Soluciones a la hoja de ejercicios 7

Ejercicio 1

El estimador por el método de los momentos lo obtenemos de la ecuación

$$\hat{\mu} = \frac{\theta}{\theta + 1}$$

Despejando

$$\hat{\theta}_{momentos} = \frac{\hat{\mu}}{1 - \hat{\mu}}$$

Para obtener el estimador máximo-verosímil necesitamos la función de verosimilitud

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} \theta x_i^{\theta - 1} \\ &= \theta^n \prod_{i=1}^{n} x_i^{\theta - 1} \end{aligned}$$

La función de log-verosimilitud y sus dos primeras derivadas con respecto a θ son

$$\begin{aligned} l(\theta) &= n \log \theta + (\theta - 1) \sum_{i=1}^{n} \log x_i \\ l'(\theta) &= \frac{n}{\theta} + \sum_{i=1}^{n} \log x_i \\ l''(\theta) &= -\frac{n}{\theta^2} \end{aligned}$$

A partir de la ecuación de estimación, $l'(\theta) = 0$, obtenemos el estimador

$$\hat{\theta} = -\frac{n}{\sum_{i=1}^{n} \log x_i}$$

El estimador de θ es necesariamente positivo porque X está comprendido entre 0 y 1, por lo que $\log X < 0$ y el denominador $\sum_{i=1}^{n} \log x_i$ es menor de 0. En consecuencia $l''(\theta) < 0$ y el punto encontrado es un máximo de $l(\theta)$.

Utilizando la información de Fisher, la varianza del estimador es

$$Var(\theta) = \frac{\theta^2}{n}$$

Con los datos del ejemplo, la media muestral es $\hat{\mu} = 0,624$ y $\sum_{i=1}^{n} \log x_i = -5,40$, por lo que

$$
\begin{aligned}
\hat{\theta}_{momentos} &\approx 1,66 \\
\hat{\theta} &\approx 1,85 \\
Var(\theta) &\approx 0,34 \\
Se &\approx 0,586
\end{aligned}
$$

Utilizando la aproximación normal a la distribución del estimador, los límites del intervalo de confianza al 95 % para θ son

$$
\begin{aligned}
L_s &= \hat{\theta} + |Z_{\alpha/2}|Se = 1,85 + 1,96 \times 0,586 = 3,000 \\
L_i &= \hat{\theta} - |Z_{\alpha/2}|Se = 1,85 - 1,96 \times 0,586 = 0,704
\end{aligned}
$$

Por tanto, mantenemos $H_0 : \theta = 2$ y no podemos descartar que este sea el verdadero valor de θ.

Ejercicio 2

El código R es el siguiente. Los resultados que produce coinciden con los obtenidos mediante cálculo matemático.

```
x <- c(0.48, 0.38, 0.83, 0.51, 0.35, 0.80, 0.86, 0.35, 0.73, 0.95)
n <- length(x)

thetaMomentos <- mean(x)/(1-mean(x))

l <- function(p){
  logLk <- 0
  for(i in 1:n){
    logLk <- logLk + log(p[1]) + (p[1]-1)*log(x[i])
  }
  return(logLk)
}

fit <- optim(1, f=l, method="Brent", lower=0, upper=10, control=list(fnscale=-1), hessian=TRUE)

thetaML <- fit$par
info <- -fit$hessian
Se <- sqrt(1/info)

cat(sprintf("El estimador por el método de los momentos es theta = %5.3f
El estimador máximo-verosímil es theta = %5.3f, Se = %5.3f\n", thetaMomentos, thetaML, Se))
```

```
## El estimador por el método de los momentos es theta = 1.660
## El estimador máximo-verosímil es theta = 1.852, Se = 0.586
```

Ejercicio 3

Programos la función de log-verosimilitud como la suma de la contribución de cada una de las observaciones de la muestra, matemáticamente la función a programar es

$$
l(\alpha, \beta) = \sum \log f(\alpha, \beta)
$$

donde $f(\alpha, \beta)$ es una función de densidad beta. El código R es

```
h <- function(p){
  logLk <- 0
  for(i in 1:n){
    logLk <- logLk + log(dbeta(x[i], p[1], p[2]))
  }
  return(logLk)
}

fitBeta <- optim(c(1,1), f=h, method="L-BFGS-B", lower=0, control=list(fnscale=-1),hessian=TRUE)
```

Al ejecutar el programa obtenemos los siguientes estimadores para α y β

```
print(fitBeta$par)
```

```
## [1] 2.572801 1.506321
```

Los errores típicos son las raices cuadradas de la matriz de varianzas-covarianzas de los estimadores:

```
Informacion <- - fitBeta$hessian
VarCovar <- solve(Informacion)
Se <- sqrt(diag(VarCovar))
print(Se)
```

```
## [1] 1.133109 0.621452
```

La función *cov2cor* de R convierte una matriz de varianzas-covarianzas a matriz de correlaciones. El resultado es

```
R <- cov2cor(VarCovar)
print(R)
```

```
##           [,1]      [,2]
## [1,] 1.0000000 0.7764304
## [2,] 0.7764304 1.0000000
```

Los resultados indican que tanto los errores típicos como la correlación entre estimadores son elevados. Esto significa que puede haberse producido *sobreajuste*. Es decir, haber estimado un modelo demasiado complejo para lo que requieren estos datos.