Inferencia Estadística II: Trabajo T1

Sergio García Prado
sergio.garcia.prado@alumnos.uva.es

20 de octubre de 2018

1. Se lanzan seis monedas en cien ocasiones y se anota el número de caras en cada lanzamiento. Los resultados fueron:

```
coins <- data.frame(tosses = c(0, 1, 2, 3, 4, 5, 6),
                    freq = c(2, 8, 10, 12, 16, 30, 22))
```

Número de caras	0	1	2	3	4	5	6
Frecuencias	2	8	10	12	16	30	22

1.1. Obtener el *pvalor* del test de razón de verosimilitud para contrastar la siguiente hipótesis:

$$H_0 : \text{Todas las monedas tienen la misma probabilidad de cara}$$

Vamos a definir las siguientes variables aleatorias:

$$Y_1, ..., Y_6 \; iid \quad | \; Y_i \sim Bin(n, p_i)$$

Donde $n = 100$ es el número de realizaciones de la muestra, tal y como se indica en el enunciado y la variable Y_i representa el número de caras obtenidas por la moneda i-ésima. Por tanto, el contraste se puede reescribir utilizando esta notación de tal manera que la hipótesis sea:

$$H_0 : p_i = p_j \quad \forall i, j \in \{1, ..., 6\}$$

Sin embargo, esto no es lo que se proporciona en la muestra obtenida

```
n <- coins %>%
    summarise((max(tosses) * sum(freq))) %>%
    pull()
```

$$n = 600$$

```
p.hat <- coins %>%
    mutate(total = freq * tosses) %>%
    summarise(sum(total) / n) %>%
    pull()
```

$$\hat{p} = 0.6833$$

```
coins <- coins %>%
    mutate(freq.rel = freq / sum(freq))
```

Número de caras	0.00	1.00	2.0	3.00	4.00	5.0	6.00
Frecuencias	2.00	8.00	10.0	12.00	16.00	30.0	22.00
Frecuencias Relativas	0.02	0.08	0.1	0.12	0.16	0.3	0.22

```
coins <- coins %>%
    mutate(expected.freq.rel = dbinom(0:6, 6, p.hat))
```

Número de caras	0.000	1.0000	2.0000	3.0000	4.000	5.0000	6.0000
Frecuencias	2.000	8.0000	10.0000	12.0000	16.000	30.0000	22.0000
Frecuencias Relativas	0.020	0.0800	0.1000	0.1200	0.160	0.3000	0.2200
Frecuencias Relativas Esperadas	0.001	0.0131	0.0704	0.2026	0.328	0.2831	0.1018

```
G <- coins %>%
    summarise(2 * sum(expected.freq.rel * log(expected.freq.rel / freq.rel))) %>%
    pull()
```

$$G = 0.3906$$

```
pvalue <- pchisq(G, df=6)
```

$$p - valor = 0.0011$$

1.2. En el modelo que define la hipótesis nula obtener intervalos de confianza (95 %) para el parámetro, basados en los estadísticos de Wald (W) y de razón de verosimilitud (VR).

```
alpha <- 0.05
```

Cálculo por ecuación explícita

```
W.var <- p.hat * (1 - p.hat) / n
W.IC <- p.hat + c(-1, 1) * qnorm(1 - alpha / 2) * sqrt(W.var)
```

$$(0.64611, 0.72055)$$

Cálculo por optimización numérica

```
nloglhood <- function(p, n = 600, y = 410) {
    return( -(log(choose(n, y)) + y * log(p) + (n - y) *log(1 - p)) )
}
```

```
opt <- optim(0.5, nloglhood,lower = 0.0001,upper = 0.9999,
             hessian = TRUE, method = "L-BFGS-B")
phat <- opt$par
phat.var <- as.numeric(1 / opt$hessian)
```

1. 0.646111848300026 2. 0.720553688712397

```
control.list=list(label="p",est=p.hat,low=0,upp=1)
invisible(capture.output(LR.ci <- plkhci(control.list, nloglhood, "p")))
```

1. 0.645386613311019 2. 0.719718791742196

2. Considerar el vector aleatorio $X = (X_1, ..., X_5)$ con distribución multinomial, tal que $p = \left(\frac{1}{2}, \frac{\theta}{4}, \frac{1-\theta}{4}, \frac{1-\theta}{4}, \frac{\theta}{4}\right)$

Si el valor observado de Y es $y = (125, 18, 20, 34)$, usar 3 iteraciones del algoritmo EM para aproximar el estimador máximo verosímil de θ, partiendo del valor inicial $\theta^{(0)} = 0.5$.

```
# TODO
```