$f\left(p|x\right)=\frac{f\left(p\right)P\left(x|p\right)}{P\left(x\right)}\propto\frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)}p^{\alpha-1}\left(1-p\right)^{\beta-1}\cdot\binom{n}{x}p^{x}\left(1-p\right)^{n-x}$$

$$\propto p^{x+\alpha-1}\left(1-p\right)^{n-x+\beta-1},$$

siendo $p\in\left(0,1\right)$ y $f\left(p|x\right)=0$ para $p\notin\left(0,1\right)$.

Resulta, por tanto, que la distribución a posteriori de p sigue una beta de parámetros $\left(x+\alpha\right)$ y $\left(n-x+\beta\right)$.

Se puede emplear esa distribución para resolver los problemas típicos que se presentan en inferencia.

Estimación puntual

Se trata de dar un valor-resumen representativo de la distribución a posteriori. Las más habituales son:

1. **Media a posteriori**,
$$\frac{x+\alpha}{n+\alpha+\beta}$$
2. **Moda a posteriori,**
$$\frac{x+\alpha-1}{n+\alpha+\beta-2}$$
3. **Mediana a posteriori**, que es la solución p^{*} de la ecuación
$$\int_{0}^{p^{*}}\frac{\Gamma\left(n+\alpha+\beta\right)}{\Gamma\left(\alpha+x\right)\Gamma\left(n-x+\beta\right)}p^{x+\alpha-1}\left(1-p\right)^{n-x+\beta-1}dp=\frac{1}{2}.$$
En este caso, se hace necesario usar cálculo numérico.

Ejemplo en R:

```
# Se calcula el cuantil del 50% de una beta 10, 15
qbeta(0.5,10,15)
```

```
[1] 0.3972924
```

Estimación por intervalos

De manera informal, se trata de dar un intervalo que, con cierta probabilidad r, cubra el valor de p, esto es, un intervalo con extremos a y b que satisfaga lo siguiente

$$\int_a^b f\left(p|x\right) dp = r.$$

Normalmente, se escogen a y b de manera que a deja probabilidad $\frac{1-r}{2}$ a su izquierda y $\frac{1-r}{2}$ a su derecha.

Por ejemplo, si $r = 0{,}90$, a y b satisfacen

$$\int_0^a f\left(p|x\right) dp = 0{,}05$$
$$\int_b^1 f\left(p|x\right) dp = 0{,}05.$$

Por ejemplo, en R:

```
a = qbeta(0.05,4,5)
b = qbeta(0.05,4,5,lower.tail=F)
```

Contraste de hipótesis

En muchas ocasiones se está interesado en un modelo específico o subconjuntos de modelos que denominamos *hipótesis nula* (H_0) frente al resto de modelos que se denominan *hipótesis alternativa* (H_1). Por ejemplo, podemos contrastar

$$H_0 : p = 0{,}5$$

frente a

$$H_1 : p \neq 0{,}5;$$

o podemos contrastar

$$H_0 : p \leq 0{,}6$$

frente a

$$H_1 : p > 0{,}6.$$

Contrastar H_0 frente a H_1, significa calcular la probabilidad a posteriori de cada una de las hipótesis y quedarse con la hipótesis más probable.

En el segundo ejemplo propuesto se calcularía

$$P\left(H_0|x\right) = \int_0^{0,6} f\left(p|x\right) dp,$$
$$P\left(H_1|x\right) = \int_{0,6}^{1} f\left(p|x\right) dp,$$

y se diría que los datos apoyan H_0 (frente a H_1) si

$$P\left(H_0|x\right) > P\left(H_1|x\right)$$

(o equivalentemente, $P\left(H_0|x\right) > 0{,}5$). El primer ejemplo es algo más delicado, pues $P\left(H_0|x\right) = 0$, al ser una distribución continua. Una solución rigurosa requiere cálculos más sofisticados. Alternativamente, se puede calcular un intervalo I de probabilidad r, centrado en la media a posteriori de p, y si $0{,}5 \in I$, se dice que hay evidencia a favor de la hipótesis nula. En otro caso, se dice que hay evidencia a favor de la hipótesis alternativa.

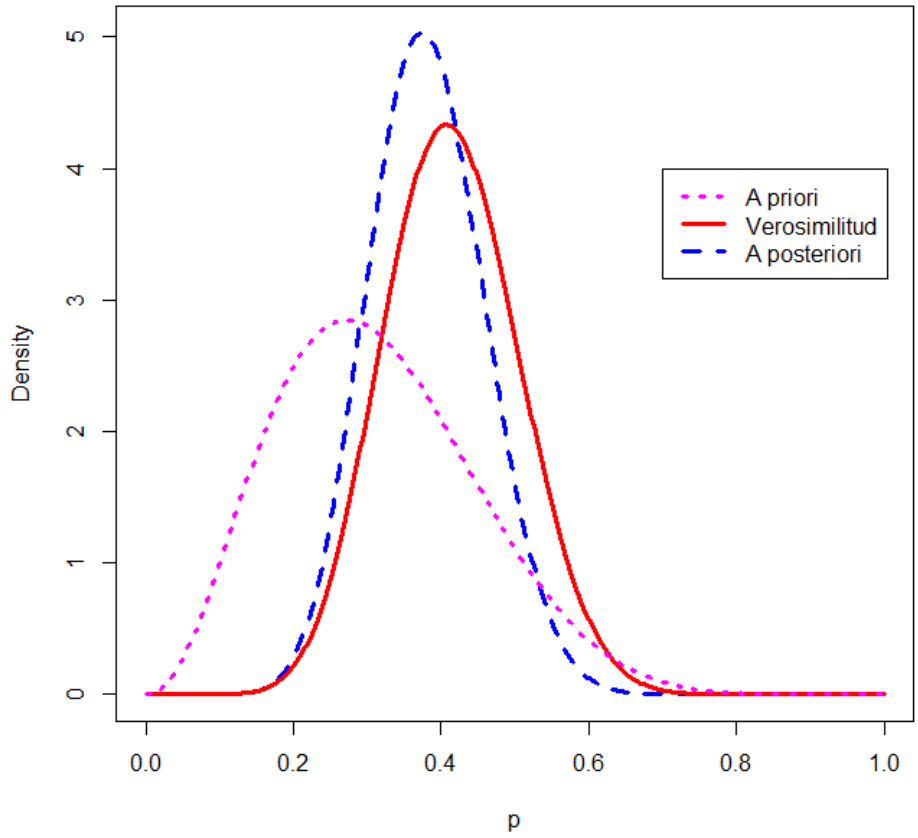
Ejemplo

Supongamos que estamos interesados en estudiar los hábitos de sueño de los estudiantes de un cierto centro. Parece ser que los médicos recomiendan un mínimo de 8 horas de sueño para una persona adulta, con lo cual el estudio se plantea en términos de averiguar la proporción de de estudiantes que duermen al menos 8 horas. Llamaremos p a dicha proporción.

Se toma una muestra de 27 estudiantes de modo que 11 de ellos duermen al menos 8 horas y el resto no. Nos planteamos hacer inferencias sobre la proporción p teniendo en cuenta la información previa que se tiene. Además interesa predecir el número de estudiantes que duermen al menos 8 horas si se toma una nueva muestra de 20 estudiantes.

Supongamos que consideramos una distribución a priori beta:

```
p=seq(0,1,length=500)
s=11
f=16
a=3.4
b=7.4

prior = dbeta(p, a, b)
like = dbeta(p, s+1, f+1)
post = dbeta(p,a+s,b+f)

plot(p,post,type="l",ylab="Density",lty=2,lwd=3, col=4)
lines(p,like,lty=1,lwd=3,col=2)
lines(p,prior,lty=3,lwd=3,col=6)
legend(0.7, 4, c("A priori","Verosimilitud","A posteriori"),
lty=c(3,1,2), lwd=c(3,3,3), col=c(6,2,4))
```



Se pueden considerar resúmenes de la distribución: ¿Cuál es $P(p \geq 0{,}5|\text{datos})$? ¿Cuál es el intervalo de confianza al 90 % para el verdadero valor de p?

```
s = 11
f = 16
a = 3.4
b = 7.4
```

```
1-pbeta(0.5, a+s, b+f)
```

```
[1] 0.06842569
```

```
qbeta(c(0.05, 0.95), a+s, b+f)
```

```
[1] 0.2562364 0.5129274
```

Otra alternativa es hacerlo mediante simulación:

```
s = 11
f = 16
a = 3.4
b = 7.4

ps = rbeta(1000, a+s, b+f)
```

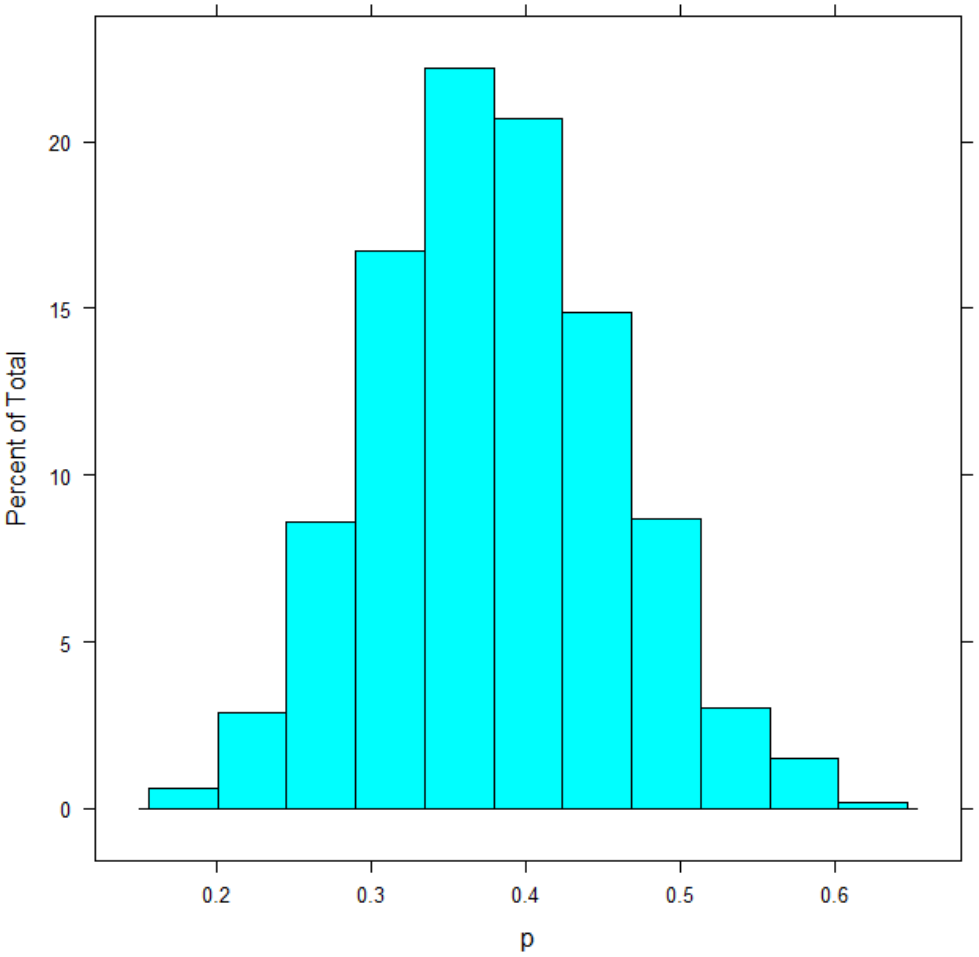
```
sum(ps >= 0.5)/1000
```

```
[1] 0.072
```

```
quantile(ps, c(0.05, 0.95))
```

```
       5%       95%
0.2538683 0.5107199
```

```
lattice::histogram(ps, xlab="p", main="")
```

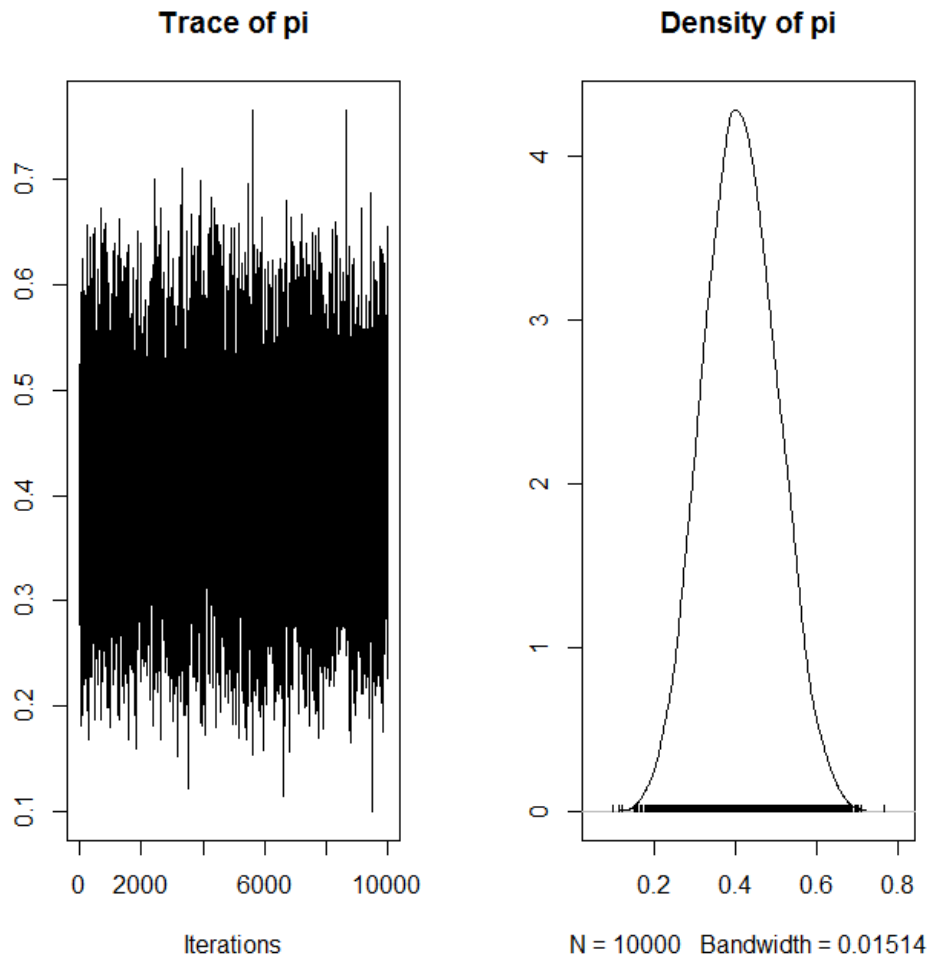


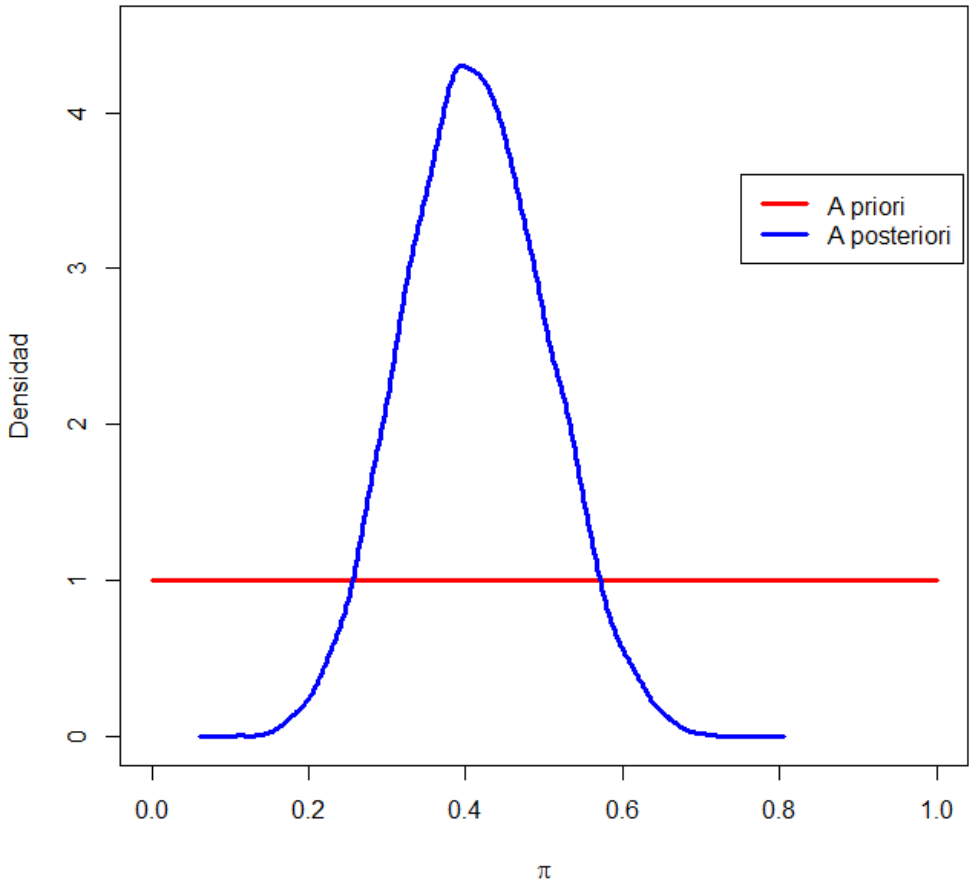
Resolución con MCMCpack

```
library(MCMCpack)

posterior = MCbinomialbeta(11, 27, mc=10000)
summary(posterior)

X11()
plot(posterior)

grid = seq(0,1,0.01)

X11()
plot(grid, dbeta(grid, 1, 1), type="l", col="red",
lwd=3, ylim=c(0,4.5), xlab=expression(pi), ylab="Densidad")
lines(density(posterior), col="blue", lwd=3)
legend(.75, 3.6, c("A priori", "A posteriori"),
lwd=3, col=c("red", "blue"))
```

Se obtiene

```
Iterations = 1:10000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

          Mean             SD       Naive SE Time-series SE
     0.4130949      0.0901232      0.0009012      0.0009012

2. Quantiles for each variable:

  2.5%    25%    50%    75%  97.5%
0.2407 0.3501 0.4111 0.4750 0.5923
```


Trace of pi
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
2000
6000
10000
Iterations
Density of pi
0
1
2
3
4
0.2
0.4
0.6
0.8
N = 10000 Bandwidth = 0.01514


A priori
A posteriori
Densidad
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0
π

Resolución con SAS

Entrar en

https://odamid.oda.sas.com/SASStudio

```
OPTIONS nodate;
TITLE 'Ejemplo BetaBinomial Bayes';

DATA cosas;
INPUT sopa @@;
DATALINES;
1 1 1 1 1 1 1 1 1 1 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0
;

PROC mcmc data=cosas nmc=10000 outpost=sale;

/* Valor inicial de la cadena */
parm p 0.2;

/* A priori no informativa:
prior p ~ uniform(0,1); */

/* A priori conjugada */
prior p ~ beta(4,4);

model sopa ~ Binary(p);
RUN;
```

Ejemplo

Un distribuidor anuncia que el 95 % de sus ordenadores no se tiene que reparar durante el año de garantía. Desde el punto de vista bayesiano se puede modelizar esta creencia mediante una $Be\left(\alpha = 4{,}75, \beta = 0{,}25\right)$, ya que si

$$p \sim Be\left(4{,}75, 0{,}25\right),$$

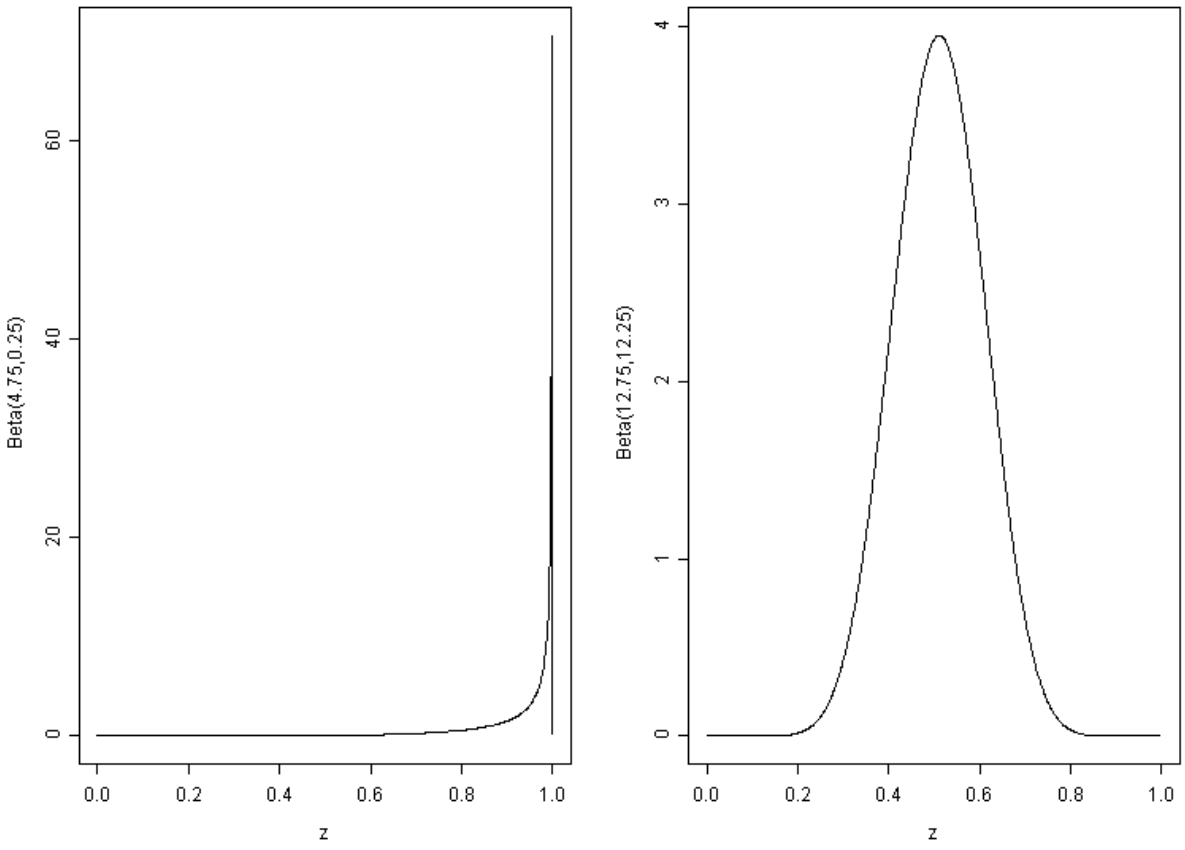
entonces

$$E\left(p\right) = \frac{4{,}75}{4{,}75 + 0{,}25} = 0{,}95,$$
$$Var\left(p\right) = 0{,}0016.$$

Debido a la aparente calidad, compramos 20 ordenadores, de los cuales 12 requieren reparación durante el año de garantía. Nuestra distribución a posteriori sobre la proporción es entonces ahora

$$Be\left(4{,}75 + 8, 0{,}25 + 12\right)$$

La figura muestra las distribuciones a priori y a posteriori.



Estimadores puntuales.

Algunos valores que resumen la distribución a posteriori son:

Media a posteriori: $E\left(p|x\right) = \frac{12{,}75}{25} = 0{,}51$

Moda a posteriori: $\frac{12{,}75-1}{25-2} = 0{,}51087$

Mediana a posteriori: 0.5102709 (en R: `qbeta(0.5,12.75,12.25)`).

Estimación por intervalos:

Un intervalo de probabilidad 0.95 es

En R:

```
p = c(0.025, 0.975)
qbeta(p, 12.75, 12.25)
```

Obteniéndose $[0{,}32; 0{,}70]$

Contraste de Hipótesis:

Contrastamos la hipótesis $H_0 : p \geq 0{,}95$ frente a $H_1 : p < 0{,}95$, que se correspondería con contrastar el anuncio.

Como, al considerar la orden en R, `pbeta(0.95,12.75,12.25)`, se obtiene 1, entonces se tiene que

$$P\left(H_1|x\right) \approx 1$$
$$P\left(H_0|x\right) \approx 0,$$

por lo que los datos sugieren que el distribuidor está exagerando.

Predicción

Dados los datos $\mathbf{x} = (x_1, \ldots, x_n)'$, supongamos que se quiere predecir el valor de una futura observación X_{n+1}. Para ello, hay que calcular la distribución predictiva. Se puede considerar tanto la distribución predictiva *a priori* como *a posteriori* dependiendo de si se usa una u otra distribución de los parámetros.

Sus expresiones son

$$f(x_{n+1}|\mathbf{x}) = \int f(x_{n+1}|\theta, \mathbf{x}) f(\theta)\, d\theta$$
$$f(x_{n+1}|\mathbf{x}) = \int f(x_{n+1}|\theta, \mathbf{x}) f(\theta|\mathbf{x})\, d\theta.$$

En nuestra situación, los $X_i|\theta$ son intercambiables, es decir, su distribución es igual cuando se les cambia de orden y lo anterior se puede simplificar como

$$f(x_{n+1}|\mathbf{x}) = \int f(x_{n+1}|\theta) f(\theta)\, d\theta$$
$$f(x_{n+1}|\mathbf{x}) = \int f(x_{n+1}|\theta) f(\theta|\mathbf{x})\, d\theta.$$

Ejemplo:

Supongamos que se tira una moneda equilibrada y que salen 9 cruces y 3 caras (considero como *éxito* $\equiv$ *cruz*). Suponiendo una distribución a priori uniforme, se tiene que la distribución a posteriori de p sigue una distribución beta de parámetros $(x+\alpha)$ y $(n-x+\beta)$: Beta$(10,4)$.

A continuación, se quiere predecir el número de cruces que se obtendrían en diez tiradas más.

Se tiene que, ahora,

$$X_{fut}|\theta \sim \text{Bin}(10,\theta),$$

entonces

$$\begin{aligned} f(x_{fut}|\mathbf{x}) &= \int_0^1 \begin{pmatrix} 10 \\ x_{fut} \end{pmatrix} \theta^{x_{fut}}(1-\theta)^{10-x_{fut}} \times \\ &\times \frac{\Gamma(14)}{\Gamma(10)\Gamma(4)} \theta^{10-1}(1-\theta)^{4-1}\, d\theta \\ &= \begin{pmatrix} 10 \\ x_{fut} \end{pmatrix} \frac{\Gamma(14)}{\Gamma(10)\Gamma(4)} \times \\ &\times \int_0^1 \theta^{10+x_{fut}-1}(1-\theta)^{14-x_{fut}-1}\, d\theta \\ &= \begin{pmatrix} 10 \\ x_{fut} \end{pmatrix} \frac{\text{Be}(10+x_{fut}, 14-x_{fut})}{\text{Be}(10,4)} \end{aligned}$$

donde se ha usado la notación

$$\text{Be}(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

En general, si estamos interesados en r éxitos en los m siguientes ensayos haremos

$$P\left[r \text{ éxitos en } m \text{ ensayos } | \mathbf{x}\right] = \int_0^1 \binom{m}{r} p^r (1-p)^{m-r} f(p|x)\, dp =$$

$$= \int_0^1 \binom{m}{r} p^r (1-p)^{m-r} \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+x)\Gamma(n-x+\beta)} p^{x+\alpha-1}(1-p)^{n-x+\beta-1}\, dp =$$

$$= \binom{m}{r} \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+x)\Gamma(n-x+\beta)} \frac{\Gamma(r+x+\alpha)\Gamma(m-r+n-x-\beta)}{\Gamma(m+n+\alpha+\beta)}.$$

Se puede evaluar la media de $X_{fut}|\mathbf{x}$ sin tener que evaluar la distribución predictiva.

NOTA: se tenía que,

$$E_z[Z] = E_y[E_z[Z|Y]]$$

para cualquier variable Z y Y.

Así, respecto al ejemplo anterior,

$$E_x[X_{fut}|\mathbf{x}] = E_\theta[E_x[X_{fut}|\theta]]$$

como $X_{fut}|\theta \sim \text{Bin}(10,\theta)$ entonces $E_x[X_{fut}|\theta] = 10\theta$, y como $\theta|\mathbf{x} \sim \text{Be}(10,4)$ entonces $E_\theta[\theta|\mathbf{x}] = \frac{10}{14}$, por tanto

$$E_x[X_{fut}|\mathbf{x}] = E_\theta[E_x[X_{fut}|\theta]] = 10 \times \frac{10}{14} \approx 7{,}141$$

Para evaluar la varianza predictiva, se usa la fórmula

$$\begin{aligned} V_z[Z] &= E_y[V_z[Z|Y]] + V_y[E_z[Z|Y]] \\ &= E_\theta[V_x[X_{fut}|\theta]] + V_\theta[E_x[X_{fut}|\theta]] \end{aligned}$$

Evaluamos ambas partes.

$$\begin{aligned} V_x[X_{fut}|\theta] &= 10\theta(1-\theta) = 10(\theta - \theta^2) \\ E_\theta[V_x[X_{fut}|\theta]] &= 10\left(E_\theta[\theta|\mathbf{x}] - E_\theta[\theta^2|\mathbf{x}]\right) \\ &= 10\left(E_\theta[\theta|\mathbf{x}] - (E_\theta[\theta|\mathbf{x}]^2 + V_\theta[\theta|\mathbf{x}])\right) \\ &= 10\left(\frac{10}{10+4} - \left(\frac{10}{10+4}\right)^2 - \right. \\ &\left. - \frac{10\cdot 4}{(10+4)^2(10+4+1)}\right) = \\ &= \frac{40}{21} \end{aligned}$$

y como $E_x[X_{fut}|\theta] = 10\theta$,

$$\begin{aligned} V_\theta[E_x[X_{fut}|\theta]] &= 100 V_\theta[\theta|\mathbf{x}] \\ &= 100 \frac{10\cdot 4}{(10+4)^2(10+4+1)} \\ &= \frac{200}{147} \end{aligned}$$

Entonces tenemos $V_x[X_{fut}|\mathbf{x}] = \frac{40}{21} + \frac{200}{147} \approx 3{,}3$.

Predicción en el ejemplo de los ordenadores:

La probabilidad de que el siguiente ordenador que adquiramos no deba repararse en el periodo de garantía es igual a la media a posteriori ya que es una $\text{Ber}(\theta)$, luego

$$f(x_{fut}|\mathbf{x}) = \int_0^1 \theta f(\theta|\mathbf{x}) d\theta = E(\theta|\mathbf{x}) = \frac{12{,}75}{25} = 0{,}51.$$

Asignación de distribuciones beta

Se trata, ahora de encontrar procedimientos para asignar la distribución a priori, supuesto que es $Be(\alpha, \beta)$. Esto implica escoger α y β, lo que implica extraer dos juicios de un experto.

Una forma de proceder es la siguiente. Primero pedimos al experto que indique la probabilidad r de éxito en el primer ensayo. Sabemos que ésta es $\frac{\alpha}{\alpha+\beta}$.

Después, se pide al experto que suponga que el primer ensayo fue un éxito y que proporcione la probabilidad r_2 de éxito en el segundo ensayo; la densidad actualizada por el primer éxito es $Be(\alpha+1, \beta)$, por lo que la probabilidad r_2 será $\frac{\alpha+1}{\alpha+\beta+1}$.

Se resuelve, entonces, el sistema

$$r_1 = \frac{\alpha}{\alpha+\beta},$$
$$r_2 = \frac{\alpha+1}{\alpha+\beta+1}$$

y se obtiene,

$$\alpha = \frac{r_1(1-r_2)}{r_2-r_1},$$
$$\beta = \frac{(1-r_1)(1-r_2)}{r_2-r_1}.$$

Nota:

Si se emplea la $\mathrm{Be}(1,1)$ se corresponde a la $U(0,1)$ modeliza la ignorancia o total incertidumbre. En este sentido, cuando tengamos mucha incertidumbre deberíamos favorecer distribuciones a priori con valores pequeños de α y β que facilitan la *absorción* de evidencia.

Ejemplo

Supongo que el experto afirma que $r_1 = 0{,}7$ y $r_2 = 0{,}75$. Entonces,

$$\alpha = \frac{0{,}7(1-0{,}75)}{0{,}75-0{,}7} = 3{,}5,$$
$$\beta = \frac{(1-0{,}7)(1-0{,}75)}{0{,}75-0{,}7} = 1{,}5.$$

Observemos que todos los cálculos que hagamos quedarán afectados por r_1 y r_2, a través de α y β, especialmente con pocos datos.

Alternativas para la distribución a priori

En una muestra aleatoria tomada de una clase de 27 alumnos, se considera la variable aleatoria X que es igual a 1 si los alumnos duermen más de 8 horas al día ó 0 en caso contario. La verosimilitud es por tanto binomial de parámetro desconocido p.

Supongamos que la distribución de probabilidad a priori para p se denota como $g(p)$. Si denominamos como éxito al hecho de dormir más de 8 horas, la verosimilitud es

$$L(p) \propto p^s(1-p)^f$$

de modo que la densidad a posteriori es

$$g\left(p|\text{datos}\right) \propto g(p) \cdot p^s(1-p)^f$$

Luego interesa determinar la distribución a priori.

Alternativa 1: A priori discreta.

Supongamos que se piensa que los valores posibles para p son

$$0{,}05;\ 0{,}15;\ 0{,}25;\ 0{,}35;\ 0{,}45;\ 0{,}55;\ 0{,}65;\ 0{,}75;\ 0{,}85;\ 0{,}95$$

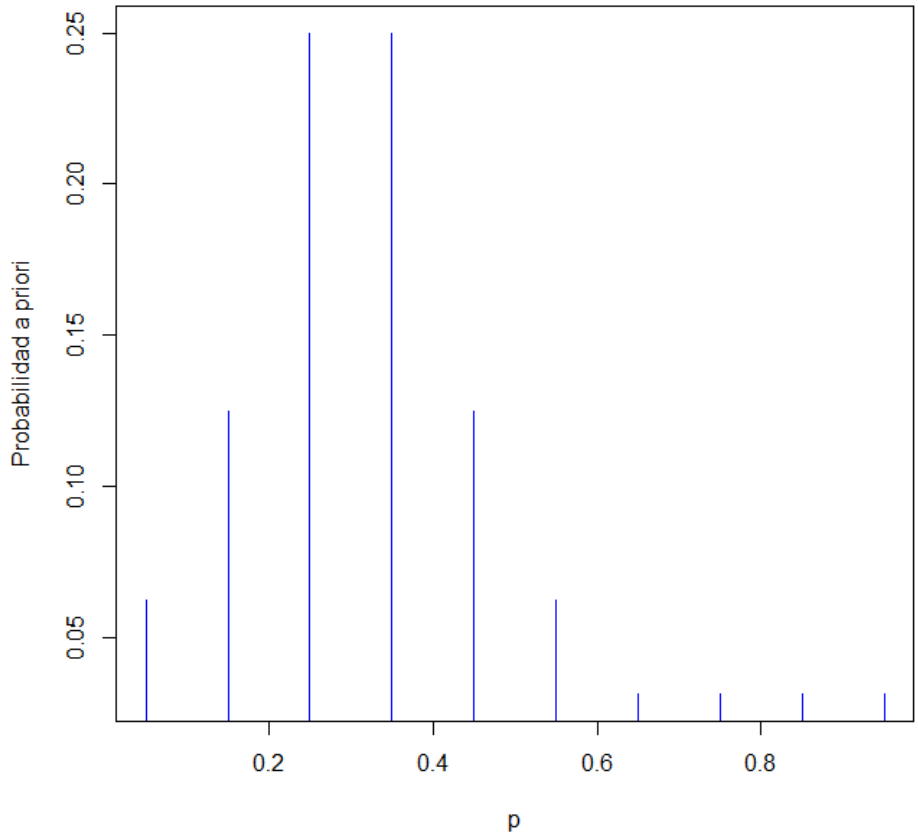
basado en su información previa le asigna los siguientes pesos:

$$2, 4, 8, 8, 4, 2, 1, 1, 1, 1$$

que se pueden convertir en probabilidades a priori normalizando los pesos para que sumen 1.

En `R` esto se puede hacer así:

```
p = seq(0.05, 0.95, by=0.1)
prior = c(2, 4, 8, 8, 4, 2, 1, 1, 1, 1)
prior = prior/sum(prior)
plot(p, prior, type="h", ylab="Probabilidad a priori", col=4)
```



En nuestro caso hay 11 estudiantes que duermen bien, por lo que

$$L(p) \propto p^{11}(1-p)^{16}$$

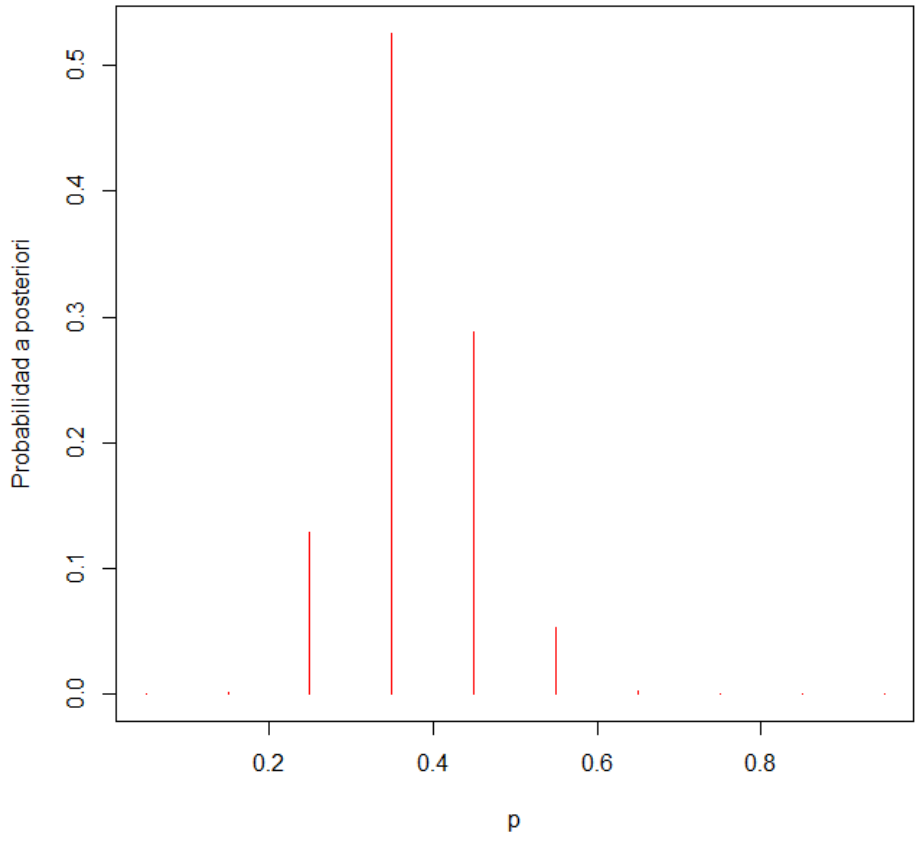
Se puede observar que es una distribución beta: $Be(12, 17)$

Para calcular las probabilidades a posteriori usamos la librería `LearnBayes`:

```
library(LearnBayes)
data = c(11, 16)
post = pdisc(p, prior, data)
cbind(p, prior, post)
plot(p, post, type="h", ylab="Probabilidad a posteriori", col=2)
```

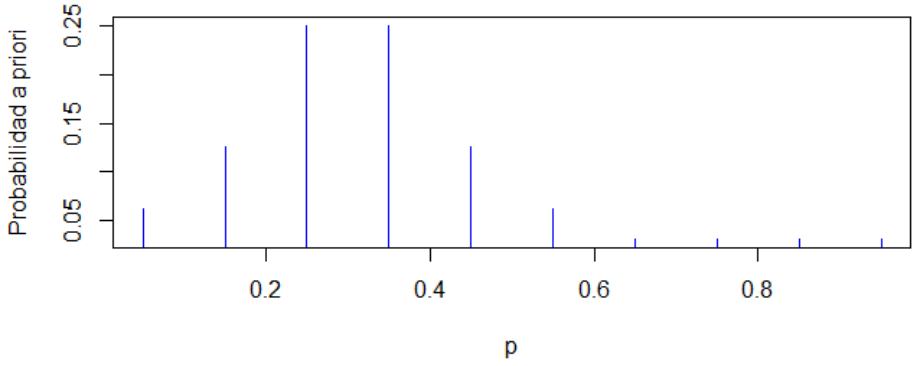
Se obtiene

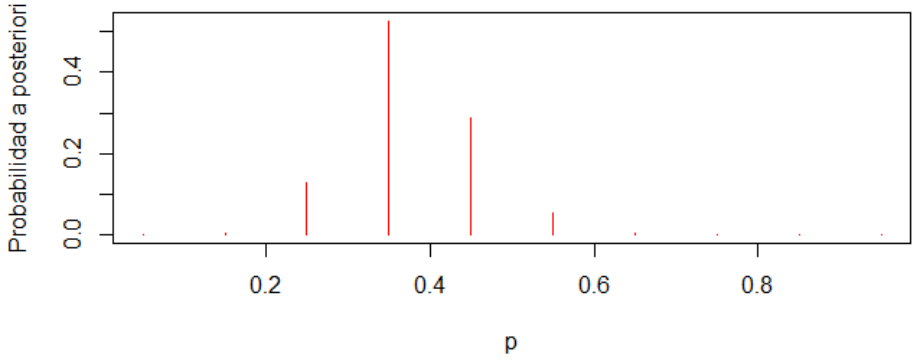
```
         p   prior          post
 [1,] 0.05 0.06250 2.882642e-08
 [2,] 0.15 0.12500 1.722978e-03
 [3,] 0.25 0.25000 1.282104e-01
 [4,] 0.35 0.25000 5.259751e-01
 [5,] 0.45 0.12500 2.882131e-01
 [6,] 0.55 0.06250 5.283635e-02
 [7,] 0.65 0.03125 2.976107e-03
 [8,] 0.75 0.03125 6.595185e-05
 [9,] 0.85 0.03125 7.371932e-08
[10,] 0.95 0.03125 5.820934e-15
```



Si dibujamos ambas para comparar cómo ha variado la distribución de p después de obtener los datos:

```
par(mfrow=c(2,1))
plot(p, prior, type="h", ylab="Probabilidad a priori", col=4)
plot(p, post, type="h", ylab="Probabilidad a posteriori", col=2)
```



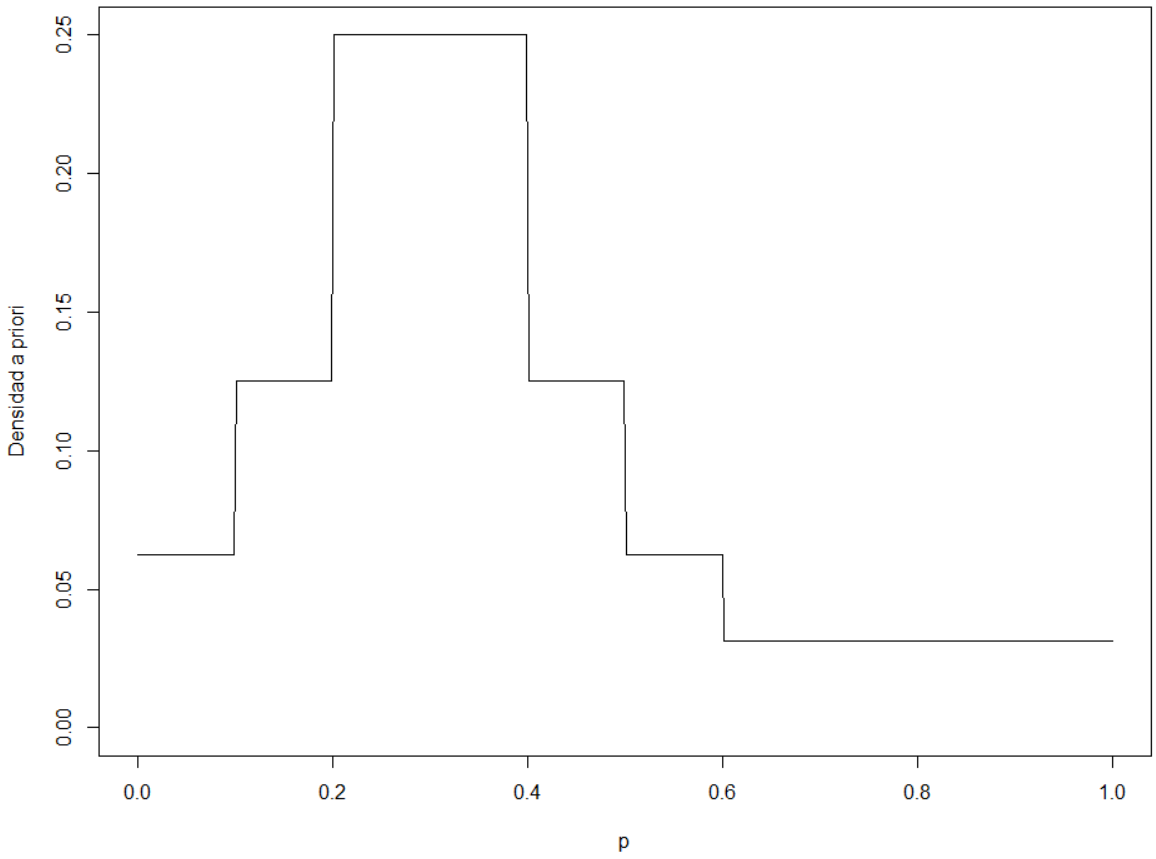


Alternativa 2: A priori por un histograma. Se tendría una versión continua de la anterior.

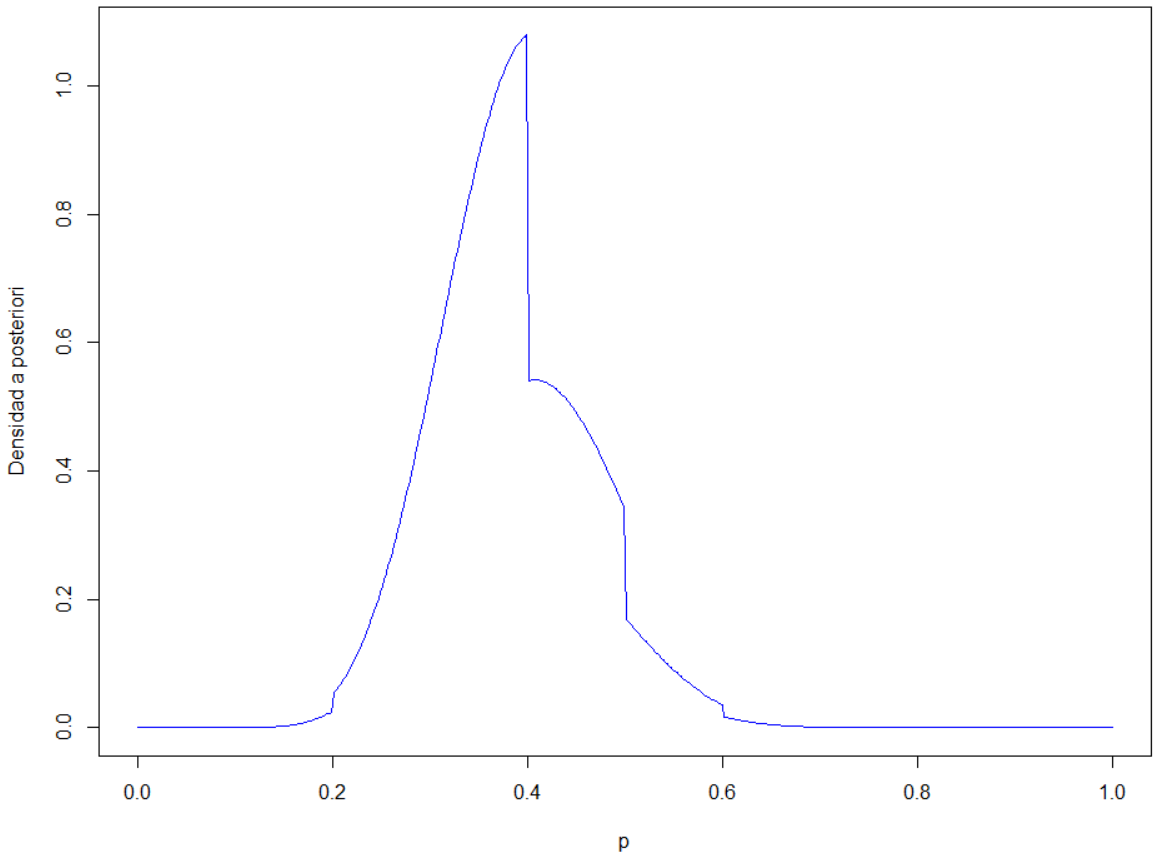
Supongamos que se puede asignar una distribución a priori sobre intervalos. Asumimos que nos basta con 10 para el rango de p: (0, 0.1), (0.1, 0.2),... (0.9, 1) de modo que asignamos como pesos:

$$2, 4, 8, 8, 4, 2, 1, 1, 1, 1$$

```
library(LearnBayes)
midpt = seq(0.05, 0.95, by = 0.1)
prior = c(2, 4, 8, 8, 4, 2, 1, 1, 1, 1)
prior = prior/sum(prior)
p = seq(0, 1, length = 500)

plot(p, histprior(p,midpt,prior), type="l",
ylab="Densidad a priori", ylim=c(0,.25))
```

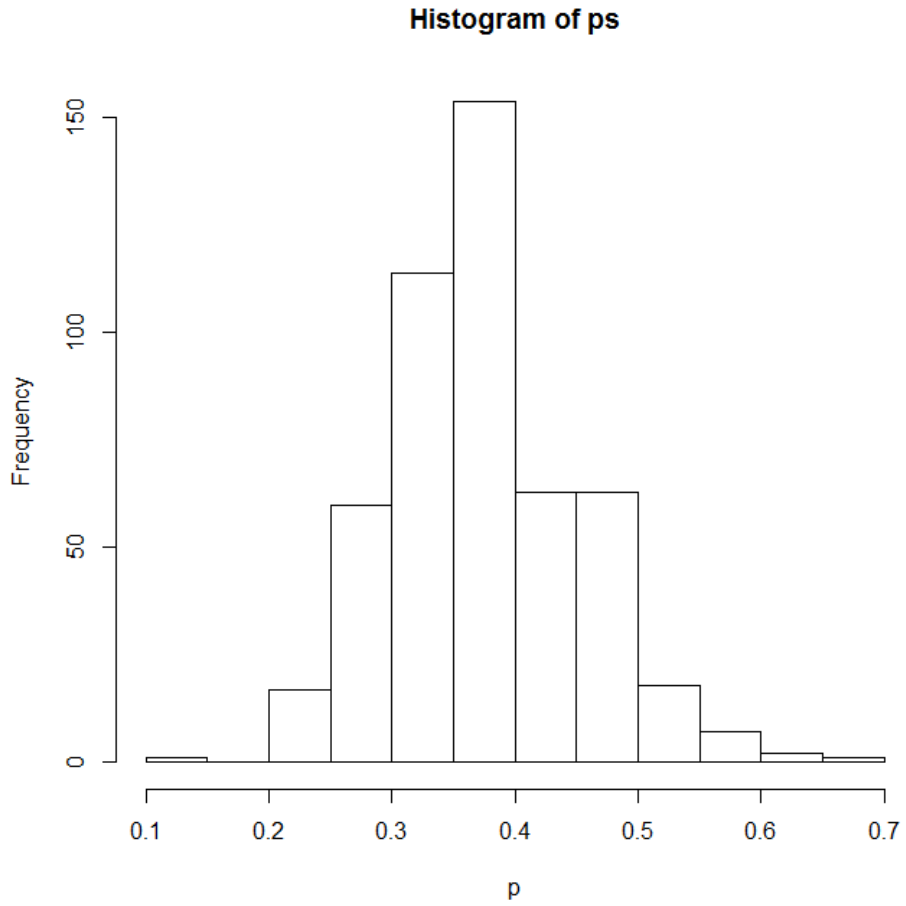



```
s = 11
f = 16
like = dbeta(p, s+1, f+1)
post = like*histprior(p, midpt, prior)
plot(p, post, type="l", ylab="Densidad a posteriori", col=4)
```



Se puede obtener una muestra simulada de la densidad a posteriori, convirtiendo los productos que aparecen en la rejilla anterior en probabilidades. Se toma una muestra con reemplazamiento de de la rejilla:

```
post = post/sum(post)
ps = sample(p, replace=TRUE, prob=post)
hist(ps, xlab="p")
```



Ejemplo

En el ejemplo de las horas de sueño se puede considerar la distribución predictiva a posteriori:

```
library(LearnBayes)
s = 11
f = 16
a = 3.4
b = 7.4
ab = c(a+s, b+f)
m = 20
ys = 0:20
pred = pbetap(ab, m, ys)
cbind(ys,pred)
```

Se obtiene

```
      ys   pred
 [1,]  0 0.0006
 [2,]  1 0.0040
 [3,]  2 0.0143
 [4,]  3 0.0348
 [5,]  4 0.0653
 [6,]  5 0.1002
 [7,]  6 0.1299
 [8,]  7 0.1456
 [9,]  8 0.1431
[10,]  9 0.1242
[11,] 10 0.0957
[12,] 11 0.0655
[13,] 12 0.0398
[14,] 13 0.0212
[15,] 14 0.0099
[16,] 15 0.0040
[17,] 16 0.0013
[18,] 17 0.0004
[19,] 18 0.0001
[20,] 19 0.0000
[21,] 20 0.0000
```

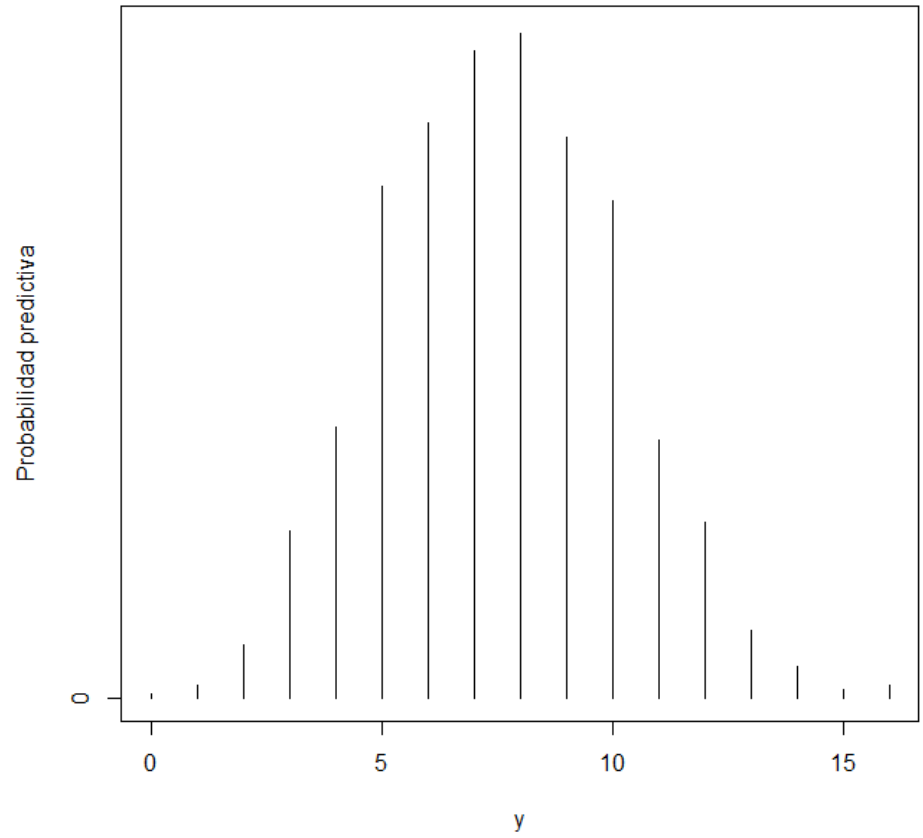
De manera alternativa, para obtener las predicciones se puede simular p^* de la distribución a posteriori, obtiéndose las predicciones a partir de una distribución binomial con parámetro igual a p^*:

```
p = rbeta(1000, a+s, b+f)
y = rbinom(1000, 20, p)
table(y)
```

```
y
  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16
  1   3  12  37  60 113 127 143 147 124 110  57  39  15   7   2   3
```

Se puede dibujar la distribución de las predicciones

```
freq=table(y)
if(names(freq)[1]=="0"){
# Si aparece alguna observacion 0
ys=c(0:max(y))
}else{
# Si NO aparece ninguna observacion 0
ys=c(1:max(y))}
predprob=freq/sum(freq)
plot(ys,predprob,type="h",xlab="y",ylab="Probabilidad predictiva")
```



Si se quiere resumir la distribución predictiva mediante un intervalo que cubra el 90 % de la función, se puede usar la función `discint`.

Se obtienen los valores que incluye el intervalo y la probabilidad exacta de los mismos.

```
# Highest probability interval for a
# discrete distribution
covprob = 0.9
dist = cbind(ys, as.vector(predprob))
discint(dist, covprob)
```

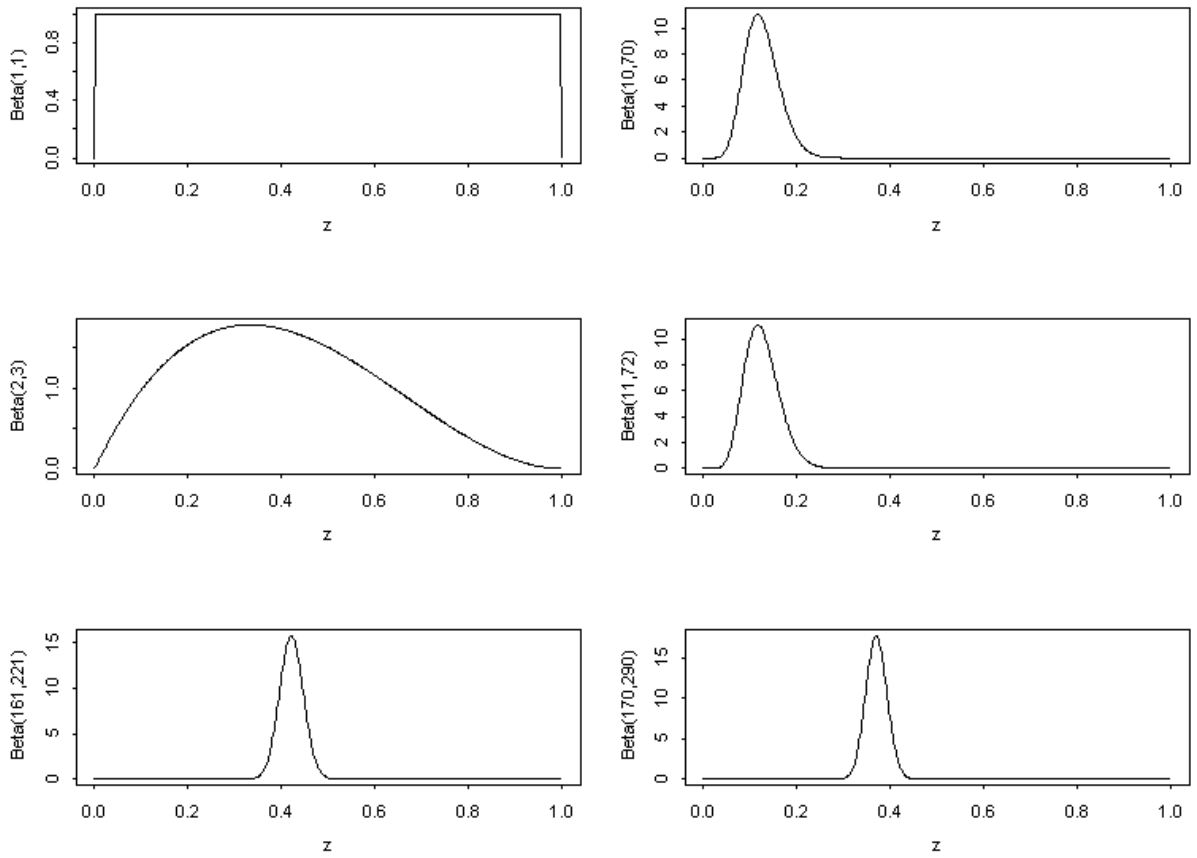
```
$prob
[1] 0.92

$set
[1]  4  5  6  7  8  9 10 11 12
```

Comportamiento del modelo con muestras grandes

Ejemplo

Dos ingenieros se enfrentan a un problema de control de calidad; para ello se fijan en la proporción de piezas defectuosas. El primero es inexperto sobre el proceso de producción de modo que se modeliza sus creencias sobre dicha proporción mediante una $Be\,(1,1)$ (es decir, con una $U(0,1)$). El segundo tiene más experiencia y modeliza sus creencias sobre la proporción mediante una $Be\,(10,70)$. Ambos observan tres piezas, siendo una de ellas defectuosa. Las creencias del primer ingeniero pasan a ser $Be\,(2,3)$, cambiando radicalmente, mientras que las del segundo pasan a ser $Be\,(11,72)$, sin cambiar prácticamente nada.

Después de observar 220 piezas en buen estado y 160 defectuosas, las creencias del primero pasan a ser $Be\,(161,221)$ y las del segundo pasan a ser $Be\,(170,290)$, pareciéndose ahora las creencias de ambos ingenieros.



El ejemplo describe una situación característica de un modelo beta-binomial que se da, también, en otros modelos: al aumentar la evidencia aportada por los datos, la influencia de la distribución a priori disminuye.

En efecto, como la distribución a posteriori es $Be\left(x+\alpha, n-x+\beta\right)$, cuando n es muy grande, la media a posteriori es

$$\frac{x+\alpha}{\alpha+\beta+n} = \frac{\frac{x}{n}+\frac{\alpha}{n}}{1+\frac{\alpha+\beta}{n}} \approx \frac{x}{n}$$

y la varianza a posteriori

$$\frac{\left(x+\alpha\right)\left(\beta+n-x\right)}{\left(\alpha+\beta+n\right)^2\left(\alpha+\beta+n+1\right)} \approx \frac{x(n-x)}{n^3} \approx \frac{x}{n}\left(1-\frac{x}{n}\right)\frac{1}{n}$$

Por tanto, para n suficientemente grande, al tender la varianza hacia 0, la distribución se concentrará más y más en torno a x/n. De hecho, el resultado es más preciso puesto que se puede probar que para n suficientemente grande, la distribución a posteriori es aproximadamente normal $N\left(\frac{x}{n};\ \sqrt{\frac{x}{n}\left(1-\frac{x}{n}\right)\frac{1}{n}}\right)$. Este resultado simplifica las cuestiones computacionales.

Así, por ejemplo, se puede considerar:

(i) Estimación puntual con la media a posteriori: $\dfrac{x}{n}$

(ii) Estimación por intervalos

$$I = \left[\frac{x}{n} - z_{\frac{\alpha}{2}}\sqrt{\frac{x}{n}\left(1-\frac{x}{n}\right)\frac{1}{n}}; \frac{x}{n} + z_{\frac{\alpha}{2}}\sqrt{\frac{x}{n}\left(1-\frac{x}{n}\right)\frac{1}{n}}\right]$$

(iii) Contraste de hipótesis (por ejemplo, $H_0 : p = 0{,}5$): Se rechaza H_0 si $0{,}5 \notin \mathbf{I}$.

El resultado anterior es bastante general y se describe técnicamente como la normalidad asintótica de la distribución a posteriori.

Una alternativa consiste en emplear una simulación de la distribución a posteriori.

Este caso es muy sencillo: basta simular de una $Be\left(\alpha+x, \beta+n-x\right)$ y sustituir momentos poblacionales por momentos muestrales.

Por ejemplo, podemos simular 1000 observaciones de tal distribución $(x_1, x_2, \ldots, x_{1000})$ y

1. Aproximar la media a posteriori por $\frac{\sum_i x_i}{n}$
2. Aproximar el intervalo de probabilidad $1-\alpha$ (pongamos, para fijar ideas que $\alpha = 0{,}1$) mediante$\left[X_{(50)}, X_{(950)}\right]$

Ejemplo

Generamos 1000 observaciones de una $Be\left(12{,}75, 12{,}25\right)$. En `R` se haría:

```
x = rbeta(1000,12.75,12.25)
mean(x)
y = sort(x)
y[975]
y[25]
```

La media muestral es 0,5110.

Al ordenar la muestra se obtiene

$$X_{(975)} = 0{,}7017$$

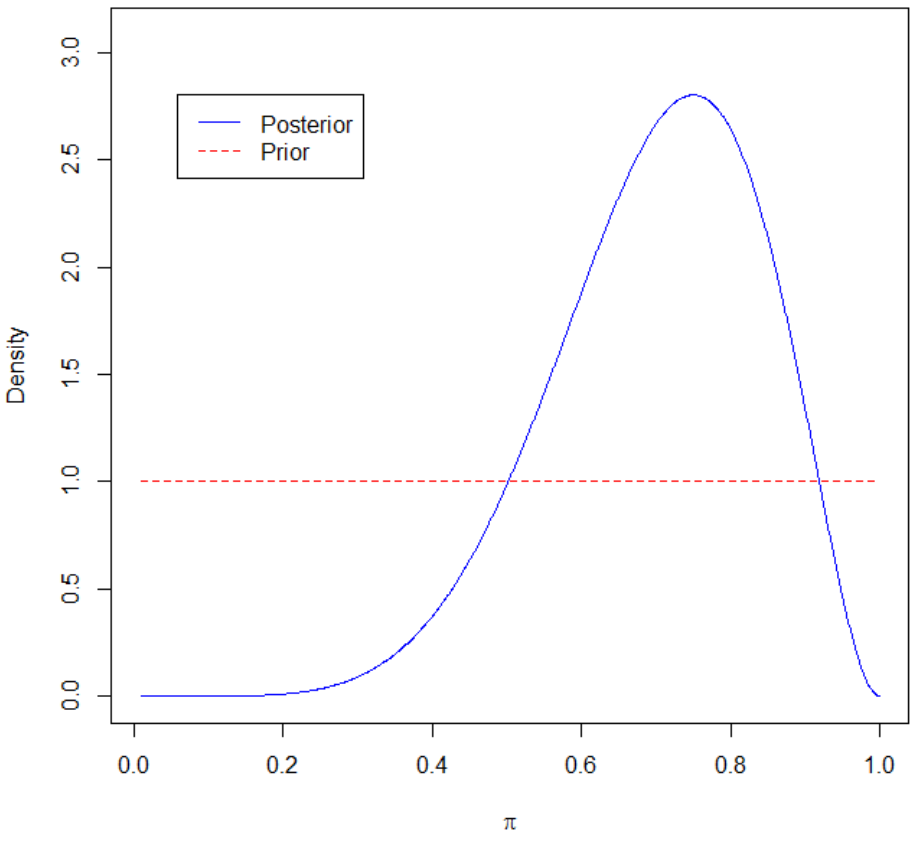
$$X_{(25)} = 0{,}3197$$

con lo que un intervalo de probabilidad a posteriori de 0,95 es $[0{,}3197; 0{,}7017]$.

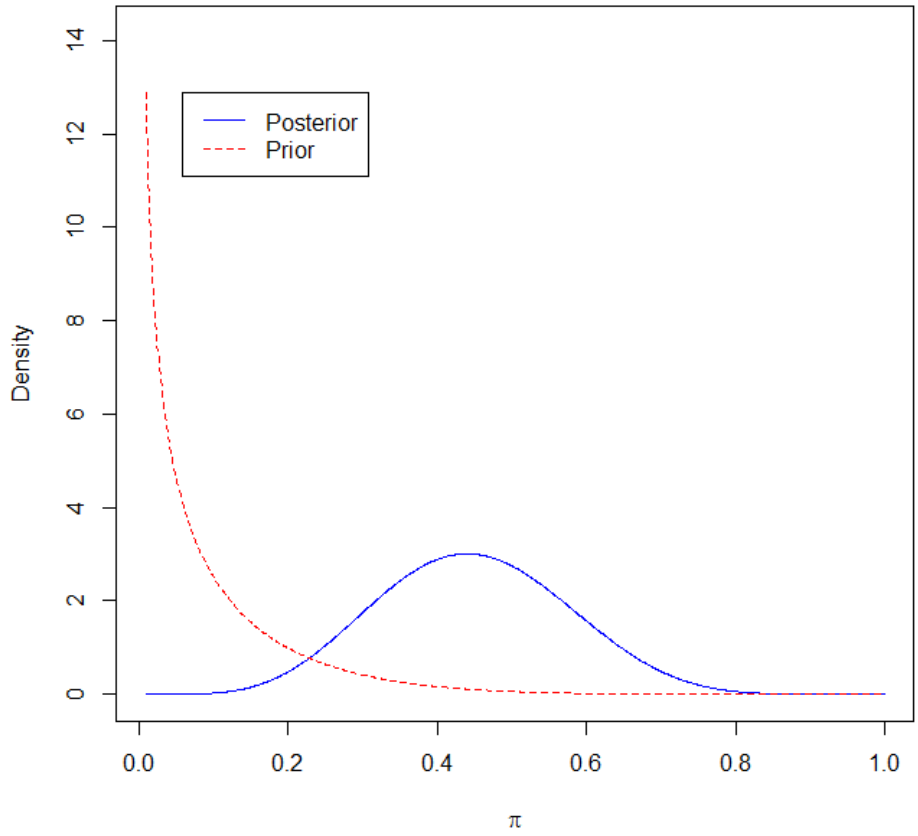
Ejemplo con R

Se necesita instalar previamente la librería `Bolsdtad` de `R`.

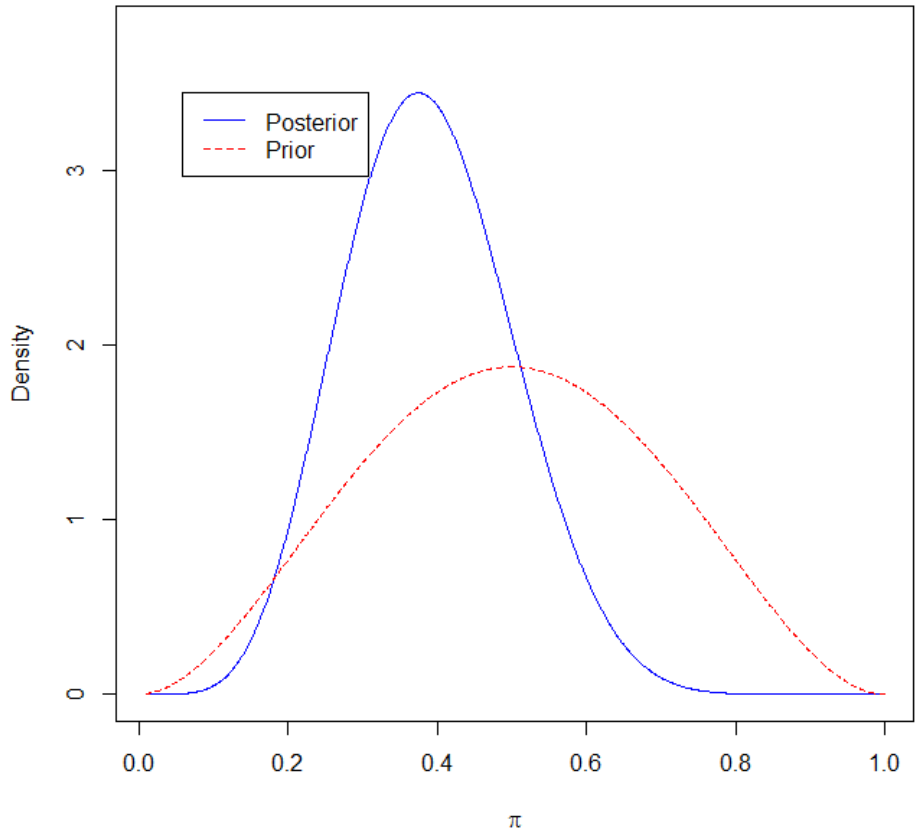
```
library(Bolstad)
# Ejemplo con 6 exitos en 8 ensayos con una a priori
# uniforme Beta(1,1)
x11()
binobp(6,8)
```



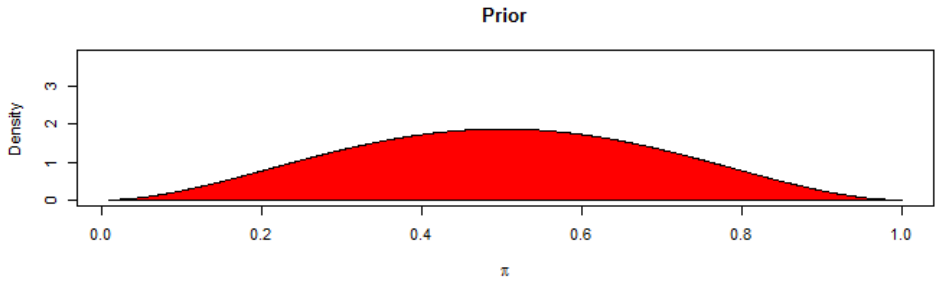
```
# 6 exitos en 8 ensayos con una a priori Beta(0.5,6)
x11()
binobp(6,8,0.5,6)
```

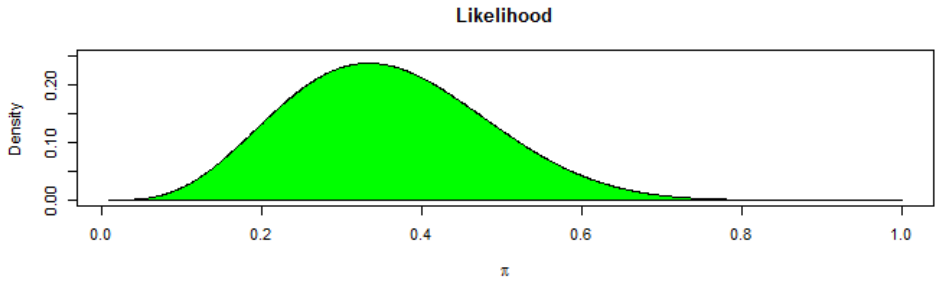


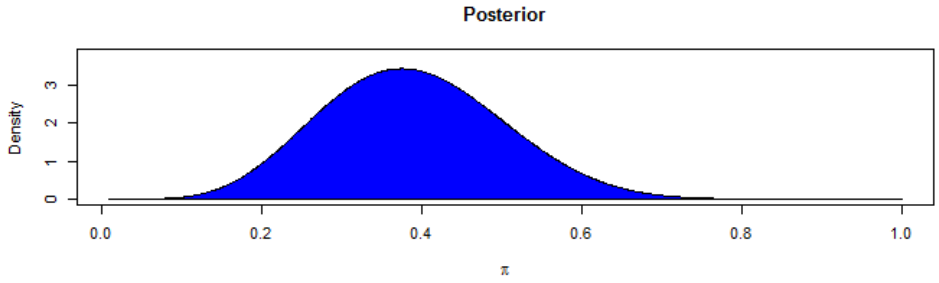

```
# 4 exitos observados in 12 ensayos con una a priori Beta(3,3)
# Se dibujan todas las graficas
x11()
results = binobp(4,12,3,3)
```




```
x11()
par(mfrow=c(3,1))

y.lims = c(0,1.1*max(results$posterior,results$prior))

plot(results$pi,results$prior,ylim=y.lims,type="l",
xlab=expression(pi),ylab="Density",main="Prior")

polygon(results$pi,results$prior,col="red")

plot(results$pi,results$likelihood,ylim=c(0,0.25),type="l",
xlab=expression(pi),ylab="Density",main="Likelihood")

polygon(results$pi,results$likelihood,col="green")

plot(results$pi,results$posterior,ylim=y.lims,type="l",
xlab=expression(pi),ylab="Density",main="Posterior")

polygon(results$pi,results$posterior,col="blue")
```


Prior
Density
0
1
2
3
0.0
0.2
0.4
0.6
0.8
1.0
π


Likelihood
Density
0.00
0.10
0.20
0.0
0.2
0.4
0.6
0.8
1.0
π


Posterior
Density
0
1
2
3
0.0
0.2
0.4
0.6
0.8
1.0
π

Comparación de proporciones

Se estudia, ahora, el problema de la comparación entre dos proporciones. Para ello se consideran dos poblaciones que dan lugar a dos experimentos, cada uno con una proporción p_1 y p_2 de éxito.

Deseamos comparar p_1 y p_2. Por ejemplo, si $p_1 \geq p_2$ o viceversa.

Suponemos que se tiene una muestra de tamaño n_1 de la primera población con x_1 éxitos. La verosimilitud es, en este caso,

$$P\left(X_1 = x_1 | p_1\right) = \binom{n_1}{x_1} p_1^{x_1} \left(1 - p_1\right)^{n_1 - x_1}.$$

Además, se tiene otra muestra de tamaño n_2 de la segunda población con x_2 éxitos. La verosimilitud es, en este caso,

$$P\left(X_2 = x_2 | p_2\right) = \binom{n_2}{x_2} p_2^{x_2} \left(1 - p_2\right)^{n_2 - x_2}.$$

Supondremos que hay independencia entre ambas muestras (a veces esta hipótesis no será adecuada, como ocurre, por ejemplo, en muestras *apareadas*), con lo cual la verosimilitud conjunta será el producto de las verosimilitudes marginales

$$P\left(X_1 = x_1, X_2 = x_2 | p_1, p_2\right) = \binom{n_1}{x_1} p_1^{x_1} \left(1 - p_1\right)^{n_1 - x_1} \binom{n_2}{x_2} p_2^{x_2} \left(1 - p_2\right)^{n_2 - x_2}.$$

Para cada p_i disponemos de información a priori que modelizamos mediante una distribución $Be\left(\alpha_i, \beta_i\right)$. Supondremos que las p_i son independientes. De nuevo esta hipótesis no será siempre adecuada; por ejemplo, podemos creer que si una proporción es grande, es probable que la otra también lo será, o, a la inversa, al ser un proporción grande, la otra podría ser pequeña. Así, la distribución a priori conjunta es

$$f\left(p_1, p_2\right) \propto p_1^{\alpha_1 - 1} \left(1 - p_1\right)^{\beta_1 - 1} p_2^{\alpha_2 - 1} \left(1 - p_2\right)^{\beta_2 - 1} I_{[0,1]}\left(p_1\right) I_{[0,1]}\left(p_2\right).$$

Calculamos la distribución a posteriori de p_1 y p_2 multiplicando por la verosimilitud,

$$f\left(p_1, p_2 | x_1, x_2\right) \propto p_1^{\alpha_1 + x_1 - 1} \left(1 - p_1\right)^{n_1 - x_1 + \beta_1 - 1} p_2^{\alpha_2 + x_2 - 1} \left(1 - p_2\right)^{n_2 - x_2 + \beta_2 - 1} I_{[0,1]}\left(p_1\right) I_{[0,1]}\left(p_2\right).$$

Se comprueba, además, que

$$\begin{aligned} f\left(p_1 | x_1, x_2\right) &\propto p_1^{\alpha_1 + x_1 - 1} \left(1 - p_1\right)^{n_1 - x_1 + \beta_1 - 1} I_{[0,1]}\left(p_1\right), \\ f\left(p_2 | x_1, x_2\right) &\propto p_2^{\alpha_2 + x_2 - 1} \left(1 - p_2\right)^{n_2 - x_2 + \beta_2 - 1} I_{[0,1]}\left(p_2\right), \end{aligned}$$

con lo que p_1 y p_2 son también independientes a posteriori. Este resultado es, de hecho, general.

La distribución a posteriori se emplea en sentido similar a como lo hemos hecho en el caso de una sola proporción. Por ejemplo, podemos emplearla para el cálculo de probabilidades de conjuntos interesantes:

Conjuntos rectangulares

$[c_1, d_1] \times [c_2, d_2]$

Utilizando la independencia de la distribución a posteriori, se tiene

$$\int_{c_1}^{d_1} \int_{c_2}^{d_2} f\left(p_1, p_2 | x_1, x_2\right) dp_1 dp_2 =$$
$$= \left[\int_{c_1}^{d_1} f\left(p_1 | x_1, x_2\right) dp_1\right] \times \left[\int_{c_2}^{d_2} f\left(p_2 | x_1, x_2\right) dp_2\right],$$

y para calcular cada una de las probabilidades aplicamos los métodos ya vistos.

Conjuntos triangulares

Se está más interesado en calcular probabilidades a posteriori de conjuntos triangulares, y más precisamente, de conjuntos triangulares de la forma $(p_1 - p_2) \geq c$. La probabilidad se puede calcular numéricamente y es igual a

$$\int \int_{\{p_1 - p_2 \geq c\}} f\left(p_1, p_2 | x_1, x_2\right) dp_1 dp_2.$$

Alternativamente, se puede emplear la normalidad asintótica: si n_1 y n_2 son suficientemente grandes

$$p_1 | x_1, x_2 \sim N\left(\frac{x_1}{n_1}, \frac{x_1}{n_1}\left(1 - \frac{x_1}{n_1}\right)\frac{1}{n_1}\right)$$
$$p_2 | x_1, x_2 \sim N\left(\frac{x_2}{n_2}, \frac{x_2}{n_2}\left(1 - \frac{x_2}{n_2}\right)\frac{1}{n_2}\right)$$

y son independientes. Se tiene, por las propiedades de las distribuciones normales, que

$$p_1 - p_2 | x_1, x_2 \sim N\left(\frac{x_1}{n_1} - \frac{x_2}{n_2}, \frac{x_1}{n_1}\left(1 - \frac{x_1}{n_1}\right)\frac{1}{n_1} + \frac{x_2}{n_2}\left(1 - \frac{x_2}{n_2}\right)\frac{1}{n_2}\right)$$

y el cálculo de la probabilidad pedida es

$$1 - \phi\left(\frac{c - \left(\frac{x_1}{n_1} - \frac{x_2}{n_2}\right)}{\sqrt{\frac{x_1}{n_1}\left(1 - \frac{x_1}{n_1}\right)\frac{1}{n_1} + \frac{x_2}{n_2}\left(1 - \frac{x_2}{n_2}\right)\frac{1}{n_2}}}\right).$$

Así, un intervalo de probabilidad a posteriori $(1 - \alpha)$ sería

$$\left(\frac{x_1}{n_1} - \frac{x_2}{n_2}\right) \pm z_{\frac{\alpha}{2}} \sqrt{\frac{x_1}{n_1}\left(1 - \frac{x_1}{n_1}\right)\frac{1}{n_1} + \frac{x_2}{n_2}\left(1 - \frac{x_2}{n_2}\right)\frac{1}{n_2}}.$$

Otra posibilidad es emplear simulación. Generamos k observaciones $\left(p_i^1, p_i^2, \ldots p_i^k\right)$ de $p_i|x_1, x_2$, para $i = 1, 2$ y se aproxima la probabilidad mediante

$$\frac{\text{cardinal}\left\{p_1^i - p_2^i \geq c\right\}}{k}.$$

Mediante simulación podemos generar, por ejemplo, 1000 observaciones de p_1 y p_2 (p_j^i, $j \in \{1, 2\}$, $i \in \{1, 2, \ldots, 1000\}$), ordenar los valores $r_i = p_1^i - p_2^i$ y emplear el intervalo $\left[r_{(50)}, r_{(950)}\right]$ para dar un intervalo aproximado de probabilidad de 0.9.

Por último, se puede contrastar hipótesis, por ejemplo,

$$H_0 : p_1 - p_2 = 0$$

frente a

$$H_1 : p_1 - p_2 \neq 0.$$

Para ello, calculamos un intervalo de probabilidad a posteriori igual 0.95; si incluye a 0, decimos que los datos justifican la hipótesis de que ambas proporciones son iguales.

Ejemplo

Un grupo de investigadores evaluó si un grupo de personas en la república del Congo tenía el virus del *sida* y registró la frecuencia con que habían empleado preservativos el año anterior. La tabla recoge el número de personas que pasaron de seronegativos a seropositivos, en función del uso de preservativos.

Uso	*Nunca*	<25 %	26-50 %	50-75 %	>75 %
Sero+	74	19	7	0	0
Sero-	214	36	15	2	6
Total	288	55	22	2	6

Agrupamos las observaciones entre personas que usan muy frecuentemente preservativos (≥75 %) y las que no lo usan (<75 %)

	<75 %	≥75 %
Sero+	100	0
Sero-	267	6
	367	6

La cuestión que nos planteamos es si el uso alto de preservativos previene la transmisión del virus del *sida.* Consideramos distribuciones a priori $Be\,(1, 1)$ independientes para ambas proporciones p_{SI}, p_{NO} y calculamos un intervalo de probabilidad a posteriori 0.95 para $d = p_{NO} - p_{SI}$.

La densidad a posteriori para p_{NO} es $Be\,(101, 268)$ y para p_{SI} es $Be\,(1, 7)$. Las medias y varianzas a posteriori son, respectivamente,

$$\begin{aligned} E\,(p_{NO}|\text{datos}) &= \frac{101}{369} = 0{,}2737 \\ Var\,(p_{NO}|\text{datos}) &= 0{,}0005 \\ E\,(p_{SI}|\text{datos}) &= \frac{1}{8} = 0{,}125 \\ Var\,(p_{SI}|\text{datos}) &= 0{,}0121, \end{aligned}$$

con lo que

$$\begin{aligned} E\,(p_{NO} - p_{SI}|\text{datos}) &= 0{,}149 \\ Var\,(p_{NO} - p_{SI}|\text{datos}) &= 0{,}0127. \end{aligned}$$

Suponiendo válido el argumento asintótico resulta que, aproximadamente,

$$p_{NO} - p_{SI}|\text{datos} \sim N\,(0{,}149, 0{,}0127)\,.$$

Como, para la $N\,(0, 1)$, $P\,(-1{,}96, 1{,}96) = 0{,}95$, un intervalo de probabilidad 0.95 es

$$\begin{aligned} &\left[0{,}149 - 1{,}96\sqrt{0{,}0127}, 0{,}149 + 1{,}96\sqrt{0{,}0127}\right] = \\ &= [-0{,}07, 0{,}37] \end{aligned}$$

y $0 \in [-0{,}07, 0{,}37]$.

Con esto ya se ha respondido a la pregunta formulada, pero podemos completar el argumento.

Como $P\,(-1, 1) = 0{,}68$, un intervalo de probabilidad 0.68 es

$$\begin{aligned} &\left[0{,}149 - 1\sqrt{0{,}0127}, 0{,}149 + 1\sqrt{0{,}0127}\right] = \\ &= [0{,}04, 0{,}26] \end{aligned}$$

y $0 \notin [0{,}04, 0{,}26]$

Además, $P\,(p_{NO} \leq p_{SI}) = P\,(p_{NO} - p_{SI} \leq 0) = \Phi\left(\frac{0-0{,}149}{\sqrt{0{,}0127}}\right) = \Phi\,(-1{,}318) = 1 - \Phi\,(1{,}318) = 0{,}1$.

Así, la hipótesis de que usar los preservativos es efectivo tiene algo de apoyo, pero no mucho.

Procedemos en forma similar mediante simulación. Generamos 1000 observaciones de una $Be\,(101, 268)$ y de una $Be\,(1, 7)$. Se obtienen las diferencias, se ordenan de menor a mayor y se calcula el intervalo correspondiente. Para una probabilidad del 0.95 se tiene

$$\left[z_{(25)}, z_{(975)}\right] = [-0{,}0981, 0{,}2875]\,.$$

Para probabilidad 0.68, se tiene

$$\left[z_{(160)}, z_{(840)}\right] = [0{,}0436, 0{,}2197]$$

y

$$P\left(p_{NO} - p_{SI} \leq 0 | \text{datos}\right) = \frac{\text{cardinal}\left\{z \leq 0\right\}}{1000} = \frac{106}{1000} = 0{,}106.$$

```
# Ejemplo
# Se tiene que p1 sigue una Be(101,268)
# Se tiene que p2 sigue una Be(1,7)

# Simulo dos distribuciones a posteriori
p1 = rbeta(1000,101,268)
p2 = rbeta(1000,1,7)

# Calculo la diferencia entre ambas muestras
dif = p1 - p2

# Calculo la probabilidad de los valores que son >=0
difmas0 = dif[dif >= 0]
(prob = length(difmas0)/1000)
```

```
[1] 0.893
```

```
# Calculo los cuantiles del 5% y el 95%
# Esto equivale a un intervalo del 0.90
quantile(dif,c(0.05,0.95))
```

```
         5%         95%
-0.09852904  0.27579033
```

```
# Calculo los cuantiles del 2.5% y el 97.5%
# Esto equivale a un intervalo del 0.95
quantile(dif,c(0.025,0.975))
```

```
      2.5%      97.5%
-0.1813385  0.2886177
```

```
# O bien ordeno valores y calculo un intervalo de 0.90
ordedif = sort(dif)

ordedif[50]; ordedif[950]
```

```
[1] -0.1007258
[1] 0.2757859
```

```
ordedif[100]; ordedif[900]
```

```
[1] -0.008156479
[1] 0.2637232
```

```
ordedif[150]; ordedif[850]
```

```
[1] 0.03205061
[1] 0.2512118
```