

Guia 12

Agustin Muñoz Gonzalez

20/5/2020

Preparamos el entorno

```
rm(list=ls())
```

Las aves parásitas de cría ponen huevos en nidos de otras especies (hospedador), las cuales incuban los huevos y crían al pichón parásito. En un bosque de talas de la provincia de Buenos Aires hay dos especies hospederas que son indistinguibles a simple vista. Una de las principales diferencias entre estas especies radica en el grado de discriminación y remoción de huevos parásitos de sus nidos. Una de las especies es “aceptadora” de huevos parásitos ($Y = 0$), ya que remueve del nido sólo el 30% de los huevos parásitos, mientras que la otra especie es “rechazadora” ($Y = 1$) ya que remueve el 80% de los huevos parásitos presentes en su nido. Además, se sabe que el 90% de los nidos del bosque corresponden a la especie “aceptadora”, mientras que apenas el 10% restante son nidos de la especie “rechazadora”. De esta manera la distribución conjunta del vector (X, Y) queda determinada por la siguiente tabla

a) Consideremos la regla de clasificación g_1 dada por :

```
tabla=data.frame(rbind(
  'Y=0'=c(0.0519,0.1779,0.2668,0.2287,0.1225,0.0420,0.0090,0.0011,0.0001),
  'Y=1'=c(0,0,0.0001,0.0009,0.0046,0.0147,0.0294,0.0336,0.0168)))
names(tabla)=(0:8)
tabla
```

```
##           0           1           2           3           4           5           6           7           8
## Y=0 0.0519 0.1779 0.2668 0.2287 0.1225 0.0420 0.0090 0.0011 0.0001
## Y=1 0.0000 0.0000 0.0001 0.0009 0.0046 0.0147 0.0294 0.0336 0.0168
```

Calcular

$$L(g_1) = \mathbb{P}(g_1(X) \neq Y)$$

y completar con el valor obtenido en el archivo compartido.

Resolución:

Vamos a usar el teo de proba total

$$\begin{aligned} P(g(X) \neq Y) &= 1 - P(g(X) = Y) \\ &= 1 - \sum_x P(g(X) = Y, X = x) \\ &= 1 - \sum_x P_{XY}(x, g(x)) \end{aligned}$$

Vamos a definir la proba puntual conjunta como una función.

```

g=function(x){
  if(x<=3){0}else{1}
}
p_XY=function(x,y){
  tabla[y+1,x+1]
  # otra forma: tabla[y+1,as.character(x)]
  # porque el nombre de c/col es el valor de x
}
error_medio=1
for(x in (0:8)){
  error_medio=error_medio-p_XY(x,g(x))
}
error_medio

## [1] 0.1756

```

- b) El archivo `depredadosclasificadas.txt` contiene mediciones correspondientes a nidos elegidos al azar en el bosque, registrándose en cada uno de ellos el número de huevos parásitos removidos (primera columna) y la especie del hospedador (segunda columna), determinada mediante un análisis genético. Tenemos entonces observaciones (X, Y) donde X indica la cantidad de huevos removidos del nido, mientras que $Y = 0$ si el hospedador es aceptador, $Y = 1$ indica que el hospedador es rechazador. Computar el Error de Clasificación Empírico de la regla de clasificación g_1 a partir de los datos del archivo `depredadosclasificadas.txt` y completar con el valor obtenido en el archivo compartido.

```
depredados=read.csv('depredadosclasificadas.csv',sep=' ',header = TRUE)
# le tengo que poner sep=' ' para que entienda que tiene que las columnas
# están divididas por un espacio.
```

Resolución:

Vimos en las diapos que el Error de Clasificación Empírico es la proporción de pares mal clasificados según g_1 . Es decir, el promedio de los casos en los que $g_1(X) \neq Y$.

```
X=depredados$removidos
Y=depredados$especie
error_empirico=mean(lapply(X,g)!=Y)
```

- c) Comparar Error de Clasificación Empírico con el valor obtenido en a).

La diferencia de errores es de

```
abs(error_medio-error_empirico)
```

```
## [1] 0.0164
```

Es decir hay un error de solo el 1.6%.