

Inferencia Estadística II: Trabajo T1

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

22 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(hits = 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 *p*valor del test de razón de verosimilitud para contrastar la siguiente hipótesis:

H_0 : Todas las monedas tienen la misma probabilidad de cara.

Vamos a definir las siguientes variables aleatorias:

$$Y_1, \dots, Y_6 \text{ iid} \mid Y_i \sim \text{Bin}(n, p_i)$$

Donde $n = 100$ es el número de realizaciones de la muestra, tal 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 nula H_0 sea:

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

Por la propiedad de independencia de las variables Y_i entre si, podemos redefinir dicho test apoyándonos en:

$$X = \sum_{i=1}^6 Y_i$$

Donde $m = n \cdot 6 = 600$ y $p_0 = p_i \quad \forall i \in \{1, \dots, 6\}$. Por lo tanto, podemos redefinir el contraste como de bondad de ajuste, donde la hipótesis nula se transforma en:

$$H_0 : X \sim \text{Bin}(m, p_0)$$

Frente a la hipótesis alternativa de que el X se distribuya de manera diferente.

Para contrastar esta hipótesis utilizaremos un *test G* (o *test χ^2*) basado en la comparación entre frecuencias observadas y esperadas, para lo cual nos apoyaremos en la distribución multinomial.

Para ello, lo primero es calcular el número de observaciones:

```
m <- coins %>%
  summarise((max(hits) * sum(freq))) %>%
  pull()
```

$$m = 600$$

Puesto que vamos a realizar el test utilizando frecuencias relativas, es necesario calcular estas sobre los datos observados

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

Número de caras	0	1	2	3	4	5	6
Frecuencias	2	8	10	12	16	30	22
Frecuencias Rel.	0.02	0.08	0.1	0.12	0.16	0.3	0.22

El siguiente paso es obtener el *Estimador Máximo Verosimil* bajo la hipótesis nula. Esto se puede hacer tanto de manera analítica, como de manera numérica a partir de la función `optim`.

Se puede comprobar fácilmente que el EMV para una distribución binomial se construye como $\hat{p}_0 = \frac{\sum x_i}{m}$, cuya varianza aproximada es $\widehat{Var}[\hat{p}_0] = \frac{\hat{p}_0(1-\hat{p}_0)}{m}$. A continuación se calculan dichos valores:

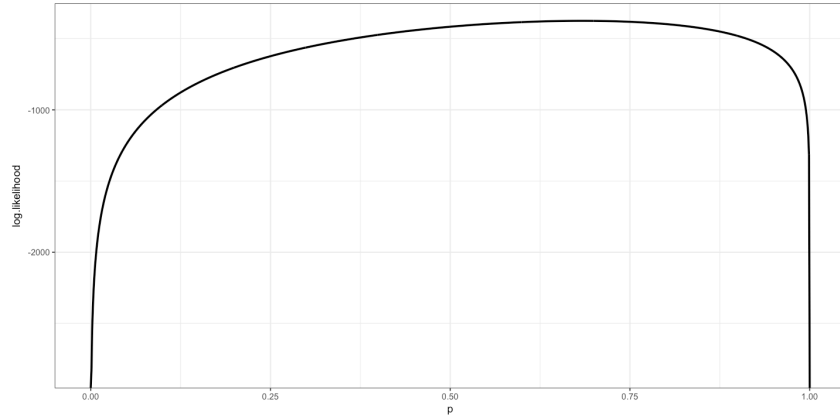
```
p.zero.hat <- coins %>%
  mutate(total = freq * hits) %>%
  summarise(sum(total) / m) %>%
  pull()
p.zero.hat.var <- p.zero.hat * (1 - p.zero.hat) / m
```

$$\hat{p}_0 = 0.68333 \quad \widehat{Var}[\hat{p}_0] = 0.00036$$

Sin embargo, en este caso vamos a optimizar el ajuste numérico a partir de la función `optim`. Para ello es necesario definir la función $-\log L(p; x_1, \dots, x_m)$.

```
loglikelihood <- function(p, y, m) {
  l <- y * log(p) + (m - y) * log(1 - p)
  return(l)
}
negative.loglikelihood <- function(p, y, m) {
  - loglikelihood(p, y, m)
}
```

A continuación se muestra la función $\log L(p; x_1, \dots, x_m)$. Es fácil apreciar que esta toma su valor máximo en torno a $p = 0.68$, tal y como hemos indicado anteriormente.



Una vez definida la función de verosimilitud, basta con asignar unos valores frontera, junto con el valor inicial, para obtener el óptimo, así como el *hessiano*, a partir del cual aproximaremos la varianza del estimador \hat{p}_0 .

```
opt <- optim(runif(1), negative.loglikelihood, hessian = TRUE,
            y = sum(coins$hits * coins$freq), m = m,
            lower = 10e-4, upper = 1-10e-4, method = "L-BFGS-B")
```

```
p.zero.hat <- opt$par
p.zero.hat.var <- c(solve(opt$hessian))
```

$$\hat{p}_0 = 0.68333 \quad \widehat{Var}[\hat{p}_0] = 0.00036$$

Como vemos, estos valores son equivalente a los obtenidos mediante el método analítico. Sin embargo, hay muchos casos en que dichos métodos son inescalables cuando el estadístico en cuestión es demasiado complejo de definir.

Una vez hemos calculado el *EMV* bajo la hipótesis nula, ya podemos calcular el las frecuencias esperadas que deberían seguir nuestras observaciones aproximadamente para poder verificar que todas ellas pertenecen a una misma distribución Binomial con parámetro p_0 .

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

Número de caras	0	1	2	3	4	5	6
Frecuencias	2	8	10	12	16	30	22
Frecuencias Rel.	0.02	0.08	0.1	0.12	0.16	0.3	0.22
Frecuencias Rel. Esperadas	3e-04	0.0048	0.0312	0.1123	0.2423	0.3138	0.2257

El siguiente paso es obtener el estadístico test G. Este se obtiene a partir del test de razón de verosimilitud de la siguiente forma:

Sea:

$$L(p; x_1, \dots, x_m) \propto p^{\sum x_i} \cdot (1 - p)^{m - \sum x_i}$$

Por lo que:

$$\log L(p; x_1, \dots, x_m) \propto \sum x_i \cdot \log(p) + (m - \sum x_i) \cdot \log(1 - p)$$

El estadístico test de razón de verosimilitud se puede escribir por tanto como:

$$\begin{aligned} G &= -2 \cdot \log(\Delta(x_1, \dots, x_m)) \\ &= -2 \cdot \log\left(\frac{L(\hat{p}_0; x_1, \dots, x_m)}{L(\hat{p}_{obs}; x_1, \dots, x_m)}\right) \\ &= 2 \cdot (L(\hat{p}_{obs}; x_1, \dots, x_m) - L(\hat{p}_0; x_1, \dots, x_m)) \\ &= 2 \cdot \sum_i y_i \cdot \log\left(\frac{\hat{p}_{obs,i}}{\hat{p}_{0,i}}\right) \end{aligned}$$

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

$$G = 9.903$$

Entonces, para calcular el *pvalor* del test, basta comparar este estadístico con su distribución bajo H_0 , en este caso una χ^2 con $7 - 1 = 6$ grados de libertad, ya que el espacio paramétrico de partida es de 7 valores, mientras que bajo la hipótesis nula, este queda reducido a 1 único parámetro. Por tanto:

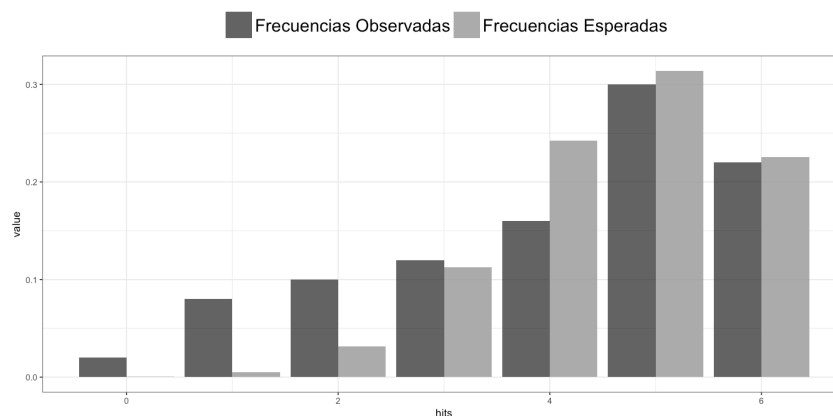
$$pvalor = P(G \geq \chi_6^2)$$

```
pvalue <- 1 - pchisq(G, df=6)
```

$$pvalor = 0.1288$$

Tal y como indica el *pvalor*, con una confianza del 95 % no tenemos indicios suficientes como para rechazar la hipótesis de que todas las observaciones obtenidas provienen todas ellas de una distribución de 6 monedas, todas ellas distribuidas de la misma forma (igual tasa de acierto).

A continuación se muestra un gráfico de barras que representa de manera conjunta las frecuencias observadas y esperadas:



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 (RV)*.

```
alpha <- 0.05
```

Ahora vamos a calcular los intervalos de confianza basados en la estimación de *Wald* y en *razón de verosimilitud*. Estos nos proporcionarán un entorno de puntos sobre los cuales el verdadero valor de p se encontraría bajo la hipótesis nula con una probabilidad del $(1 - \alpha) * 100 \%$. Estos intervalos se contruyen de manera diferente, tal y como indicaremos a continuación:

1.2.1. Intervalo de confianza de Wald:

El Intervalo de confianza basado en inferencias de *Wald* se construye de la siguiente manera. Sabemos que la distribución asintótica estimada es:

$$\hat{p} \sim N(p, V(p)) \quad (1)$$

$$\frac{\hat{p} - p}{\sqrt{V(p)}} \sim N(0, 1) \quad (2)$$

$$(3)$$

También sabemos que:

$$\frac{\hat{p} \cdot (1 - \hat{p})}{n} \rightarrow V(p)$$

Luego:

$$\frac{\hat{p} - p}{\sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}} \sim N(0, 1)$$

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

(0.64611, 0.72055)

1.2.2. Intervalo de confianza de Razón de Verosimilitud:

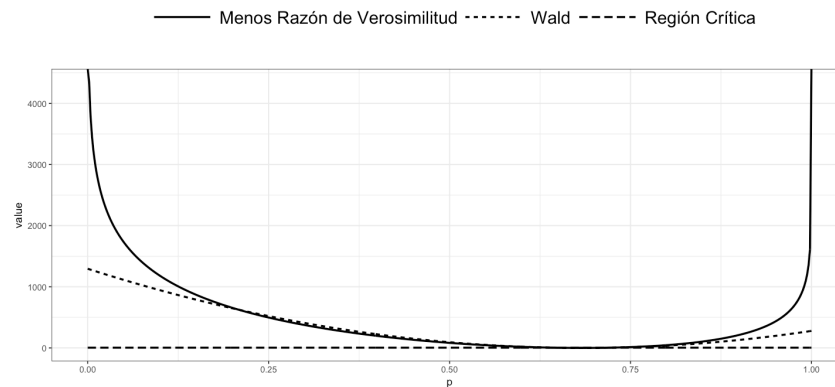
[TODO explicar intervalo TRV]

```
control.list <- list(label = "p", est = p.zero.hat, low = 0, upp = 1)
LR.ci <- Plkhci(control.list, negative.loglikelihood, "p",
               prob = 0.95, y = 410, m=600)
```

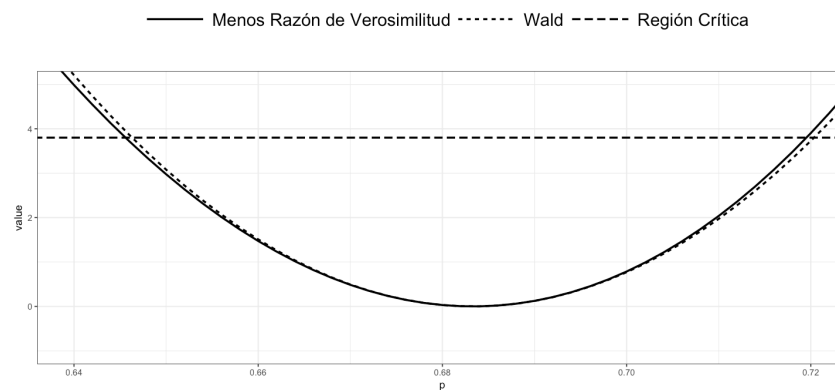
(0.64539, 0.71972)

```
coins.test <- data.frame(p = seq(0, 1, by=0.002)) %>%
  mutate(LR = - 2 * (loglikelihood(p, 410, 600)
    - loglikelihood(p.zero.hat, 410, 600)),
    W = (p - p.zero.hat) ^ 2 / c(p.zero.hat.var),
    chi = sqrt(qchisq(1 - alpha / 2, df = 6))
  )
```

[TODO explicar gráfico]



[TODO explicar gráfico]



2. Considerar el vector aleatorio $X = (X_1, \dots, X_5)$

Supongamos que X (vector aleatorio de datos completos) se distribuye sobre una distribución multinomial, tal que:

$$X = (X_1, X_2, X_3, X_4, X_5) \sim \text{multinomial} \left(n, \left(\frac{1}{2}, \frac{\theta}{4}, \frac{1-\theta}{4}, \frac{1-\theta}{4}, \frac{\theta}{4} \right) \right)$$

La función de verosimilitud en este caso será:

$$\begin{aligned}
L(\theta; x_1, \dots, x_5) &= \left(\frac{1}{2}\right)^{x_1} \cdot \left(\frac{\theta}{4}\right)^{x_2} \cdot \left(\frac{1-\theta}{4}\right)^{x_3} \cdot \left(\frac{1-\theta}{4}\right)^{x_4} \cdot \left(\frac{\theta}{4}\right)^{x_5} \\
&\propto \left(\frac{\theta}{4}\right)^{x_2} \cdot \left(\frac{1-\theta}{4}\right)^{x_3} \cdot \left(\frac{1-\theta}{4}\right)^{x_4} \cdot \left(\frac{\theta}{4}\right)^{x_5} \\
&= \left(\frac{\theta}{4}\right)^{x_2+x_4} \cdot \left(\frac{1-\theta}{4}\right)^{x_3+x_5} \\
&\propto \theta^{x_2+x_5} \cdot (1-\theta)^{x_3+x_4}
\end{aligned}$$

Por lo que:

$$\log L(\theta; x_1, \dots, x_5) \propto (x_2 + x_5) \cdot \log(\theta) + (x_3 + x_4) \cdot \log(1 - \theta)$$

Se define Y (vector de datos observados o incompletos) a partir de X como:

$$\begin{aligned}
Y &= (Y_1, Y_2, Y_3, Y_4) \\
&= (X_1 + X_2, X_3, X_4, X_5) \sim \text{multinomial} \left(n, \left(\frac{1}{2} + \frac{\theta}{4}, \frac{1-\theta}{4}, \frac{1-\theta}{4}, \frac{\theta}{4} \right) \right)
\end{aligned}$$

La función de verosimilitud en este caso será:

$$\begin{aligned}
L(\theta; y_1, \dots, y_4) &= L(p; x_1 + x_2, x_3, x_4, x_5) \\
&= \left(\frac{1}{2} + \frac{\theta}{4}\right)^{y_1} \cdot \left(\frac{1-\theta}{4}\right)^{y_2} \cdot \left(\frac{1-\theta}{4}\right)^{y_3} \cdot \left(\frac{\theta}{4}\right)^{y_4} \\
&= \left(\frac{1}{2} + \frac{\theta}{4}\right)^{y_1} \cdot \left(\frac{1-\theta}{4}\right)^{y_2+y_3} \cdot \left(\frac{\theta}{4}\right)^{y_4} \\
&\propto (2 + \theta)^{y_1} \cdot (1 - \theta)^{y_2+y_3} \cdot \theta^{y_4}
\end{aligned}$$

Por lo que:

$$\begin{aligned}
\log L(\theta; y_1, \dots, y_4) &= \log L(\theta; x_1 + x_2, x_3, x_4, x_5) \\
&= (y_1) \cdot \log(2 + \theta) + (y_2 + y_3) \cdot \log(1 - \theta) + (y_4) \cdot \log(\theta)
\end{aligned}$$

Si intentamos reescribir la función de verosimilitud de X en función de Y tenemos que:

$$\log L(\theta; x_1, x_2, y_2, y_3, y_4) \propto (x_2 + y_4) \cdot \log(\theta) + (y_2 + y_3) \cdot \log(1 - \theta)$$

Ahora necesitamos calcular la esperanza de la función de log verosimilitud conjunta:

$$\begin{aligned}
E_{\theta^{(0)}} [X \mid Y = y] &= E[(X_2 + X_5) \cdot \log(\theta) + (X_3 + X_4) \cdot \log(1 - \theta) \mid Y = y] \\
&= (E[X_2 \mid Y_1 = y_1] + y_4) \cdot \log(\theta) + (y_2 + y_3) \cdot \log(1 - \theta) \\
&= \left(\frac{\frac{\theta}{4}}{\frac{1}{2} + \frac{\theta}{4}} \cdot y_1 + y_4 \right) \cdot \log(\theta) + (y_2 + y_3) \cdot \log(1 - \theta) \\
&= \left(\frac{\theta^{(0)}}{2 + \theta^{(0)}} \cdot y_1 + y_4 \right) \cdot \log(\theta) + (y_2 + y_3) \cdot \log(1 - \theta)
\end{aligned}$$

Por último, es necesario maximizar dicho valor. Para ello tomaremos derivadas, igualaremos a cero y despejaremos el parámetro θ :

$$\begin{aligned}\frac{\partial}{\partial \theta} E_{\theta^{(0)}}[X | Y = y] &= \frac{\partial}{\partial \theta} \left(\left(\frac{\theta^{(0)}}{2 + \theta^{(0)}} \cdot y_1 + y_4 \right) \cdot \log(\theta) + (y_2 + y_3) \cdot \log(1 - \theta) \right) \\ &= \frac{\frac{\theta^{(0)}}{2 + \theta^{(0)}} \cdot y_1 + y_4}{\theta} + \frac{-(y_2 + y_3)}{1 - \theta}\end{aligned}$$

Ahora despejamos el parámetro θ para llegar a una ecuación explícita:

$$\begin{aligned}0 &= \frac{\frac{\theta^{(0)}}{2 + \theta^{(0)}} \cdot y_1 + y_4}{\theta} + \frac{-(y_2 + y_3)}{1 - \theta} \\ \frac{y_2 + y_3}{1 - \theta} &= \frac{\frac{\theta^{(0)}}{2 + \theta^{(0)}} \cdot y_1 + y_4}{\theta} \\ \theta \cdot \left(\frac{\theta^{(0)}}{2 + \theta^{(0)}} \cdot y_1 + y_4 + y_2 + y_3 \right) &= \frac{\theta^{(0)}}{2 + \theta^{(0)}} \cdot y_1 + y_4 \\ \theta &= \frac{\frac{\theta^{(0)}}{2 + \theta^{(0)}} \cdot y_1 + y_4}{\frac{\theta^{(0)}}{2 + \theta^{(0)}} \cdot y_1 + y_4 + y_2 + y_3}\end{aligned}$$

```
update.theta <- function(theta, y) {
  temp <- (theta * y[1]) / (2 + theta)
  theta <- (temp + y[4]) / (temp + y[2] + y[3] + y[4])
  return(theta)
}
```

Ahora pondremos en práctica el algoritmo EM. Se observa:

$$y = (125, 18, 20, 34)$$

Se pide aproximar el estimador máximo verosímil de θ , utilizando 3 iteraciones del *algoritmo EM*, partiendo del valor inicial $\theta^{(0)} = 0.5$.

```
y <- c(125, 18, 20, 24)
theta.zero <- 0.5
iterations <- 3
```

Por tanto, ejecutamos el algoritmo:

```
theta <- runif(1)
for (i in 1:iterations) {
  theta <- update.theta(theta, y)
}
```

[TODO]

$$\hat{\theta}^{(3)} = 0.5792$$