Práctica 1

Pablo Marcos Parra / Sofía Mara Rivas Cuevas

3/11/2020

Enunciado de la práctica

Considerar una muestra aleatoria de tamaño n = 30 de una población cuya distribución es una mixtura de dos distribuciones exponenciales con funciones de densidad:

$$f_1(x) = exp(-x)$$

$$f_2(x) = 2exp(-2x)$$

y proporciones p y 1-p respectivamente los datos observados son los siguientes:

```
y<-c(0.92169370,0.20110924,0.08299092,1.27148296,0.08975299,2.49922718
     ,4.34097682,0.39260263,0.06973844,0.05284850,0.40770048,0.03917915
     ,0.19068404,1.26898667,0.53213247,0.52049674,0.22417266,0.18774498
     ,0.16727780,0.44944121,1.10100809,0.84404590,0.66023800,2.86944266
     ,0.08869227,0.85046707,0.41026355,0.28243983,0.07341746,0.10278472)
```

Apartado A

Obtener intervalos de confianza, de Wald, para p, con confianza 0.95.

Para calcular el EMV de p vamos a utilizar la función *optim* de R. La función de la mixtura es:

$$f(y_i, p) = pf_1(y_i) + (1-p)f_2(y_i)$$

La función de verosimilitud es:

$$L(p, y_i) = \prod_{i=1}^{n} pf_1(y_i) + (1-p)f_2(y_i)$$

A la hora de pasarlo a R usaremos la menos log-verosimilitud (de aquí en adelante nos referiremos a ella como *mlv*) y utilizando la función mencionada anteriormente calcularemos el EMV:

```
f1<-function(y){
  return(dexp(y,1))
}
```

```
f2<-function(y){
  return(dexp(y,2))
}
mlv<-function(p){
  (-sum(log(p*f1(y)+(1-p)*f2(y))))
}
minimo<-optim(0.25,mlv,method="Brent",lower = 0, upper = 1,hessian = T)
EMV<-minimo$par
V<-as.numeric(solve(minimo$hessian))
I.C<-c(EMV-qnorm(0.975,0,1)*sqrt(V),EMV+qnorm(0.975,0,1)*sqrt(V))
```

EMV=

```
## [1] 0.3658762
```

I.C de Wald=

```
## [1] -0.08646918  0.81822162
```

La raíz de la varianza no está acotada y se nos puede ir a menos infinito, como se trata de una proporción el intervalo tiene que estar entre 0 y 1.

Apartado B

Obtener el p-valor del test de razón de verosimilitud para contrastar la hipótesis nula, *H0: p=0.5*.

La razón de verosimilitud es:

$$\Delta(X) = \frac{Sup_{\theta \in \Theta_0} L(\theta, X)}{Sup_{\theta \in \Theta} L(\theta, X)} = \frac{L(\widehat{\theta}_0, X)}{L(\widehat{\theta}, X)}$$

El estadístico razón de verosimilitud viene dado por:

$$\mathrm{T(X)} = -2\log\Delta(X) = 2\left[\log L\left(\hat{\theta}, X\right) - \log L(\hat{\theta}_0, X)\right]$$

Obtenemos el p-valor pedido con el siguiente código:

```
Tobs<--2*(minimo$value -mlv(0.5))
pvalor<-1-pchisq(Tobs,1)
```

P-Valor=

```
## [1] 0.583717
```

Apartado C

Utilizar el algoritmo EM para aproximar el estimador máximo verosímil de p, utilizando el valor inicial 0.5 y 25 iteraciones.

Obtendremos la densidad de los datos completos. Como tenemos solo los datos observados no sabemos la probabilidad con la que aparecerán *f1* y *f2* así que para sacar los datos completos usaremos *bi*:

$$f(y_i, b_i, p) = \begin{cases} pf_1(y_i) & \text{si } b_i = 1 \\ (1-p)f_2(y_i) & \text{si } b_i = 0 \end{cases}$$

Luego la función de verosimilitud de los datos completos es:

$$L(p,y,b) = \prod_{i=1}^{n} \left((pf_1(y_i))^{b_i}((1-p)f_2(y_i))^{1-b_i}\right)$$

Y su función log-verosimilitud:

$$logL(p,y,b) = (\sum_{i=1}^{n} b_i)log(p) + (n - \sum_{i=1}^{n} b_i)log(1-p) + \sum_{i=1}^{n} (b_i log(f_1(y_i)) + (1-b_i)log(f2(y_i)))$$

Paso E

$$E_{p^0}[logL(p,y,b) \mid Y_i = y_i] = log(p)(\sum_{i=1}^{n} E_{p^0}[b_i \mid Y_i = y_i] + log(1-p)(\sum_{i=1}^{n} E_{p^0}[b_i \mid Y_i = y_i] + \sum_{i=1}^{n} log(f_1(y_i))E_{p^0}[I(b_i \mid Y_i = y_i]$$

Y tras sacar la esperanza condicionada de la fórmula obtenemos:

$$b^* log(p) + (n - b^*)log(1-p) + \sum_{i=1}^{n} (log(f_1(y_i)b_i^*)) + \sum_{i=1}^{n} (log(f_2(y_i)(1-b_i^*))$$

Paso M

Para el paso M hay que derivar con respecto de p y para maximizar, igualar a 0:

$$\frac{b^*}{p} - \frac{n-b^*}{1-p} = 0$$

$$p^1 = \frac{b^*}{n}$$

$$p^{k+1} = \frac{1}{n}\sum_{i=1}^{n} \frac{p^0 f_1(y_i)}{p^0 f_1(y_i) + (1-p^0)f_2(y_i)}$$

En R queda:

```
DO<-function(y,p){
  return(p*f1(y)+(1-p)*f2(y))
}
p<-0.5 #valor inicial de p
n<-length(y)#tamaño datos
for(i in 1:25){
  p<-(1/n)*sum(p*f1(y)/DO(y,p))
}
```

```
## [1] 0.3687765
```