

Capítulo 7. Propiedades de los estimadores

Índice

1. Error cuadrático medio	1
2. Consistencia de los estimadores	3
3. Propiedades asintóticas de los estimadores máximo-verosímiles	3
3.1. Media, varianza y distribución de los estimadores	4
3.2. Información observada	4
4. Intervalos de confianza	6
5. Cálculo de la varianza del estimador en lenguaje R	7
6. Máxima verosimilitud con dos o más parámetros en lenguaje R	9
7. Estimación de un modelo de regresión en R	10
8. Ejercicios	11

1. Error cuadrático medio

Desde el punto de vista de la estadística clásica se asume que los parámetros, θ , son cantidades fijas aunque desconocidas. En un experimento aleatorio se calcula el estimador $\hat{\theta}$ a partir de los datos disponibles en la muestra, por lo que el estimador es una variable aleatoria y como tal está caracterizada por su valor esperado y su varianza. Estas cantidades, el valor esperado del estimador y su varianza, se utilizan para evaluar las propiedades del estimador y si es “razonable” esperar que el valor de $\hat{\theta}$ esté próximo a θ .

El estadístico principal para evaluar el funcionamiento de un estimador es el error cuadrático medio (mean squared error, MSE). Supongamos que θ es el valor verdadero de un parámetro y $\hat{\theta}_j$ es su valor estimado en la muestra j . Si el número de muestras diferentes que pueden tomarse de la población es J , el MSE es la media de las diferencias al cuadrado entre el estimador y el parámetro

$$MSE = \frac{1}{J} \sum_{j=1}^J (\hat{\theta}_j - \theta)^2$$

El MSE no es la varianza del estimador porque la media de $\hat{\theta}_j$ no tiene por qué ser igual a θ . En concreto, supongamos que el valor esperado del estimador es

$$\mu_{\hat{\theta}} = \frac{1}{J} \sum_{j=1}^J \hat{\theta}_j$$

Entonces podemos desarrollar el MSE del siguiente modo

$$\begin{aligned}
MSE &= \frac{1}{J} \sum_{i=1}^J (\hat{\theta}_j - \mu_{\hat{\theta}} + \mu_{\hat{\theta}} - \theta)^2 \\
&= \frac{1}{J} \sum_{i=1}^J [(\hat{\theta}_j - \mu_{\hat{\theta}})^2 + (\mu_{\hat{\theta}} - \theta)^2 - 2(\hat{\theta}_j - \mu_{\hat{\theta}})(\mu_{\hat{\theta}} - \theta)] \\
&= S_{\hat{\theta}}^2 + b_{\hat{\theta}}^2
\end{aligned}$$

Esta ecuación indica que MSE es igual a la varianza del estimador más su sesgo al cuadrado. En primer lugar, debe advertirse que $\sum_{i=1}^J 2(\hat{\theta}_j - \mu_{\hat{\theta}})(\mu_{\hat{\theta}} - \theta) = 0$, por lo que este término desaparece del desarrollo. La varianza del estimador

$$S_{\hat{\theta}}^2 = \frac{1}{J} \sum_{i=1}^J (\hat{\theta}_j - \mu_{\hat{\theta}})^2$$

indica cuanto difiere el valor del estimador entre las muestras, y no depende de θ . El error típico del estimador es la raíz cuadrada de la varianza $S_e = \sqrt{S_{\hat{\theta}}^2}$. El sesgo al cuadrado

$$b_{\hat{\theta}}^2 = (\mu_{\hat{\theta}} - \theta)^2$$

indica cuanto se aleja la media de los estimadores del valor verdadero del parámetro.

Ejemplo 1. Vamos a evaluar las propiedades del estimador mínimo-cuadrático de la pendiente de la regresión. La regresión lineal de Y sobre X en un modelo en el que el intercepto es cero es

$$y'_i = \beta x_i$$

El estimador mínimo-cuadrático de β es

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$$

Realizaremos un estudio de con las siguientes características, el tamaño muestral es 10, X toma los valores del 1 al 10, el valor verdadero del parámetro es $\beta = 2$, la varianza error de la regresión es $\sigma_e^2 = 36$, tomaremos 1000 muestras simuladas. El código R es

```

n <- 10
x <- 1:n
J <- 1000
beta <- 2
sigma2_e <- 36
beta_estimada <- rep(0, J)
for(j in 1:J){
  y <- rnorm(n, beta*x, sqrt(sigma2_e))
  beta_estimada[j] <- sum(x*y) / sum(x*x)
}
S2 <- mean((beta_estimada - mean(beta_estimada))^2)
b2 <- (mean(beta_estimada) - beta)^2
MSE <- S2+b2
Se <- sqrt(S2)
cat(sprintf("S2 = %10.8f, Se = %10.8f, sesgo^2 = %10.8f, MSE = %10.8f\n", S2, Se, b2, MSE))

```

```
## S2 = 0.10302345, Se = 0.32097266, sesgo^2 = 0.00001019, MSE = 0.10303364
```

En un estudio de simulación no se toman todas las posibles muestras que podrían extraerse de la población bajo estudio sino sólo unas pocas. De hecho, en el caso de la distribución normal, como en este ejemplo, el número de muestras distintas es infinito. Por esto, la simulación no proporciona el valor real del MSE y las otras propiedades del estimador sino una estimación.

2. Consistencia de los estimadores

La consistencia es una propiedad deseable de los estimadores, que dicho de manera informal significa que a medida que el tamaño muestral aumenta el estimador $\hat{\theta}$ converge en probabilidad al valor verdadero θ . En otras palabras, un procedimiento estadístico es consistente si proporciona la respuesta correcta cuando dispone de infinita información.

Dos condiciones suficientes para que un estimador sea consistente es que sea insesgado y que su varianza tienda a cero. No son condiciones necesarias porque un estimador puede ser consistente aunque no las cumpla.

Ejemplo 2. La media muestral es un estimador consistente de la media poblacional. El valor esperado de la media muestral es

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{1}{n} \sum_{i=1}^n x_i\right) \\ &= \frac{1}{n} \sum_{i=1}^n E(x_i) \\ &= \frac{1}{n} \sum_{i=1}^n \mu \\ &= \mu \end{aligned}$$

Por tanto es un estimador insesgado, $b_{\hat{\mu}} = E(\bar{X}) - \mu = 0$. Además, la varianza de la media es

$$\begin{aligned} Var(\bar{X}) &= E\left(\frac{1}{n} \sum_{i=1}^n x_i\right)^2 \\ &= \frac{1}{n^2} \sum_{i=1}^n Var(x_i) \\ &= \frac{1}{n^2} \sum_{i=1}^n \sigma^2 \\ &= \frac{\sigma^2}{n} \end{aligned}$$

Por tanto, asintóticamente la varianza es cero, $\lim_{n \rightarrow \infty} \sigma^2/n = 0$.

3. Propiedades asintóticas de los estimadores máximo-verosímiles

Los estimadores máximo-verosímiles cumplen determinadas propiedades asintóticas que los hacen ser muy convenientes. Propiedades asintóticas son aquellas que se cumplen en el límite $n \rightarrow \infty$, por lo que no están garantizadas en muestras de tamaño finito, aunque cabe esperar que se cumplan en muestras de elevado tamaño. Las propiedades óptimas de los estimadores máximo-verosímiles se consiguen a costa de un precio: los supuestos tan fuertes que asume el método.

Con otros métodos de estimación, como mínimos cuadrados o la estimación por momentos, no existe una teoría que informe de las propiedades de los estimadores, aunque pueden analizarse las propiedades de algunos casos específicos.

3.1. Media, varianza y distribución de los estimadores

Sea θ un parámetro y $\hat{\theta}$ su estimador máximo-verosímil. Cuando $n \rightarrow \infty$ se cumple que

1. El estimador $\hat{\theta}$ es insesgado, es decir $E(\hat{\theta}) \rightarrow \theta$.
2. La varianza del estimador cumple que $Var(\hat{\theta}) \rightarrow 1/I(\theta)$, donde $I(\theta)$ es la información observada en la muestra acerca del valor del parámetro.
3. La distribución de $\hat{\theta}$ es normal.

En definitiva, en muestras grandes podemos asumir que $\hat{\theta}$ es normal $(\theta, 1/\sqrt{I(\theta)})$. Gracias a estas propiedades podemos calcular el error típico de los estimadores y el intervalo de confianza para el parámetro.

3.2. Información observada

La información observada se calcula como -1 multiplicado por la segunda derivada del logaritmo de la función de verosimilitud, es decir:

$$I(\theta) = -\frac{\partial^2}{\partial \theta^2} l(\theta) .$$

En consecuencia, podemos calcular de forma aproximada la varianza del estimador en todas las posibles muestras de tamaño n que pueden obtenerse de la población mediante la fórmula

$$Var(\theta) = \frac{1}{I(\hat{\theta})} .$$

La precisión del estimador también se evalúa mediante su error típico, que es la raíz cuadrada de la varianza

$$\sigma_{\hat{\theta}} = \frac{1}{\sqrt{I(\hat{\theta})}} .$$

La demostración de por qué la varianza del estimador puede obtenerse a partir de $I(\theta)$ excede los objetivos de este curso. Sin embargo, debemos conocer que $I(\theta)$ básicamente está indicando la curvatura de $l(\theta)$ evaluada en el punto $\hat{\theta}$. Si la curvatura es alta en este punto el estimador será más preciso. Cuando $l(\theta)$ es muy plana en torno al estimador $\hat{\theta}$ la estimación será más imprecisa porque $l(\theta)$ asigna un nivel de verosimilitud similar a puntos muy separados unos de otros. Los siguientes ejemplos ilustran estos conceptos.

Ejemplo 3. Supongamos que una variable sigue la distribución normal(μ, σ). Además, el valor de σ es conocido y el objetivo es estimar μ . La función de log-verosimilitud y su derivada con respecto a μ son

$$l(\mu) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{nS_Y^2 + n(\bar{Y} - \mu)^2}{2\sigma^2}$$

$$l'(\mu) = \frac{n(\bar{Y} - \mu)}{\sigma^2}$$

Por tanto, la ecuación de estimación $l'(\mu) = 0$ tiene como solución

$$\hat{\mu} = \bar{Y}$$

La segunda derivada de $l(\mu)$ con respecto a μ es:

$$l''(\mu) = -\frac{n}{\sigma^2}.$$

En consecuencia, la varianza asintótica y el error típico del estimador son

$$\begin{aligned} \text{Var}(\hat{\mu}) &= \frac{\sigma^2}{n} \\ \sigma_{\hat{\mu}} &= \frac{\sigma}{\sqrt{n}} \end{aligned}$$

En consecuencia, el estimador de μ es consistente porque es insesgado y su varianza tiende a cero, $\lim_{n \rightarrow \infty} \sigma^2/n = 0$.

Ejemplo 2. Supongamos que tenemos los datos $\mathbf{y} = (104, 96)$ proceden de una distribución normal($\mu, \sigma = 2$). El estimador máximo verosímil de μ y su varianza son

$$\begin{aligned} \hat{\mu} &= \bar{Y} = 100 \\ \text{Var}(\hat{\mu}) &= \frac{\sigma^2}{n} = \frac{4}{2} = 2 \end{aligned}$$

El error típico del estimador es $\sigma_{\hat{\mu}} = \sqrt{2} \approx 1,41$, que es una aproximación a la desviación típica de $\hat{\mu}$ en todas las posibles muestras de tamaño $n = 2$ que podrían extraerse de esta población.

Supongamos ahora que la muestra es $\mathbf{y} = (104, 105, 96, 99, 97, 95, 101, 103)$. Entonces el estimador y su varianza son

$$\begin{aligned} \hat{\mu} &= \bar{Y} = 100 \\ \text{Var}(\hat{\mu}) &= \frac{\sigma^2}{n} = \frac{4}{8} = 0,5 \end{aligned}$$

El error típico ahora es más pequeño, $\sigma_{\hat{\mu}} = \sqrt{0,5} \approx 0,71$, como consecuencia del mayor tamaño muestral.

La figura 1 muestra la función $l(\mu)$ para ambas muestras, después de eliminar de dicha función todas las constantes innecesarias que no afectan al estimador. Es obvio que $l(\mu)$ toma valores más pequeños en el caso de $n = 8$, lo cual se debe a que $l(\mu)$ es la suma de la contribución de cada una de las observaciones de la muestra. Como $l(\mu) = \sum_i \log f(x_i)$, cada observación contribuye a $l(\mu)$ con un término $\log f(x_i) < 0$, y la función $l(\mu)$ va disminuyendo al aumentar n . Además puede verse que $l(\mu)$ alcanza su máximo en el mismo punto para ambas muestras debido a que la media de ambas es 100.

La curvatura de $l(\mu)$ es mayor con $n = 8$ que con $n = 2$. En ambos casos el estimador está en el mismo punto, pero cuando la curvatura es mayor está más claro cual debe ser el estimador. Con $n = 2$ la verosimilitud de otros valores de μ distintos a $\hat{\mu} = 2$ es relativamente alta, lo que quiere decir que hay una mayor incertidumbre acerca del valor del estimador. Este fenómeno es el que está recogiendo la información, $I(\mu)$.

```
mu <- seq(90, 110, by=0.001)
l1 <- -2*((100 - mu)^2)/8
l2 <- -8*((100 - mu)^2)/8
par(mfrow=c(1,2))
plot(mu, l1, col="darkolivegreen4", lwd=1.5, type="l", ylab="", xlab="", ylim=c(-100,0))
plot(mu, l2, col="darkolivegreen4", lwd=1.5, type="l", ylab="", xlab="", ylim=c(-100,0))
```

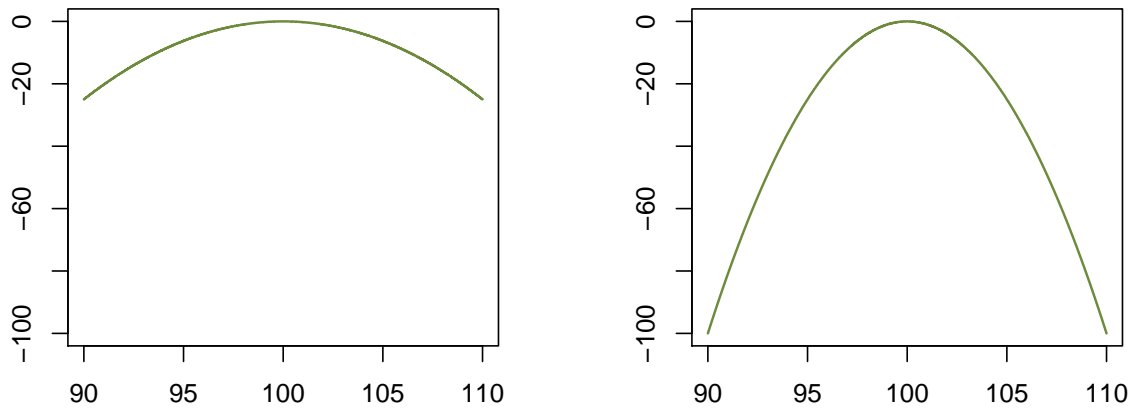


Figura 1: Función de verosimilitud normal con $n=2$ y $n=8$

4. Intervalos de confianza

En muestras grandes, la propiedad de normalidad de los estimadores máximo-verosímiles puede utilizarse para construir intervalos de confianza para el valor verdadero del parámetro. Como el estimador $\hat{\theta}$ es asintóticamente normal $(\theta, \sigma_{\hat{\theta}})$, donde $\sigma_{\hat{\theta}} = \sqrt{1/I(\hat{\theta})}$, tipificando el estimador tenemos que el siguiente estadístico es normal estándar en muestras grandes

$$Z = \frac{\hat{\theta} - \theta}{S_e}.$$

La obtención de un intervalo de confianza para θ consiste en buscar dos valores, $Z_{\alpha/2}$ y $Z_{1-\alpha/2}$, de la distribución normal estándar que dejan dentro de sí una probabilidad de $1 - \alpha$. Por tanto, la probabilidad de encontrar valores de Z dentro del intervalo es

$$P\left(Z_{\alpha/2} \leq \frac{\hat{\theta} - \theta}{S_e} \leq Z_{1-\alpha/2}\right) = 1 - \alpha$$

Despejando θ se obtiene que el intervalo de confianza para el parámetro es

$$P(L_i \leq \theta \leq L_s) = 1 - \alpha,$$

donde los límites superior e inferior del intervalo

$$\begin{aligned} L_s &= \hat{\theta} + |Z_{\alpha/2}| S_e \\ L_i &= \hat{\theta} - |Z_{\alpha/2}| S_e \end{aligned}$$

Los valores más habituales de $1 - \alpha$ son 0,95 o 0,99.

El intervalo de confianza es un resultado asintótico, si la muestra no es grande el intervalo puede ser incorrecto al no cumplirse los supuestos en que se basa el desarrollo, normalidad, estimador insesgado y con el S_e que se obtiene a partir de la función de información.

El intervalo de confianza se basa en una la definición frecuentista de probabilidad. Si el valor $1 - \alpha$ es, por ejemplo, 0,99, con el 99 % de las muestras que se tomen de la población se construirá un intervalo de confianza que incluya dentro de sí el valor verdadero de θ . Esto no significa que la probabilidad de que el intervalo calculado con una muestra concreta contenga θ sea del 99 %, porque dicha muestra incluye θ o no lo incluye, pero no existe una frecuencia de veces en que lo incluye. Por esto se dice *intervalo de confianza*, cuando trabajamos con una sola muestra en un estudio real, tenemos una *confianza* del $(1 - \alpha) \%$ de que sea una de las muestras que incluyen θ , no una probabilidad de que lo incluya. Los términos *verosimilitud* y *confianza* involucrados en el procedimiento de máxima-verosimilitud se utilizan para precisar aquellos aspectos en que no estamos utilizando una probabilidad frecuentista, sino algo más subjetivo.

Ejemplo 3. En una distribución normal el error típico del estimador de μ es

$$S_e = \frac{\sigma}{\sqrt{n}}.$$

Queremos obtener un intervalo de confianza al 99 %. Entonces $1 - \alpha = 0,99$, por lo que $\alpha = 0,01$ y $\alpha/2 = 0,005$. Por tanto $|Z_{0,005}| = 2,575$. Los límites del intervalo de confianza para μ son

$$\begin{aligned} L_s &= \bar{X} + 2,575 \frac{\sigma}{\sqrt{n}} \\ L_i &= \bar{X} - 2,575 \frac{\sigma}{\sqrt{n}}. \end{aligned}$$

Supongamos que $\sigma^2 = 15$ y se ha obtenido la muestra $\mathbf{x} = (102, 97, 110)$. El estimador de μ es $\hat{\mu} = 103$, y su varianza es:

$$\frac{\sigma^2}{n} = \frac{15}{3} = 5.$$

Entonces, con un nivel de confianza del 99 %, el valor verdadero de μ está dentro del intervalo:

$$\begin{aligned} L_s &= \bar{X} + 2,575 \frac{\sigma}{\sqrt{n}} = 103 + 2,575 \sqrt{5} = 108,75 \\ L_i &= \bar{X} - 2,575 \frac{\sigma}{\sqrt{n}} = 103 - 2,575 \sqrt{5} = 97,24 \end{aligned}$$

Esto no significa que exista una probabilidad de 0,99 de que μ esté entre 97,24 y 108,75. El valor verdadero de μ estará o no dentro de este intervalo. Lo que significa el intervalo de confianza es que en el 99 % de las muestras que tomemos de esta población, el intervalo resultante contendrá el valor verdadero de μ .

5. Cálculo de la varianza del estimador en lenguaje R

Para calcular la varianza del estimador en lenguaje R debemos utilizar la función *optim* y pedirle que nos de el valor de la segunda derivada evaluada en el punto correspondiente al estimador. Para pedirle a *optim* que nos de la segunda derivada utilizamos el argumento *hessian*. Supongamos que tenemos la muestra $\mathbf{x} = (2, 7, 3)$ procedente de una distribución Weibull. El siguiente código busca el estimador de forma numérica

```
x <- c(2, 7, 3)
lk <- function(lambda) -3*log(lambda) - sum(x) / lambda
fit <- optim(1, f=lk, method="Brent", lower=0, upper=10, control=list(fnscale=-1),
            hessian=TRUE)
print(fit)
```

```
## $par
## [1] 4
##
## $value
## [1] -7.158883
##
## $counts
## function gradient
##      NA      NA
##
## $convergence
## [1] 0
##
## $message
## NULL
##
## $hessian
##      [,1]
## [1,] -0.1875001
```

En la salida de resultados, objeto *fit* en el ejemplo, tenemos el valor de la segunda derivada en el campo *hessian*. A partir de ella podemos calcular la varianza, el error típico del estimador y el intervalo de confianza para el parámetro

```
lambda <- fit$par
var_lambda <- -1/fit$hessian
Se <- sqrt(var_lambda)

Z <- abs(qnorm(0.025))

Ls <- lambda + Z * Se
Li <- lambda - Z * Se

cat(sprintf(
  "El estimador es lambda = %5.3f con Se = %5.3f\nIntervalo de confianza: (%5.3f, %5.3f)\n",
  lambda, Se, Li, Ls))

## El estimador es lambda = 4.000 con Se = 2.309
## Intervalo de confianza: (-0.526, 8.526)
```

En el cálculo del intervalo de confianza se ha utilizado la función *qnorm* para obtener $|Z_{\alpha/2}|$. El comando *cat* muestra un mensaje por pantalla, y con *sprintf* se construye el mensaje de texto que muestra *cat*. Los argumentos *%5.3f* y *\n* pasados a *sprintf* tienen el siguiente sentido, *%5.3f* indica la posición del mensaje donde insertar el valor de una variable real y el formato en que se mostrará esta en cuanto a longitud y número de decimales, *\n* inserta un salto de línea en el mensaje.

6. Máxima verosimilitud con dos o más parámetros en lenguaje R

Cuando el modelo probabilístico incluye dos o más parámetros, θ no es un escalar sino un vector de parámetros, e $\mathbf{I}(\theta)$ es una matriz de información. La matriz de varianzas-covarianzas asintóticas de es la inversa de la información $\Sigma_{\theta} = \mathbf{I}^{-1}(\theta)$.

Supongamos que tomamos los datos de una distribución normal (μ, σ) , $\mathbf{y} = (90, 110, , 85, 105, 102, 108)$. Podemos estimar μ y σ con el siguiente código, que también proporciona la matriz de las derivadas segundas en el campo *hessian*.

```
y <- c(90, 110, 85, 105, 102, 120)

l <- function(p){
  logLk <- 0
  for(i in 1:length(y)){
    logLk <- logLk + log(dnorm(y[i], p[1], p[2]))
  }
  return(logLk)
}

fit <- optim(c(100,10), f=l, method="L-BFGS-B", lower=0, control=list(fnscale=-1),
            hessian=TRUE)

print(fit)

## $par
## [1] 102.00005 11.76153
##
## $value
## [1] -23.30263
##
## $counts
## function gradient
##      8      8
##
## $convergence
## [1] 0
##
## $message
## [1] "CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH"
##
## $hessian
##           [,1]      [,2]
## [1,] -4.337339e-02  3.397282e-07
## [2,]  3.397282e-07 -8.674647e-02
```

La matriz de información es $\mathbf{I}(\theta) = -\mathbf{H}$, siendo \mathbf{H} el hessiano o matriz de derivadas segundas. El siguiente código R calcula la información, la matriz de varianzas-covarianzas de los estimadores y la correlación entre estimadores. Para invertir la información se ha utilizado el comando *solve* de R.

```
H <- fit$hessian
Informacion <- - H
VarCovar <- solve(Informacion)
R <- cov2cor(VarCovar)
print(VarCovar)
```

```
##           [,1]      [,2]
## [1,] 2.305561e+01 9.029350e-05
## [2,] 9.029350e-05 1.152785e+01

print(R)

##           [,1]      [,2]
## [1,] 1.000000e+00 5.538524e-06
## [2,] 5.538524e-06 1.000000e+00
```

7. Estimación de un modelo de regresión en R

Supongamos que un psicólogo está estudiando una tarea de condicionamiento operante. Realiza varios ensayos de aprendizaje durante cinco días seguidos y sus variables son

- F es el número de ensayos en que un animal recibe una descarga eléctrica al cabo de un día.
- X es el número de días del experimento.
- Y es la intensidad de una señal acústica en db que avisa de que va a ocurrir una descarga.

El objetivo es ver cómo cambia F en función de X e Y . Utilizaremos la regresión de Poisson al ser la variable dependiente una frecuencia. La regresión de Poisson asume que para sujeto i el valor observado de la variable dependiente, f_i , sigue una distribución de Poisson con parámetro

$$\lambda_i = \exp(a + b_1 x_i + b_2 y_i)$$

El siguiente código R implementa la estimación del intercepto, a , y las pendientes b_1 y b_2 para unos datos de ejemplo. Además muestra los resultados, incluyendo la matriz de varianzas-covarianzas asintóticas, Σ , la matriz de correlaciones entre los parámetros estimados, \mathbf{R} y los errores típicos, Se , que son las raíces cuadradas de la diagonal de la matriz Σ .

```
f <- c(10, 9, 6, 5, 4)
x <- 1:5
y <- c(60, 40, 70, 30, 50)

l <- function(p){
  logLk <- 0
  for(i in 1:length(y)){
    lambda <- exp(p[1] + p[2]*x[i] + p[3]*y[i])
    logLk <- logLk + log(dpois(f[i], lambda))
  }
  return(logLk)
}

fit <- optim(c(0,0,0), f=l, control=list(fnscale=-1), hessian=TRUE)
print(fit)

## $par
## [1] 2.789541220 -0.252615341 -0.003472655
##
## $value
## [1] -9.335719
##
## $counts
## function gradient
```

```
##      290      NA
##
## $convergence
## [1] 0
##
## $message
## NULL
##
## $hessian
##           [,1]      [,2]      [,3]
## [1,] -33.99648 -85.99266 -1730.766
## [2,] -85.99266 -280.39814 -4165.238
## [3,] -1730.76572 -4165.23761 -94512.317

H <- fit$hessian
Informacion <- - H
VarCovar <- solve(Informacion)
R <- cov2cor(VarCovar)
print(VarCovar)

##           [,1]      [,2]      [,3]
## [1,]  0.75374113 -0.0756317125 -0.0104698020
## [2,] -0.07563171  0.0179161197  0.0005954343
## [3,] -0.01046980  0.0005954343  0.0001760686

print(R)

##           [,1]      [,2]      [,3]
## [1,]  1.0000000 -0.6508349 -0.9088379
## [2,] -0.6508349  1.0000000  0.3352517
## [3,] -0.9088379  0.3352517  1.0000000

Se <- sqrt(diag(VarCovar))
print(Se)

## [1] 0.86818266 0.13385111 0.01326908
```

8. Ejercicios

Ejercicio 1. Se ha tomado una m.a.s. de la distribución de Poisson y el resultado es $\mathbf{y} = \{4, 2, 6, 4\}$. Además sabemos que la función de log-verosimilitud de Poisson es

$$l(\lambda) = n\bar{Y} \log \lambda - n\lambda$$

1. Obtenga el estimador máximo-verosímil mediante desarrollo matemático, su varianza y el intervalo de confianza al 95 %.
2. Demuestre que el estimador es consistente.

Ejercicio 2. Se ha tomado una m.a.s. de la distribución de Poisson y el resultado es $\mathbf{y} = \{4, 2, 6, 4\}$. Además sabemos que la función de log-verosimilitud de Poisson es

$$l(\lambda) = n\bar{Y} \log \lambda - n\lambda$$

Obtenga el estimador máximo-verosímil de λ , su varianza y el intervalo de confianza al 95 % mediante optimización numérica en R.

Ejercicio 3. En una muestra aleatoria de la distribución exponencial se ha encontrado $\mathbf{x} = \{2, 5, 1, 5, 1, 25, 0, 75\}$. La función de log-verosimilitud exponencial es

$$l(\omega) = n \log \omega - \omega n \bar{X}$$

Obtenga el estimador máximo-verosímil mediante desarrollo matemático, su varianza y el intervalo de confianza al 99 %.

Ejercicio 4. En una muestra aleatoria de la distribución exponencial se ha encontrado $\mathbf{x} = \{2, 5, 1, 5, 1, 25, 0, 75\}$. La función de log-verosimilitud exponencial es

$$l(\omega) = n \log \omega - \omega n \bar{X}$$

Obtenga el estimador máximo-verosímil de ω , su varianza y el intervalo de confianza al 99 % mediante optimización numérica en R.

Ejercicio 5. Una variable sigue la distribución normal y sabemos que $\sigma = 2$. Un investigador desea realizar una estimación por intervalos de μ , con un nivel de confianza del 99 % y una anchura del intervalo de 0,5. ¿Cual debe ser el tamaño muestral del experimento?

Ejercicio 6. Un investigador está analizando la percepción de un estímulo luminoso en función de su nivel de saturación de color. La variable Y indica si el estímulo se percibe correctamente, y X es el grado de saturación. Los datos que ha obtenido son

$$\begin{aligned}\mathbf{y} &= \{0, 1, 1, 1, 0, 1, 0, 0, 0, 0\} \\ \mathbf{x} &= \{1, 7, 6, 4, 5, 3, 2, 5, 4, 2\}\end{aligned}$$

El investigador quiere aplicar una regresión logística, según la cual la función de probabilidad de x_i es Bernoulli con parámetro

$$\pi_i = \frac{1}{1 + \exp(\alpha + \beta x_i)}$$

Obtenga los estimadores de α y β , la matriz de varianzas-covarianzas asintóticas, la matriz de correlaciones entre los estimadores y los errores típicos utilizando R.