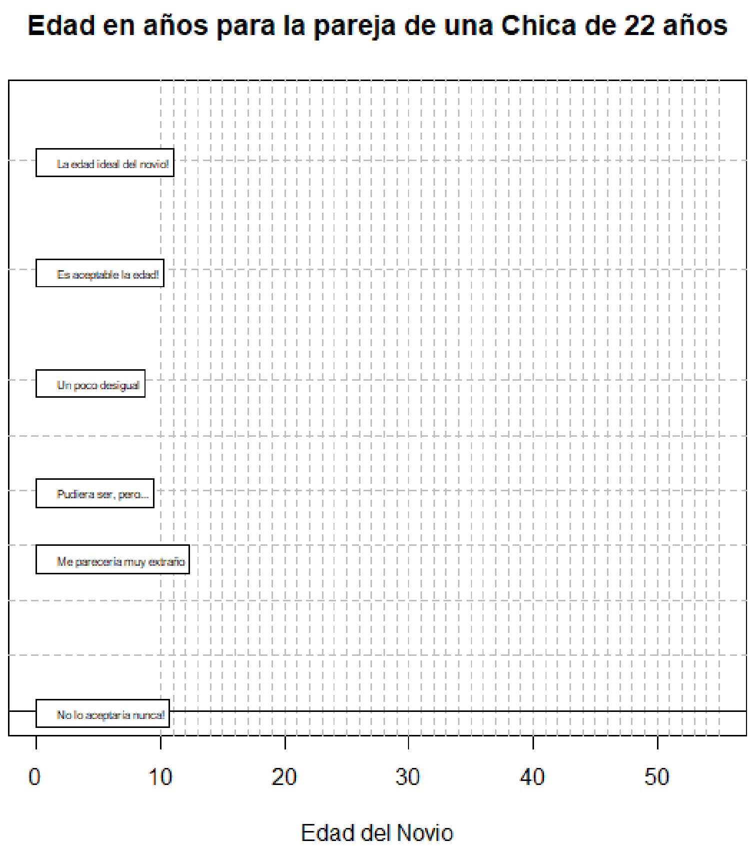
Estadística Bayesiana: Clase 5

Juan Carlos Correa

10 de marzo de 2021

Tarea 2: Determinación de la edad ideal de la pareja de una mujer de 22 años

```
par(yaxt="n")
plot(c(0,55),c(0,55),type="n",ylab="",xlab='Edad del Novio',
main="Edad en años para la pareja de una Chica de 22 años")
abline(v=c(10:55),lty=2,col="grey")
abline(h=0)
abline(h=c(5,10,15,20,25,30,40,50),lty=2,col="grey")
legend(0,1,"No lo aceptaría nunca!",cex=0.5,bg="white")
legend(0,15,"Me parecería muy extraño",cex=0.5,bg="white")
legend(0,21,"Pudiera ser, pero...",cex=0.5,bg="white")
legend(0,31,"Un poco desigual",cex=0.5,bg="white")
legend(0,41,"Es aceptable la edad!",cex=0.5,bg="white")
legend(0,51,"La edad ideal del novio!",cex=0.5,bg="white")
```


Edad en años para la pareja de una Chica de 22 años
La edad ideal del novio!
Es aceptable la edad!
Un poco desigual
Pudiera ser, pero...
Me parecería muy extraño
No lo aceptaría nunca!
0
10
20
30
40
50
Edad del Novio

Normal truncada

$$f\left(x\left|\mu,\sigma,a,b\right.\right)=\frac{1}{\sigma}\frac{\phi\left(\frac{x-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)}$$

para $a<x<b$ y $f=0$ en otro caso.

$$\phi\left(w\right)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}w^2\right)$$

y

$$\Phi\left(v\right)=\int_{-\infty}^{v}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}w^2\right)\,dw$$

La FD acumulada

$$F\left(x\right)=\frac{\Phi\left(\frac{x-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)}$$

$$Moda = \begin{cases} a & \text{si } \mu < a \\ \mu & \text{si } a < \mu < b \\ b & \text{si } \mu > b \end{cases}$$

Página www

Página de Abertura

Contenido

Regresar

Full Screen

Cerrar

Abandonar

$$E\left(X\right)=\mu+\frac{\phi\left(\frac{a-\mu}{\sigma}\right)-\phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)}\sigma$$

$$Var\left(X\right)=\sigma^2\left[1+A-B^2\right]$$

donde

$$A = \frac{\phi\left(\frac{a-\mu}{\sigma}\right)-\frac{b-\mu}{\sigma}\phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)}$$

$$B = \frac{\phi\left(\frac{a-\mu}{\sigma}\right)-\phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)}$$

Paquete truncnorm in R

```
dtruncnorm(x, a=-Inf, b=Inf, mean = 0, sd = 1)
ptruncnorm(q, a=-Inf, b=Inf, mean = 0, sd = 1)
qtruncnorm(p, a=-Inf, b=Inf, mean = 0, sd = 1)
rtruncnorm(n, a=-Inf, b=Inf, mean = 0, sd = 1)
etruncnorm(a=-Inf, b=Inf, mean=0, sd=1)
vtruncnorm(a=-Inf, b=Inf, mean=0, sd=1)
```

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

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

(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 $\xi(\theta)$.

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

Distribución Binomial con Una Apriori Normal Truncada

- Supongamos que la distribución apriori de π es una normal truncada con parámetros μ y σ.

$$\xi\left(\pi\,|\mu,\sigma\right) = \frac{1}{\sigma}\frac{\phi\left(\frac{\pi-\mu}{\sigma}\right)}{\Phi\left(\frac{1-\mu}{\sigma}\right)-\Phi\left(\frac{-\mu}{\sigma}\right)}$$

- 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

$$\xi\left(\theta\,|Datos\right) \propto \pi^{\sum_i X_i}(1-\pi)^{n-\sum_i X_i}\exp\left(-\frac{1}{2}\left(\frac{\pi-\mu}{\sigma}\right)^2\right)$$

Distribuciones No Informativas

- El uso de distribuciones apriori no informativas buscan que ellas tengan un impacto mínimo sobre la distribución posterior del parámetro de interés y que sea relativamente plana con relación a la verosimilitud.
- Esto busca que sean los datos los que tengan un claro dominio en la distribución posterior, y, por lo tanto, en todas las inferencias que de ellas se obtengan.
- También se conocen como vagas, difusas, planas o de referencia.

Estas distribuciones no informativas se reunen en dos grupos:

Propias: Cuando la distribución de probabilidad integra a una constante finita, se dice que es propia. Por ejemplo, para el caso de la distribución binomial, su parámetro π, que denota el porcentaje de éxitos en la población, podemos asumir como apriori la $U(0,1)$, lo cual refleja nuestra ignorancia total, al asumir que cualquier valor en este intervalo es igualmente posible como valor.

Impropias: Una distribución apriori $\xi(\theta)$ es impropia si

$$\int_{\Theta} \xi(\theta)\, d\theta = \infty$$

Winkler (1967a) dice

> “Los términos difuso y no-difuso son relativos en este contexto, no términos absolutos. Cuando decimos que nuestra información es difusa realmente queremos decir que es difusa relativa a la información muestral. También queremos decir que es localmente difusa (i.e., difusa solo dentro dentro de un cierto rango). Así, ‘difuso’ puede depender no solo de la precisión de la información muestral sino también de los valores específicos de la información muestral. En muchos casos el uso de distribuciones apriori difusas por parte del bayesiano puede ser sicológicamente iluminador, bien sea para otros o para él mismo, aún si su distribución apriori no es difusa. ”

Notas:

1. Una distribución apriori impropia puede terminar en una aposteriori impropia y por lo tanto no se podrán hacer inferencias.
2. Una distribución apriori impropia puede llevar a una aposteriori propia.

Ejemplo:
Asumamos que $y_1, \cdots, y_n|\theta$ son variables distribuidas normal e independientemente con media θ y con varianza conocida σ^2. Asumamos que $\xi(\theta) \propto 1$ es la distribución apriori uniforme (impropia) sobre los números reales. La verosimilitud es

$$L(\theta|y) \propto \exp\left(-\frac{n}{2}\frac{(\bar{y}-\theta)^2}{\sigma^2}\right)$$

y la distribución posterior es

$$\theta|y \sim N\left(\bar{y}, \frac{\sigma^2}{n}\right)$$

la cual es una distribución propia.

Yang y Berger (1998) presentan varias razones por las cuales es importante considerar las distribuciones no informativas. Tenemos entre ellas

- Con frecuencia la elicitación de las distribuciones apriori es imposible, por múltiples razones, por ejemplo, limitaciones de costo o tiempo, o resistencia o falta de entrenamiento de los clientes.
- El análisis estadístico debe aparecer como "objetivo".
- La elicitación subjetiva puede producir malas distribuciones subjetivas, por ejemplo si la elicitación es sesgada.
- En problemas de alta dimensión, lo más que se puede esperar es obtener buenas distribuciones subjetivas para algunos pocos parámetros, y a los parámetros de perturbación se les asignan distribuciones no informativas.
- El análisis bayesiano con distribuciones no informativas puede utilizarse para obtener procedimientos clásicos buenos.

Aún cuando un investigador tenga creencias apriori fuertes, puede ser más convincente analizar los datos utilizando una apriori de referencia dominada por la verosimilitud. Además podemos automatizar el proceso de hallar aprioris. Yang y Berger (1998) proporcionan un amplio catálogo de distribuciones no informativas que es útil en el trabajo aplicado.

1. El Principio de la Razón Insuficiente de Laplace

Si el espacio parametral es finito se puede utilizar una distribución apriori uniforme para reflejar ignorancia total.

2. Apriori de Jeffreys

La distribución apriori de Jeffreys satisface la propiedad local de uniformidad para distribuciones apriori no informativas. Esta apriori está basada en la matriz de información de Fisher. Jeffreys la propuso como una "regla general" para determinar la distribución apriori (Kass y Wasserman, 1994).

Definición: Sea $f\left(x|\theta\right)$ la densidad de x dado θ. La información de Fisher es definida como

$$\mathcal{I}(\theta) = -E\left[\frac{\partial^2 \log\left(f\left(x|\theta\right)\right)}{\partial\theta^2}\right]$$

Si θ es un vector de p componentes, entonces

$$\mathcal{I}(\theta) = -E\left[\frac{\partial^2 \log\left(f\left(x|\theta\right)\right)}{\partial\theta_i\,\partial\theta_j}\right]_{p\times p}$$

y entonces $\mathcal{I}(\theta)$ será una matriz de dimensión $p \times p$.

Definición: La distribución apriori de Jeffreys se define como

$$\xi(\theta) \propto |\mathcal{I}(\theta)|^{1/2}$$

La distribución apriori de Jeffreys es localmente uniforme y por lo tanto no informativa. Esta propiedad es importante ya que nos proporciona un esquema automatizado para hallar distribuciones apriori no informativas para cualquier modelo paramétrico (Ibrahim, 2002). Esta distribución es impropia para muchos modelos, sin embargo, es propia para algunos.

2.1. Ejemplo

Asumamos que $y_1, \cdots, y_n$ son variables distribuidas independientemente Bernoulli(π). Encontremos la distribución apriori de Jeffreys para π.
La densidad para una variable Bernoulli(π) es

$$p(y|\pi) = \pi^y (1-\pi)^{1-y}$$

Entonces tenemos

$$\begin{aligned}
\log\left(p(y|\pi)\right) &= y\log(\pi) + (1-y)\log(1-\pi) \\
\frac{\partial}{\partial\pi}\log\left(p(y|\pi)\right) &= \frac{y}{\pi} - \frac{1-y}{1-\pi} \\
\frac{\partial^2}{\partial\pi^2}\log\left(p(y|\pi)\right) &= -\frac{y}{\pi^2} - \frac{1-y}{(1-\pi)^2} \\
\mathcal{I}(\pi) &= -E\left[\frac{\partial^2}{\partial\pi^2}\log\left(p(y|\pi)\right)\right] \\
&= \frac{E(y)}{\pi^2} + \frac{1-E(y)}{(1-\pi)^2} = \frac{1}{\pi} + \frac{1-\pi}{(1-\pi)^2} \\
&= \frac{1}{\pi} + \frac{1}{1-\pi} = \frac{1}{\pi(1-\pi)}.
\end{aligned}$$

Por lo tanto la distribución apriori de Jeffreys es

$$\begin{aligned} \xi(\pi) &\propto \mathcal{I}(\pi)^{1/2} \\ &= \left(\frac{1}{\pi(1-\pi)}\right)^{1/2} \\ &= \pi^{-1/2}(1-\pi)^{-1/2} \\ &= \pi^{1/2-1}(1-\pi)^{1/2-1} \end{aligned}$$

Así $\pi \sim Beta\left(\frac{1}{2}, \frac{1}{2}\right)$. Por lo que vemos en este caso la distribución apriori de Jeffreys es propia.

2.2. Caso normal

Asumamos que $y_1, \cdots, y_n|\mu$ 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.