

01) Introducción a la inferencia bayesiana

Eduardo García Tapia

1.1] Proceso de aprendizaje bayesiano

Recuerde que en el enfoque frecuentista se considera una función de verosimilitud $p(x|\theta)$, la cual se busca maximizar respecto a θ para lograr el estimador que haga más probable el haber observado la muestra (x_1, x_2, \dots, x_n) . Esto representa un problema ya que no se pueden hablar de rangos de probabilidad para θ , solo de confianza. Igualmente, no es posible incorporar información externa adicional a la muestra observada.

El enfoque bayesiano, por su parte, hace inferencias sobre los parámetros a partir de las leyes y propiedades de la probabilidad. El teorema de Bayes sirve como regla de aprendizaje:

$$p(\theta|x) = \frac{p(\theta)p(x|\theta)}{\int p(\theta)p(x|\theta) dx}$$

Sin embargo, se puede demostrar que la proporcionalidad en funciones de distribución de probabilidad es útil para hacer inferencia sobre éstas. Entonces,

$$p(\theta|x) \propto p(\theta)p(x|\theta)$$

Observer que al seguir las reglas de la probabilidad θ no se ve más como un valor fijo, sino que sigue su propia función de distribución de probabilidad, la cual nos permitirá hablar de intervalos de probabilidad, al mismo tiempo que al definir su forma inicial, se puede integrar información externa a la función de verosimilitud.

El enfoque bayesiano permite seguir un procedimiento más lógico sobre el proceso de inferencia. La única solución bayesiana a cualquier problema es:

Encontrar la distribución condicional de todas aquellas cantidades de interés cuyo valor desconocemos, dado el valor conocido de las variables observadas.

Con el marco teórico establecido, el proceso de aprendizaje bayesiano puede describirse en 4 pasos:

1. Especificación de un modelo muestral: $p(\theta|x)$
2. Especificación de una distribución inicial: $p(\theta)$
3. Cálculo de la distribución final: $p(\theta|x) \propto p(\theta)p(x|\theta)$
4. Resumen de la información obtenida y hacer inferencias sobre las cantidades de interés

1.2] Ejemplo distribución binomial

Para este ejemplo suponga un fenómeno que al ser observado sigue una distribución binomial.

$$Bin(\theta, n) \sim P(x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$$

Se observa que las combinaciones no son funciones de θ , por lo que podemos utilizar la proporcionalidad para establecer la función de distribución muestral

$$p(x|\theta) \sim Bin(\theta, n) \propto \theta^x (1 - \theta)^{n-x}$$

De los libros de texto se sabe que la distribución $Beta(a, b)$ es “conjugada” de la binomial, lo que quiere decir que los cálculos aritméticos son fáciles cuando se asume que la distribución inicial tiene la forma;

$$p(\theta) \sim \text{Beta}(\theta|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1} \propto \theta^{a-1} (1-\theta)^{b-1}$$

Observe que las proporcionalidades de la distribución muestral e inicial son similares, por lo tanto, la regla de actualización se calcula como:

$$p(\theta)p(x|\theta) \propto \theta^{x+a-1} (1-\theta)^{n-x-b-1}$$

Por lo tanto:

$$p(\theta|x) \propto \text{beta}(\theta|x+a, n-x+b)$$

A partir de este resultado, se puede conocer la distribución del parámetro θ de una distribución muestral binomial, cuando se conocen las observaciones $(x_1, x_2, x_3, \dots, x_n)$. Observe que a y b nos permiten agregar información a la distribución final, a es similar al número de éxitos y b al número de fracasos en un experimento con $n = a + b$ observaciones ficticias.

Los valores de a y b son definidos por el estadístico al incorporar información de un experto en el área, la cual debe ser traducida para definir una distribución inicial. En el caso en el que no se conozca información adicional los hiperparámetros se establecen $a = b = 0$, de esta manera no se añade un experimento ficticio a la distribución final. La especificación de los hiperparámetros se le conoce como no informativa.

Ejemplo: Se tiene una

```
# Distribución muestral Binomial.
x = 3 # Número de éxitos
n = 10 # Número de ensayos

# Distribución no informativa
a = 9
b = 10

# Particionar 0 a 1 para calcular la inversa de la probabilidad
secuencia = seq(0,1,length.out=100)

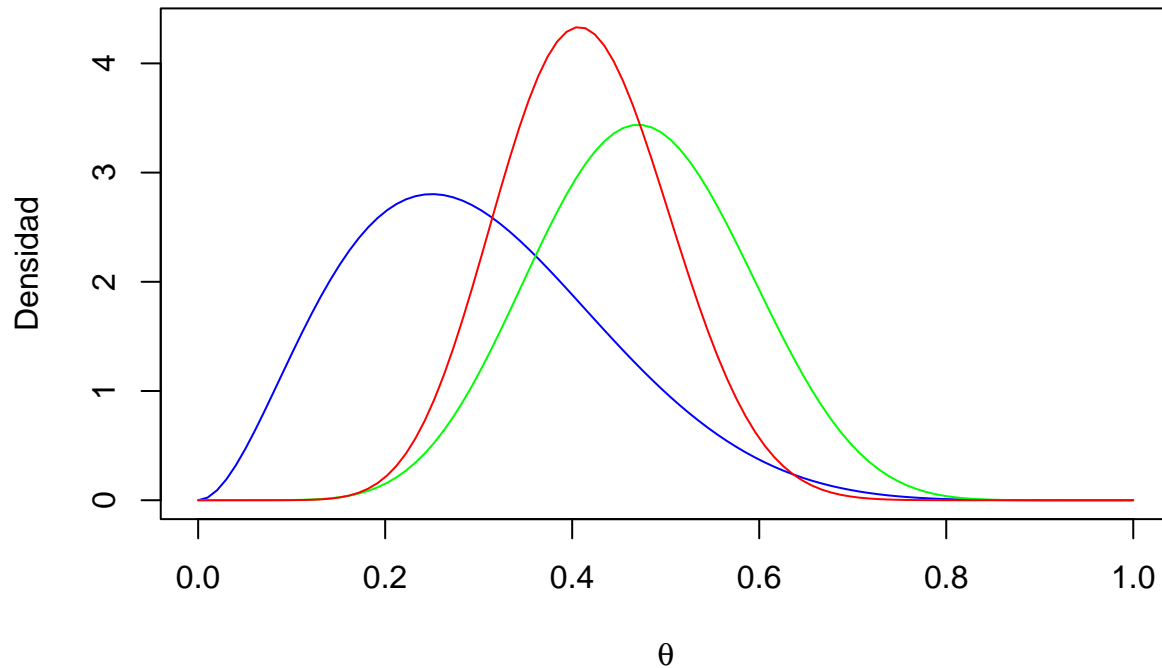
# Solución frecuentista
distr_max_vero = dbeta(secuencia, x, n-x)

# Distribución inicial de theta
distr_ini = dbeta(secuencia, a, b)

# Distribución final
distr_fin = dbeta(secuencia, x+a, n-x+b)

# Generar gráfica
y_max = max(distr_max_vero, distr_ini, distr_fin)
plot(secuencia, distr_max_vero, type="l", col="blue", ylim=c(0,y_max),
      xlab=expression(theta),ylab="Densidad")
lines(secuencia, distr_ini, col="green")
lines(secuencia, distr_fin, col="red")
title("Binomial: Azul:Max. Vero. | Verde:Inicial | Rojo:Final")
```

Binomial: Azul:Max. Vero. | Verde:Inicial | Rojo:Final



1.3] Reparametrización

Cuando se trabaja con distribuciones binomiales como distribución muestral, en ocasiones se busca trabajar con alguna reparametrización de θ , por ejemplo, el logaritmo de los momios $\omega = \ln\left(\frac{\theta}{1-\theta}\right)$. Debido a que ya se conoce la función de distribución de probabilidad del parámetro, conocer la PDF de la reparametrización es una tarea trivial, pues se puede generar una muestra del parámetro y aplicarle la transformación a esos datos.

Recuerde que los momios pueden tomar valores en $(-\infty, \infty)$. En la figura debajo se observa que la reparametrización, en efecto, sirvió para poder transformar el dominio del parámetro $(0, 1)$. A partir de esta distribución generada aleatoriamente se pueden hacer inferencias sobre el valor de ω . El proceso de generar distribuciones finales a partir de simulación es la base de la inferencia bayesiana, debido a que muchas de las soluciones del proceso de aprendizaje no son cerradas, las integrales no pueden ser resueltas analíticamente, por lo que solo se pueden obtener distribuciones finales de este modo.

```
# Generar una muestra de la distribución final
N = 10000
muestra = rbeta(N, x+a, n-x+b)

# Aplicar reparametrización a la muestra
muestra_omega = log(muestra / (1 - muestra))

# Graficar distribución de la reparametrización
plot(density(muestra_omega, adjust=1.65), type="l", main="Distribución final de logaritmo de momios",
      xlab="Valor", ylab="Densidad")
```

Distribución final de logaritmo de momios

