Página www

Página de Abertura

Contenido

Página 1 de 27

Regresar

Full Screen

Cerrar

Abandonar

Estadística Bayesiana: Clase 7

Juan Carlos Correa

17 de marzo de 2021

Tarea 3: Mezclando las distribuciones aprioris informativas de tres expertos

Pídale a tres (3) personas del mismo sexo y de aproximadamente la misma edad que le colaboren con la plantilla de le edad ideal del novio de una mujer de 22 años. Resuma estas tres distribuciones en una sola distribución usando una simulación parecida a la que está a continuación:

```
# Combinación de dos distribuciones apriori sobre el mismo parámetro
# Espacio parametral discretizado

# problema de la edad ideal del novio para una
# mujer de 22 años

datos.gráfico.a.mano.alzada.1<-scan()
10 0
15 1
20 15
25 50
30 40
35 15
40 10
45 5
50 1
55 0

experto1<-matrix(datos.gráfico.a.mano.alzada.1,ncol=2,byrow=T)
```

```
datos.gráfico.a.mano.alzada.2<-scan()
10 0
15 5
20 20
25 60
30 60
35 50
40 15
45 10
50 5
55 2

experto2<-matrix(datos.gráfico.a.mano.alzada.2,ncol=2,byrow=T)
```

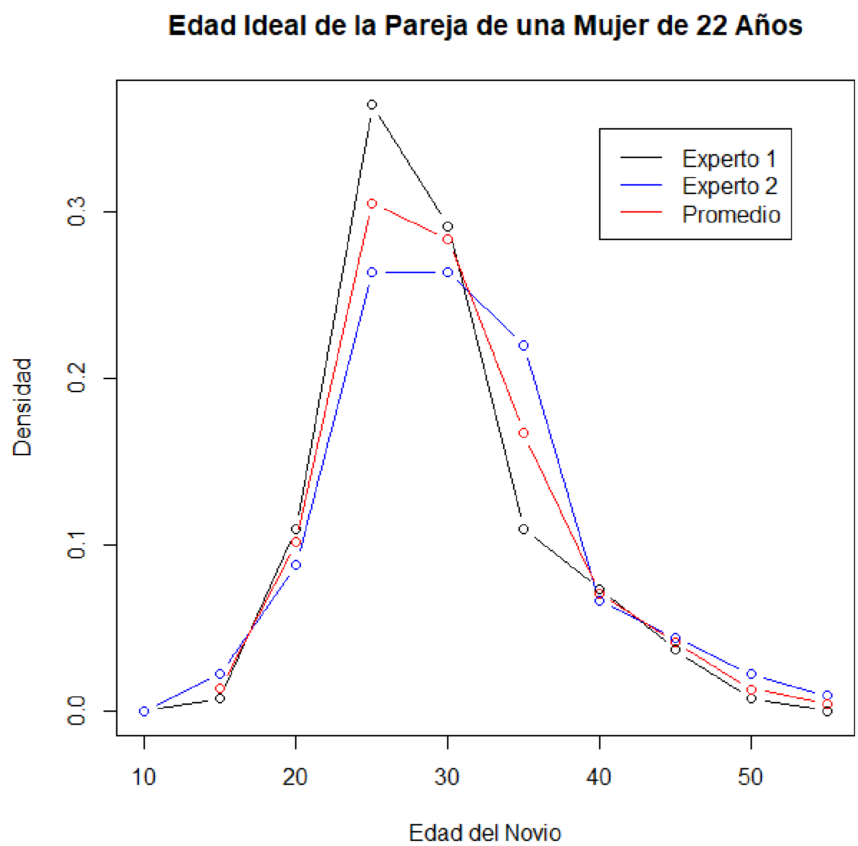
```
Nsim.1<-10000
muestra1<-sample(experto1[,1],Nsim.1,replace=T,prob=experto1[,2])

Nsim.2<-10000
muestra2<-sample(experto2[,1],Nsim.2,replace=T,prob=experto2[,2])

res1<-prop.table(experto1[,2])
res2<-prop.table(experto2[,2])
res3<-prop.table(table(c(muestra1,muestra2)))

edad<-seq(from=10,to=55,by=5)

plot(edad,res1,type='b',ylab='Densidad',xlab='Edad del Novio')
title(main='Edad Ideal de la Pareja de una Mujer de 22 Años')
points(edad,res2,type='b',col='blue')
points(sort(unique(c(muestra1,muestra2))),res3,type='b',col='red')
legend(40,0.35,c('Experto 1','Experto 2','Promedio'),lty=c(1,1,1),
col=c('black','blue','red'))
```


Edad Ideal de la Pareja de una Mujer de 22 Años
Experto 1
Experto 2
Promedio
Densidad
0.0
0.1
0.2
0.3
10
20
30
40
50
Edad del Novio

Apriori no informativa de Jeffreys para Caso normal

Asumamos que $y_1, \cdots, y_n|\mu,\ \sigma^2$ son variables distribuidas normal e independientemente con media μ y con varianza σ^2 desconocidas. calculemos la distribución apriori de Jeffreys para (μ, σ)

$$
\begin{aligned}
f(x|\mu,\sigma) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \\
\log(f(x|\mu,\sigma)) &= -\frac{1}{2}\log(2\pi) - \log(\sigma) - \frac{1}{2\sigma^2}(x-\mu)^2 \\
\frac{\partial \log(f(x|\mu,\sigma))}{\partial\mu} &= \frac{1}{\sigma^2}(x-\mu) \\
\frac{\partial^2 \log(f(x|\mu,\sigma))}{\partial\mu^2} &= -\frac{1}{\sigma^2} \\
\frac{\partial \log(f(x|\mu,\sigma))}{\partial\sigma} &= -\frac{1}{\sigma} + \frac{1}{\sigma^3}(x-\mu)^2 \\
\frac{\partial^2 \log(f(x|\mu,\sigma))}{\partial\sigma^2} &= \frac{1}{\sigma^2} - \frac{3}{\sigma^4}(x-\mu)^2 \\
\frac{\partial^2 \log(f(x|\mu,\sigma))}{\partial\mu\partial\sigma} &= -\frac{2}{\sigma^3}(x-\mu)
\end{aligned}
$$

Tomando la esperanza obtenemos

$$\mathcal{I}\left(\begin{pmatrix} \mu \\ \sigma \end{pmatrix}\right) = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{bmatrix}$$

Así la distribución apriori será

$$\begin{aligned} \xi(\mu,\sigma) &\propto \left|\mathcal{I}\left(\begin{pmatrix} \mu \\ \sigma \end{pmatrix}\right)\right|^{1/2} \\ &= \left(\frac{1}{\sigma^2} \times \frac{2}{\sigma^2}\right)^{1/2} \\ &\propto \frac{1}{\sigma^2} \end{aligned}$$

Esta distribución apriori de Jeffreys es impropia.

Distribuciones No Informativas: Otras Alternativas

Distribución Apriori de Máxima Entropía

Cuando θ es univariable y puede tomar cualquier valor sobre la recta real, y la media y la varianza apriori están especificadas, la distribución apriori de máxima entropía es la Normal con la media y la varianza especificadas.

Kass y Wasserman (1994) presentan la definición planteada por Novick y Hall:

Distribución Apriori Indiferente

Se define una distribución apriori indiferente si identificando una clase de conjugadas se selecciona una apriori de esta clase que satisfaga:

- La apriori debe ser impropia y
- una “muestra mínima necesaria” debe inducir una posterior propia.

Un ejemplo de la anterior definición es claro en el problema binomial, con la clase conjugada de las Betas, la distribución apriori $\{\pi(1-\pi)\}^{-1}$ es una apriori indiferente.

Box y Tiao (1973) proponen el uso de distribuciones *apriori localmente uniformes*, las cuales consideran el comportamiento local de la apriori en una región donde la verosimilitud es apreciable, pero la apriori no se asume grande por fuera de esa región.

Marginalización

Eliminación un término *de molestia* o parámetro *nuisance*

- En muchas situaciones tenemos un vector de parámetros, pero solo estamos interesados realmente en unos pocos.
- Debemos por lo tanto proceder a “eliminar” aquellos términos de molestia.
- Esto lo hacemos mediante la marginalización[a].

[a] En estadística clásica se hace vía el principio de condicionalidad

- Suponga que $x_1, \cdots, x_n$ es una muestra aleatoria de una $N\left(\mu, \sigma^2\right)$,
- donde $\left(\mu, \sigma^2\right)$ son desconocidos.
- Sea $\tau = 1/\sigma^2$.
- Suponga que especificamos una apriori no informativa de Jeffreys

$$\xi\left(\mu, \sigma^2\right) \propto \tau$$

- Ahora,

$$\xi\left(\mu, \tau \left| x\right.\right) \propto \tau^{\frac{n}{2}-1} \exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}\left(x_i - \mu\right)^2\right\}.$$

Así, para eliminar el término nuisance τ marginalizamos

$$\xi\left(\mu\,|x\right) \propto \int_0^\infty \tau^{\frac{n}{2}-1} \exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2\right\}\,d\tau.$$

No es difícil llegar a

$$\xi\left(\mu\,|x\right) \propto \int_0^\infty \tau^{\frac{n}{2}-1} \exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2\right\} \exp\left\{-\frac{n\tau}{2}\left(\bar{x}-\mu\right)^2\right\}\,d\tau.$$

Sea

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Entonces

$$\begin{aligned}
\xi(\mu\,|x) &\propto \int_0^\infty \tau^{\frac{n}{2}-1}\exp\left\{-\frac{\tau}{2}\left((n-1)s^2 + n(\mu-\bar{x})^2\right)\right\} d\tau \\
&\propto \left((n-1)s^2 + n(\mu-\bar{x})^2\right)^{-n/2} \\
&\propto \left(1 + \frac{n}{(n-1)s^2}(\mu-\bar{x})^2\right)^{-(n-1+1)/2}
\end{aligned}$$

Así

$$\mu|x \sim t\left(n-1, \bar{x}, \frac{s^2}{n}\right)$$

Por lo tanto

$$\frac{\mu-\bar{x}}{s/\sqrt{n}} \sim t_{(n-1)}$$

A pesar de haber llegado a un resultado que es de uso común en la estadística clásica, la interpretación aquí es diferente.

Eliminando otro término *de molestia*

En el ejemplo anterior supongamos que el término de molestia es μ. Debemos por lo tanto halla $\xi\left(\tau|x\right)$. procedemos de manera similar

$$\begin{aligned}\xi\left(\tau|x\right) &\propto \int_{-\infty}^{\infty} \tau^{\frac{n}{2}-1} \exp\left\{-\frac{\tau}{2}\left((n-1)s^2+n(\mu-\bar{x})^2\right)\right\} d\mu \\ &\propto \tau^{\frac{n-1}{2}-1} \exp\left\{-\frac{\tau}{2}\left((n-1)s^2\right)\right\}\end{aligned}$$

Así

$$\tau | x \sim Gamma\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$$

De lo anterior obtenemos que

$$(n-1)s^2\tau \sim \chi^2_{n-1}$$

Distribuciones Conjugadas

Dada la magnitud de la tarea de determinar una distribución apriori que refleje de una manera clara nuestra información bayesiana, uno intuitivamente piensa en limitar la búsqueda a familias de distribuciones apriori que posean ciertas características, tales como:

1. Tratabilidad analítica:

    a) Facilidad de determinación de la distribución posterior de la muestra y de la apriori.

    b) Facilidad para obtener características de interés, por ejemplo, valores esperados.

    c) La apriori y aposteriori deben ser miembros de la misma familia (cerrada).

2. Flexibilidad y riqueza: Debe permitir modelar una gran variedad de información apriori y creencias.

3. Interpretabilidad: Los parámetros deben ser de tal forma que el analista pueda relacionarlos fácilmente con sus creencias e información.

Las distribuciones conjugadas juegan un papel importante en los métodos bayesianos, ya que

- su uso puede simplificar el procedimiento de integración requerido para la marginalización.
- Ya que al pertenecer la apriori y la aposteriori a la misma familia, el proceso de actualización de parámetros se simplifica, lo cual es una gran ventaja para los sistemas inteligentes.

La conjugación nos limita a la selección de una clase de aprioris limitada y la información apriori solo puede utilizarse para la selección de los hiperparámetros. Si la clase es lo suficientemente grande esto puede no ser un gran problema.

Distribución Binomial

Teorema 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. También supongamos que la distribución apriori de π es una beta con parámetros $\alpha(> 0)$ y $\beta(> 0)$. 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 $\beta + n - \sum_{i=1}^{n} x_i$.

- 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 $\beta + n - \sum_i X_i$.

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

Distribución Binomial Negativa

Resultado

- Suponga que $X_1, \cdots, X_n$ es una muestra aleatoria de una distribución binomial negativa con parámetros r y π, donde r tiene una valor específico $(r > 0)$ y el valor de π es desconocido.
- También supongamos que la distribución apriori de π es una beta con parámetros $\alpha(> 0)$ y $\beta(> 0)$.
- Entonces la distribución posterior de π cuando $X_i = x_i$, para $i = 1, \cdots, n$ es una beta on parámetros $\alpha + rn$ y $\beta + \sum_{i=1}^{n} x_i$.

Mostrar

Distribución Geométrica

Otra distribución de conteo popular es la geométrica, la cual cuenta el número de fracasos antes de obtener el primer éxito. Su función de probabilidad está dada por

$$P(X = k) = (1 - \pi)\pi^k \qquad k = 0, 1, 2, \cdots$$

Su media es $\pi/(1 - \pi)$ y su varianza $\pi/(1 - \pi)^2$. El sesgo es $(1 + \pi)/\sqrt{\pi}$.