**Seminario 3**

# Estimadores de máxima Verosimilitud

**La verosimilitud de una muestra.**

A la densidad conjunta evaluada en la muestra, se la llama **función de verosimilitud (versemblança / likelyhood) de la muestra** .

Es una V.A. con valores en $\mathbf{R}^+$.

Denotaremos por

$$L(X_1, \ldots, X_n; \theta) = \prod_{i=1}^{n} f(X_i; \theta)$$

donde $f(x;\theta)$ es la densidad de las $X_i$ (o la función de probabilidades en el caso discreto), a la función de verosimilitud, y por

$$\lambda(X_1, \ldots, X_n; \theta) = \sum_{i=1}^{n} \log f(X_i; \theta)$$

al logaritmo de $L$.

El método de máxima verosimilitud consiste en utilizar como estimador de $\theta$ el valor $\hat{\theta}$ que maximiza la verosimilitud de la muestra.

El **estimador de máxima verosimilitud**[1] $\hat{\theta}$ es aquél que verifica

$$L(X, \hat{\theta}) \geq L(X, \theta)$$

para todo $\theta$ en el espacio de parámetros.

La función de densidad $f(x;\theta)$ (o en su caso la probabilidad $p(x;\theta) = \mathbf{P}\{X = x\}$) nos da una medida intuitiva de cuánto podemos confiar en que ocurra $x$ una vez que conocemos $\theta$, pero también nos permite comparar esas medidas para distintos valores del parámetro $\theta$. Particularmente, una vez que hemos observado $x$, $f(x;\theta)$ nos indica hasta qué punto era esperable el resultado observado, cuando el parámetro vale $\theta$.

Encontraremos entonces que el modelo probabilístico dado por $f(x;\theta')$ resulta más creíble (o más verosímil) que el modelo definido por $f(x;\theta'')$ cuando ocurra $f(x;\theta') > f(x;\theta'')$.

En este contexto, podemos decir que el método de máxima verosimilitud propone utilizar como estimador del parámetro, el valor que vuelve más creíble o más *esperable* el resultado que de hecho ha sido observado experimentalmente.

## 1 Cálculo de los estimadores de máxima verosimilitud con R

La librería `MASS` tiene una función (`fitdistr`) que calcula el MLE para algunas distribuciones:

[1] puede no ser único, puede que no exista

```
library(MASS)
?fitdistr    #  help sobre la función

############################  exponenciales
x=rexp(100,rate=2)#rate es lambda
fitdistr(x,"exponential")
1/mean(x) ## comprobamos que el estimador es el que conocemos
########################### normales
n=100
y=rnorm(100,3,6) #media=3, desviación=6
fitdistr(y,"normal")
## comprobamos que los estimadores son los que conocemos:
mean(y); var(y)*(n-1)/n; sqrt(var(y)*(n-1)/n)
```

- Hay otra librería (`fitdistrplus`) que incluye más modelos.
- El cálculo del máximo de la verosimilitud también se puede hacer con `nlm` o con `optim` (rutinas de optimización de `R`). Estas rutinas minimizan funciones, por lo tanto hay que escribir una función que contenga $-\log L(X_1, \ldots, X_n; \theta)$, donde $\theta \in \mathrm{R}^p$.

### 1.1 Ejemplo - el MLE para el parámetro $\lambda$ de una distribución exponencial:

Generar una muestra de tamaño 100 y $\lambda = 2$,

```
x=rexp(100,rate=2)
mean(x);1/mean(x)
##definimos una funcion con -log(verosimilitud)
mlogL=function(x,lam=1){return(-(length(x)*log(lam)-lam*sum(x)))}
```

**En realidad no hace falta escribir en este caso la verosimilitud, pues R la puede calcular:**

```
mlogl=function(x,lam=1){return(-sum(dexp(x,rate=lam, log = TRUE))) }
##### d son las densidades, log=TRUE su logaritmo

##### Gráfico de log(verosimilitud)
npoint <- 101
lams <- seq(min(x), max(x)+2, length = npoint)
logls <- numeric(npoint)
for (i in 1:npoint)
   logls[i] <- mlogl(x,lams[i])
plot(lams, -logls, type = "l",
    xlab = expression(lams), ylab = expression(l(lams)))

#### Cálculo del máximo usando nml y optim

nlm(mlogl, 1, x = x)
optim(1,mlogl,x=x)
1/mean(x)
```

## Ejercicios

1. Escribir una función que calcule la -log-verosimilitud para una muestra de variables aleatorias con distribución normal de media y varianza desconocidas. Usarla para estimar los parámetros de una muestra de tamaño 100 generada con $\mu = 3, \sigma = 2$ con `nlm`.

2. Calcular los estimadores de máxima verosimilitud de $\mu$ y $\sigma^2$ para los datos de Michelson de la librería `MASS`. ¿Es razonable pensar que son normales? Repetir el cálculo eliminando el posible outlier.

3. La següent taula mostra distribució del nombre de pars de bambes per a un grup de 60 corredors aficionats.

| | | | | | |
|---|---|---|---|---|---|
| Nombre de pars: | 1 | 2 | 3 | 4 | 5 |
| Freqüencia: | 18 | 18 | 12 | 7 | 5 |

   La distribució de Poisson no pot ser un bon model per a aquestes dades, perquè se suposa que els corredors tenen almenys un par de bambes. La distribució de Poisson truncada (en zero) amb funció de massa de probabilitats

$$p(x;\theta) = \frac{\theta^x e^{-\theta}}{x!(1-e^{-\theta})}, \quad x = 1, 2, \ldots, \quad \theta > 0$$

   pot ser un model més adient.

   (a) Escriure la versemblança, el Score i la informació observada.

   (b) Trobar (numèricament) l'estimador de màxima versemblança de $\theta$.

4. En la librería `evir` hay un conjunto de datos llamado `danish` que corresponde a los montos de reclamaciones por incendios hechas a aseguradoras en Dinamarca entre el 3 de enero de 1980 y el 31 de diciembre de 1990. Supongamos que provienen de una distribución de Pareto, con densidad

$$f(x) = \frac{\alpha x_m^\alpha}{x^{\alpha+1}} \quad \text{si } x \geq x_m$$

   y 0 si $x < x_m$.

   (a) Calcular el EMV para los parámetros $\alpha$ y $x_m$ de la distribución de Pareto. Observar que $x_m$ está en la frontera del espacio de parámetros.

5. Las densidades Gamma pertenecen a una familia de distribuciones continuas con dos parámetros. La exponencial y la $\chi^2$ son casos particulares de densidades Gamma. Hay tres parametrizaciones de uso común:

   - Con un parámetro de forma (shape) $k$ y uno de escala (scale)$\theta$.
   - Con un parámetro de forma (shape) $\alpha = k$ y un parámetro (rate) que es la inversa del parámetro de escala, $\beta = 1/\theta$
   - Con un parámetro de forma (shape) $k$ y la media $\mu = k/\beta$.

   En cada una de estas tres formas, los parámetros son números reales positivos.

   Si usamos la segunda de las parametrizaciones, una V.A. $X \sim \Gamma(\alpha, \beta)$ la densidad de $X$ es

$$f(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \mathrm{e}^{-\beta x}$$

   donde $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \mathrm{e}^{-x} dx$

Los estimadores de máxima verosimilitud no admiten una forma cerrada, así que hay que calcularlos numéricamente.

(a) Generar una muestra de tamaño 100 de una $\Gamma(10, 3)$ `x=rgamma(100,shape=10,rate=3)`

(b) Encontrar los estimadores que da el método de los momentos, para usarlos como valores iniciales en el cálculo de los MLE.

(c) Calcular los estimadores de máxima verosimilitud.