Estadística Bayesiana: Clase 4

Juan Carlos Correa

9 de marzo de 2021

Estadística bayesiana cuando se pueden conseguir datos muestrales

- Esta es la situación que se asume en la mayoría de los textos de estadística bayesiana.
- Se puede realizar un experimento que genere datos bajo una distribución bien definida, por ejemplo Normal, Poisson, etc.
- Muchos bayesianos asumen que esta información es la más importante!

Teorema de Bayes

El Teorema de Bayes es ahora una de las piedras fundamentales del trabajo estadístico y sigue siendo de cierta discusiones tanto de sus orígenes como de sus implicaiones filosóficas (Dawid, 2004). Este teorema fue publicado varios años después de la muerte de reverendo Thomas Bayes por un amigo.

(Teorema de Bayes)

Sean $B_1, B_2, \cdots, B_k$ eventos mutuamente excluyentes y exhaustivos. Para cualquier evento nuevo A, tenemos

$$P\left(B_i|A\right) = \frac{P\left(B_i \bigcap A\right)}{P(A)} = \frac{P\left(A|B_i\right)P\left(B_i\right)}{\sum_{i=1}^{k} P\left(A|B_i\right)P\left(B_i\right)}$$

Prueba: (Ejercicio)

(Teorema de Bayes para Variables Aleatorias) Sean X y θ variables aleatorias con fdp's $f\left(x|\theta\right)$ y $\xi(\theta)$.

$$\xi\left(\theta|x\right) = \frac{f\left(x|\theta\right)\xi(\theta)}{\int_{\Theta} f\left(x|\theta\right)\xi(\theta)\,d\theta}$$

Dentro del marco bayesiano tenemos que:

- X : Datos (escalar o vector o matriz)
- θ: Parámetro desconocido (escalar o vector o matriz)
- $f\left(x_1, \cdots, x_n|\theta\right)$: Función de probabilidad conjunta de los datos dado el parámetro (desconocido) θ.
- $\xi(\theta)$: Distribución apriori de θ.

Problema!

Esta aproximación nos exige tener formas paramétricas para

- la distribución apriori y
- la distribución que genera los datos!!

Aunque existe trabajo no paramétrico en estadística bayesiana, es difícil y exige el manejo de procesos que no son intuitivos y su aplicación es muy difícil a problemas reales!

(Otra forma de representar el Teorema de Bayes para Variables Aleatorias) Sean X y θ variables aleatorias con fdp's $f\left(x|\theta\right)$ y $\xi(\theta)$.

$$\xi\left(\theta|x\right) = \frac{L\left(\theta|x\right)\xi(\theta)}{\int_{\Theta} L\left(\theta|x\right)\xi(\theta)\,d\theta}$$

Dentro del marco bayesiano tenemos que:

- X : Datos (escalar o vector o matriz)
- θ: Parámetro desconocido (escalar o vector o matriz)
- $L\left(x_1, \cdots, x_n|\theta\right)$: Verosimilitud de los datos dado el parámetro (desconocido) θ.
- $\xi(\theta)$: Distribución apriori de θ.

Por el teorema anterior

$$\xi\left(\theta|x_1,\cdots,x_n\right)=\frac{L\left(x_1,\cdots,x_n|\theta\right)\xi(\theta)}{\int_\Theta L\left(x_1,\cdots,x_n|\theta\right)\xi(\theta)\,d\theta}$$

Esta es llamada la **distribución posterior**. La inferencia bayesiana se deriva de esta distribución. En la práctica, el denominador de la expresión anterior no necesita ser calculado en general, y la regla de Bayes se escribe como

$$\xi\left(\theta|x_1,\cdots,x_n\right)\propto L\left(x_1,\cdots,x_n|\theta\right)\xi(\theta)$$

Por lo tanto solo necesitamos conocer la distribución posterior hasta una constante de normalización (no necesitamos tener el denominador!).

$$\propto \longleftarrow \text{denota 'Proporcional a'}$$

Muchas veces somos capaces de identificar la distribución posterior de θ mirando solamente el numerador.

El teorema de Bayes lo que hace es una “actualización” de $\xi(\theta)$ a $\xi\left(\theta|x_1,\cdots,x_n\right)$.
Nota: El aprendizaje bayesiano será

$$
\begin{aligned}
\xi\left(\theta|x_1\right) &\propto L\left(x_1|\theta\right)\xi(\theta) \\
\xi\left(\theta|x_1,x_2\right) &\propto L\left(x_2|\theta\right)L\left(x_1|\theta\right)\xi(\theta) \\
&\propto L\left(x_2|\theta\right)\xi\left(\theta|x_1\right)
\end{aligned}
$$

Por lo tanto el teorema de Bayes nos muestra cómo el conocimiento acerca del estado de la naturaleza representada por θ es continuamente modificada a medida que nuevos datos muestrales son adquiridos.

Distribución Binomial

Resultado:

- Supongamos que la distribución apriori de π es una beta con parámetros $\alpha(> 0)$ y $\beta(> 0)$.
- Suponga que $X_1, \cdots, X_n$ es una muestra aleatoria de una distribución Bernoulli con parámetro π, donde el valor de π es desconocido.
- Entonces la distribución posterior de π cuando $X_i = x_i$, para $i = 1, \cdots, n$ es una beta con parámetros $\alpha + \sum_{i=1}^{n} x_i$ y $\beta + n - \sum_{i=1}^{n} x_i$.

Repaso distribución Beta

Si $W \sim Beta\left(\alpha, \beta\right)$

$$\begin{aligned}
f\left(w\left|\alpha, \beta\right.\right) &= \frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)} w^{\alpha-1}\left(1-w\right)^{\beta-1} \\
E\left(W\right) &= \frac{\alpha}{\alpha+\beta} \\
Moda\left(W\right) &= \frac{\alpha-1}{\alpha+\beta-2} \\
var\left(W\right) &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}
\end{aligned}$$

Prueba:

- Sean $X_1, \cdots, X_n$ variables aleatorias independientes Bernoulli(π).
- La verosimilitud es

$$L(\theta) \propto \pi^{\sum_i X_i}(1-\pi)^{n-\sum_i X_i}$$

- El parámetro π es univariable, y restringido al intervalo $[0, 1]$.
- La distribución apriori conjugada será

$$\xi(\pi) \propto \pi^{\alpha-1}(1-\pi)^{\beta-1}, \text{ con } \alpha, \beta > 0$$

α y β son llamados hiperparámetros. Esta palabra se utiliza para distiguirlos del parámetro modelo muestral π.

- La distribución posterior es

$$\xi(\pi|X_1,\cdots,X_n) \propto \xi(\pi)L(\theta|Datos)$$

$$\xi(\pi|X_1,\cdots,X_n) \propto \pi^{\alpha+\sum_i X_i-1}(1-\pi)^{\beta+n-\sum_i X_i-1}$$

la cual es una distribución beta con hiperparámetros $\alpha + \sum_i X_i$ y $\beta + n - \sum_i X_i$.

La esperanza apriori $E(\pi)$ corresponde a la probabilidad marginal de tener un éxito antes de obtener cualquier observación:

$$E(\pi) = \int \pi\xi(\theta)\, d\pi = \int p(Y = 1|\pi)\xi(\pi)\, d\pi = p(X = 1)$$

Ya que la varianza de π es una función decreciente de $\alpha + \beta$ para una media dada, la suma de los hiperparámetros $\alpha + \beta$ es también llamada la *precisión* de la distribución.

La media aposteriori se puede expresar como

$$\frac{\alpha + \sum_{i=1}^{n} X_i}{\alpha + \beta + n} = \left(\frac{\alpha + \beta}{\alpha + \beta + n}\right)\left(\frac{\alpha}{\alpha + \beta}\right) + \left(\frac{n}{\alpha + \beta + n}\right)\left(\frac{\sum_{i=1}^{n} X_i}{n}\right)$$

lo que es una media ponderada

$$E\left(\pi | X_1, \cdots, X_n, \alpha, \beta\right) = w \cdot E(\pi|\alpha, \beta) + (1 - w) \cdot \frac{\sum_{i=1}^{n} X_i}{n}$$

donde $w = (\alpha + \beta)/(\alpha + \beta + n)$.

Elicitación de los Parámetros de la Beta para Proporciones

```
##############################################################
# Elicitación de la distribución apriori de una proporción
# Se asume que la apriori es una Beta(alfa,beta)
# El procedimiento consiste en determinar los límites más extremos p
# o sea, donde estamos casi convencidos cae la verdadera proporción
# de la moda.
# Programador JCCM
# Fecha de última modificación: marzo 8 2021

conseguir.beta.apriori<-function(){
# Entrada de la información del elicitador
LI<-as.numeric(readline("Cuál es el menor valor posible para
la proporción? "))
LS<-as.numeric(readline("Cuál es el mayor valor posible para
la proporción? "))
moda<-as.numeric(readline("Cuál es el valor más posible para
la proporción? "))
```

```
funcion.a.mini<-function(parametros,moda=0.5,mini=0.3,maxi=0.6){
a1<-parametros[1]
b1<-parametros[2]

temp<-((a1-1)/(a1+b1-2)-moda)^2+(pbeta(mini,a1,b1)-
0.01)^2+(pbeta(maxi,a1,b1)-0.99)^2
return(temp)
}
res<-optim(c(2,2),funcion.a.mini,moda=moda,mini=LI,maxi=LS,
lower=c(1,1),method='L')
parametros<-res$par
alfa<-parametros[1]
beta<-parametros[2]
x<-seq(0.01,0.99,length=200)
y<-dbeta(x,alfa,beta)
plot(x,y,type='l',ylab='Densidad',xlab=expression(pi),
main='Apriori Beta para la Prporción')
abline(v=c(LI,moda,LS),lty=2)
return(res)
} # Fin conseguir.beta.apriori
```

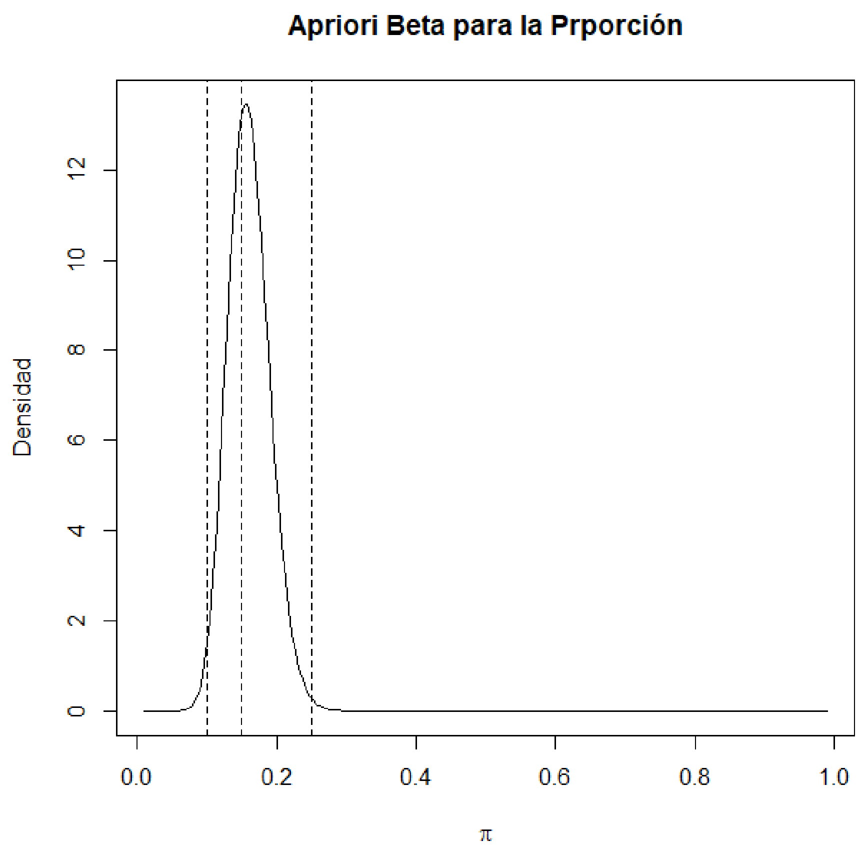
```
> resultados<-conseguir.beta.apriori()
Cuál es el menor valor posible para la proporción? 0.10
Cuál es el mayor valor posible para la proporción? 0.25
Cuál es el valor más posible para la proporción? 0.15
> resultados
> resultados
$par
[1]  24.20785 127.07272

$value
[1] 8.441759e-05

$counts
function gradient
      35       35

$convergence
[1] 0

$message
[1] "CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH"
```


Apriori Beta para la Prporción
Densidad
0
2
4
6
8
10
12
0.0
0.2
0.4
0.6
0.8
1.0
π

Ahora supongamos que sacamos una muestra de estudiantes de primer semestre y les preguntamos si en la actualidad fuman tabaco (cigarrillos, cigarros, etc.)

Fuman	Sí	No	n
tabaco?	12	58	70

Apriori

$$\xi\left(\pi\right) \propto \pi^{24{,}20785-1}\left(1-\pi\right)^{127{,}07272-1}$$

Verosimilitud

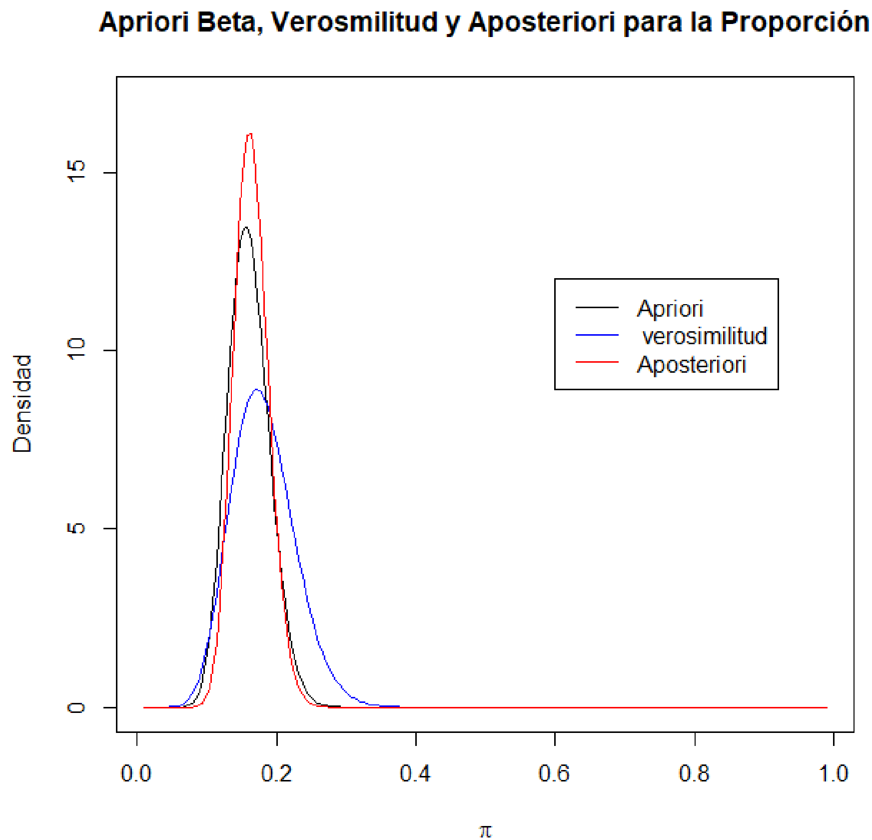
$$L\left(\pi\left|Datos\right.\right) \propto \pi^{12}\left(1-\pi\right)^{58}$$

Aposteriori

$$\xi\left(\pi\left|Datos\right.\right) \propto \pi^{24{,}20785+12-1}\left(1-\pi\right)^{127{,}07272+58-1}$$

Medidas de resumen

Promedio	valor
Media apriori	24.20785/(24.20785+127.07272)=0.1600196
Media muestral	12/70= 0.1714286
Media aposteriori	(24.20785+12)/(24.20785+12+(127.07272+58))=0.1636287


Apriori Beta, Verosmilitud y Aposteriori para la Proporción
Densidad
0
5
10
15
0.0
0.2
0.4
0.6
0.8
1.0
π
Apriori
verosimilitud
Aposteriori

```
alfa<- 24.20785
beta<-127.07272
x<-seq(0.01,0.99,length=200)
y<-dbeta(x,alfa,beta)
plot(x,y,ylim=c(0,17),type='l',ylab='Densidad',xlab=expression(pi),
main='Apriori Beta, Verosmilitud y Aposteriori para la Proporción')

alfa<-12+1
beta<-58+1
x<-seq(0.01,0.99,length=200)
y<-dbeta(x,alfa,beta)
points(x,y,type='l',col='blue')

alfa<-12+ 24.20785
beta<-58+127.07272
x<-seq(0.01,0.99,length=200)
y<-dbeta(x,alfa,beta)
points(x,y,type='l',col='red')

legend(0.6,12, legend = c("Apriori ", " verosimilitud",
"Aposteriori "),lty=c(1,1,1),col=c('black','blue','red'))
```