Practica 2

Pablo Marcos Parra / Ángel Moreno Calvo / Sofía Mara Rivas Cuevas

INTRODUCCIÓN

Como hemos visto en teoría, en el test de Kolmogorov-Smirnov necesitamos que a la hora de establecer la hipótesis de la distribución que se quiere realizar, ésta debe estar completamente especificada incluyendo el valor de sus parametros.

En algunos casos, como en esta práctica, nos interesa probar si los datos provienen de una distribución en concreto (normal, exponencial…) sin conocer el valor de los parametros.

Para ello aparecen los test de Lilliefors (test basados en el test de Kolmogorov-Smirnov), que permiten realizar un ajuste cuando los parámetros no son especificados si no estimados a partir de los datos.

En el caso de esta práctica trabajaremos con el test de Lilliefors para la distribución exponencial.

Apartado 1

Obtener, mediante simulación, los valores críticos de los tests de Lilliefors para contrastar exponencialidad, de tamaños 0.1, 0.05, 0.01, 0.001, correspondientes a una muestra de tamaño n=25. Es decir obtener la fila correspondiente al tamaño muestral n=25, de la tabla del test de Lilliefors para exponencialidad (Tabla T).

$x_1, x_2, ..., x_n$ v.a.iid con función de distribución desconocida F.

H_0: `F es Exponencial`

H_a: `F no es Exponencial`

La hipótesis nula puede transformarse en simple, sustituyendo λ por $\frac{1}{\bar{x}}$

$\widehat{D_n} = Sup|\widehat{F_n(x)} - F_0(x)|$, $F_0(x)$. es la funcion de la funcion de distribución, en x, $E(\lambda)$.

La distribución de $\widehat{D_n}$ no es la de D_n y utilizar la región crítica ($\widehat{D_n} > D_n$) produce un test muy conservador. Lilliefors obtuvo tablas adecuadas mediante simulación Monter Carlo. A traves de estas tablas comprobaremos los resultados.

Procedimiento en R

Para estos ejercicios con las tablas de Lillieford usaremos la librería de R `nortest`.

Para poder reproducir los experimentos y conseguir siempre los mismos resultados usaremos `set.seed`.

```r
library(nortest)
set.seed(420)
n<-25 #Tamaño muestral
tamsim<-10000 #Muestras para la simulacion
alfa<-c(0.1,0.05,0.01,0.001) #Valores criticos
M<-matrix(rexp(n*tamsim),n,tamsim) #Generamos la simulación
T<-rep(0,tamsim)
```

```
for(i in 1:tamsim){
  T[i]<-ks.test(M[,i],pexp,1/mean(M[,i]))$statistic
}
valorestabla<-quantile(T,1-alfa)
```

Entonces comparamos los valores obtenidos con los valores que aparecen en las tablas para el Test de Lilliefors para exponencialidad (n=25).

Valores críticos	0.1	0.05	0.01	0.001
Tabla Lilliefors	0.192	0.211	0.251	0.296
Valores simulación	0.1924575	0.2108745	0.2499075	0.2947183

Apartado 2

Considerar la muestra de tamaño n=25 dada en el vector y. Obtener, mediante simulación el p-valor del test de Lilliefors de exponencialidad

En este apartado disponemos de un vector y que es la muestra con 25 datos, entonces lo primero que debemos hacer es, a partir de estos datos, obtener una región de rechazo, D_n. Una vez hemos sacado ese valor realizamos la simulación con las 10000 muestras. Si el valor obtenido al realizar el test es mayor que D_n entonces lo sumamos a un contador, ya que significa que para esa muestra no podemos rechazar H_0.

El p-valor pedido viene dado por el número de veces que incrementamos el contador entre el número de simulaciones.

```r
y<-c(0.5088460,4.7784456,5.8640388,5.2947294,2.1367571,3.6835381,1.7578674,
1.4463360,5.4000721,8.5328018,4.2067582,5.4932311,1.1580206,0.2325523,0.6011006,
8.6225102,0.6864657,2.9142432,7.6363457,0.3445848,1.7555547,10.1100430,0.9012126,
3.2912391,0.5992558)

#Recordamos
n<-25
tamsim<-10000

dn<-ks.test(y,pexp,1/mean(y))$stat #Obtenemos el valor de dn

#Simulamos la fila de la tabla correspondiente a n=25
M<-matrix(rexp(n*tamsim),n,tamsim)
#Calculamos el ks.test para cada col
cont<-0
T<-rep(0,tamsim)
for(i in 1:tamsim){
  T[i]<-ks.test(M[,i],pexp,1/mean(M[,i]))$statistic
  if(T[i]>dn){
    cont<-cont+1
  }
}
pvalor<-cont/tamsim
```

Con el p-valor obtenido ≈ 0.8432, no podemos rechazar que y provenga de una Exponencial.

Apartado 3

Calcular, mediante simulación, la potencia, en la alternativa Gamma(2,0.7), del test de Lilliefors, de nivel 0.01, para contrastar la hipótesis nula de exponencialidad, basado en una muestra de tamaño n=50.

Primero hay que recordar que una distribución Gamma se relaciona con la distribución Exponencial de manera que sean `n` variables aleatorias independientes e igualmente distribuidas que siguen una distribución Exponencial con parámetro λ entonces $\sum_{i=1}^{n} X_i \sim \text{Gamma}(\text{n}, \frac{1}{\lambda})$

En este apartado se nos pide obtener la potencia para el siguiente contraste de hipótesis:

`H0: exponencial`

`H1: gamma(2,0.7)`

```
n<-50

#Muestras de la hipotesis nula y de la alternativa.
mH0 <- matrix(rexp(n*tamsim),n,tamsim)
mH1 <- matrix(rgamma(n*tamsim,2,0.7),n,tamsim)
```

Existen dos opciones para obtener la potencia, ya que existen dos maneras diferentes de obtener un valor para d_n.

Primera opción:

Generar d_n a partir de la simulación, tal y como hicimos en el apartado a) pero ahora para n=50 y valor crítico 0.01.

```
TH0<-rep(0,tamsim)
for (i in 1:tamsim){
  TH0[i]<-ks.test(mH0[,i],pexp,1/mean(mH0[,i]))$statistic
}
Dn<-quantile(TH0,0.99)

cont<-0
TH1<-rep(0,tamsim)
for(i in 1:tamsim){
  TH1[i]<-ks.test(mH1[,i],pexp,1/mean(mH1[,i]))$statistic
  if (TH1[i]>=Dn){
    cont<-cont+1
  }
}
potencia<-cont/tamsim
```

La potencia obtenida en esta primera opción es: **0.5565**.

Segunda opción:

Obtener d_n de las tablas de Lilliefors (valor 0.01 n=50)

```
Te<-0
cont1<-0
for(i in 1:tamsim){
  Te<-ks.test(mH1[,i],pexp,1/mean(mH1[,i]))$statistic
  if(Te>=0.179){
   cont1<-cont1+1
```

```
  }
}
potencia1 <- cont1/tamsim
```

La potencia obtenida en la segunda opción es: **0.5647**.