Estadística Actuarial I (CA-303): Laboratorio #2

Prof. Maikol Solís

Estimadores de máxima verosimilitud

Suponga que tenemos una tabla con los siguientes datos, los cuales representan la cantidad de giros hacia la derecha en cierta intersección.

```r
X <- c(rep(0,14), rep(1,30), rep(2,36), rep(3,68), rep(4, 43), rep(5,43),
       rep(6, 30), rep(7,14), rep(8,10), rep(9, 6), rep(10,4), rep(11,1), rep(12,1))
```

Queremos ajustar esta tabla a una distribución Poisson con función de densidad

$$\mathbb{P}(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Hemos visto en clase que el estimador teórico de máxima verosimilitud para λ es $\overline{X}$. Queremos estimar este parámetro alternativamente maximizando la función de verosimilitud.

Primero veamos la forma de los datos,

```r
hist(X, main="histograma del número de giros a la derecha", right=FALSE, prob=TRUE)
```

Definamos la función correspondiente a $-\log(\mathbb{P}(X = x))$

```r
n <- length(X)
negloglike <- function(lambda) {
  n * lambda - sum(X) * log(lambda) + sum( log( factorial(X) ) )
  }
```

Para encontrar el parámetro deseado, basta minimizar la función `negloglike` usando el la instrucción de optimización no lineal `nlm`.

```r
lambda.hat <- nlm(negloglike, p = c(0.5), hessian = TRUE)
```

Aquí el valor `p = c(0.5)` representa un valor inicial de búsqueda y `hessian = TRUE` determina el cálculo explícito de la segunda derivada.

Compare el resultado de `lambda.hat$estimate` con `mean(X)`.

Pruebas de hipótesis

Test con varianza desconocida

Recuerden que

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

tiene distribución t_{n-1}. Entonces, para el test $H_0 : \mu_0 < \mu$ vs $H_1 : \mu_0 \geq \mu$, rechazamos H_0 si $t \leq t_{n-1,1-\alpha}$

Queremos usar esta información en el siguiente problema:

> Un fabricante asegura que la vida media de sus bombillos supera las 10,000 horas. En una muestra de 30 bombillos se observa que estos solo duran 9,900 horas en promedio. Asuma que la desviación estandar muestral es 125 horas. Con un nivel de significancia de 5%, ¿Podemos rechazar lo que dice el fabricante?

```r
xbar <- 9900                          # media muestral
mu0 <- 10000                          # valor de la hipótesis
s <- 125                              # desviación muestral
n <- 30                               # tamaño muestral
t <- (xbar - mu0) / (s / sqrt(n)) # estadístico
alpha <- .05
t.alpha <- qt(alpha, df = n - 1)   # valor crítico
t <- t.alpha                           # prueba del estadístico vs el valor crítico
pval <- pt(t, df = n - 1)
pval <- alpha                          # prueba del p-valor vs el nivel de significancia
```

¿Qué se concluye de este análisis?

Versión alternativa con datos

Suponga que los de este problema son de la forma

```r
bulbs <- rnorm(n = 30, mean = 9900, sd = 125)
```

Podemos usar el comando `t.test` para realizar pruebas de hipotesis más fácilmente

```r
t.test(x = bulbs, alternative = "less", mu = 10000)
```

Intervalos de confianza

Varianza conocida

$$\overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Suponga que tenemos los siguientes datos

```r
library(MASS)
height.response = na.omit(survey$Height)
```

Atención: Si obtiene un error en las instrucciones anteriores ejecuten primero `install.packages(MASS)`

```r
n <- length(height.response)
sigma <- 9.48                      # asuma que esta es la desviación estandar verdadera
sem <- sigma/sqrt(n)               # desviación estándar de la media
E <- qnorm(1 - 0.05 / 2) * sem  # margen de error
xbar <- mean(height.response)   # media muestral
xbar + c(-E, E)                    # intervalo de confianza a un 95%
```

Varianza desconocida

$$\overline{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

En este caso tendríamos lo siguiente

```
n <- length(height.response)
s <- sd(height.response)                 # desviación estándar muestral
SE <- s / sqrt(n)                        # estimado del error estándar
E <- qt(1 - 0.05 / 2 , df = n - 1) * SE  # margen de error
xbar <- mean(height.response)            # media muestral
xbar + c(-E, E)                          # intervalo de confianza al 95%
```