

Prueba de rachas

López, Corral, Carleos

18 de abril de 2022

Índice

- Variable X dicótoma.
 - Toma valores A y B .
- R = número de rachas en la muestra $\vec{X} = (X_1, \dots, X_n)$.

- $\vec{X} = AAABBBBBBBBAB \implies R = 4$

- $n_A = \#\{X_i = A\} > 0$

- $n_B = \#\{X_i = B\} > 0$

- R puede tomar valores $r \in \{2, \dots, c\}$

- $n_A = n_B \implies c = n$

- $n_A \neq n_B \implies c = 1 + 2 \min\{n_A, n_B\}$

- $\Pr[R = r]$

- r par, $r = 2s$

$$\frac{2 \binom{n_A-1}{s-1} \binom{n_B-1}{s-1}}{\binom{n}{n_A, n_B}}$$

- r impar, $r = 2s + 1$

$$\frac{\binom{n_A-1}{s} \binom{n_B-1}{s-1} + \binom{n_B-1}{s} \binom{n_A-1}{s-1}}{\binom{n}{n_A, n_B}}$$

- Si $n \gg 0$, entonces $R \xrightarrow{\sim} \mathcal{N}(E[R], D[R])$

- $Y_i = [X_i \neq X_{i-1}]$ para $i \in \{2, \dots, n\}$

- $R = 1 + \sum_{i=2}^n Y_i$

- $Y_i \hookrightarrow \mathcal{B}\left(\frac{2n_A n_B}{n(n-1)}\right)$ pero dependientes pues $R \leq c$

- $i \neq j \implies E(Y_i Y_j) = \Pr[Y_i = 1 \cup Y_j = 1] =$

$$= \begin{cases} \frac{n_A}{n} \frac{n_B}{n-1} \frac{n_A-1}{n-2} + \frac{n_B}{n} \frac{n_A}{n-1} \frac{n_B-1}{n-2} = \frac{n_A n_B (n_A-1 + n_B-1)}{n(n-1)(n-2)} = \frac{n_A n_B}{n(n-1)} & |i-j| = 1 \\ \frac{4n_A n_B (n_A-1)(n_B-1)}{n(n-1)(n-2)(n-3)} & |i-j| > 1 \end{cases}$$

- $E(R) = 1 + \frac{2n_A n_B}{n}$

- $\text{Var}(R) = \frac{2n_A n_B (2n_A n_B - n)}{n^2 (n-1)}$

```

/* Obtener Var(R) con Maxima */
n : nA+nB $
EYi : (2*nA*nB)/(n*(n-1)) $
VYi : factor(EYi*(1-EYi)) $
Svar : (n-1)*VYi $
EYiYi1 : (nA*nB)/(n*(n-1)) $ /* E(Y[i]Y[i+1]) */
EYiYj : 4*nA*nB*(nA-1)*(nB-1)/(n*(n-1)*(n-2)*(n-3)) $ /* |i-j| > 1 */
nScov : binomial(n-1,2) $ /* número de sumandos con covarianzas */
Smedprod : (n-2)*EYiYi1 + (nScov-(n-2))*EYiYj $
Sprodmed : nScov * EYi * EYi $
Scov : 2 * (Smedprod - Sprodmed) $
VR : Svar + Scov $
factor(VR) ;
/* EJECUCIÓN: */
Maxima 5.43.2 http://maxima.sourceforge.net
using Lisp GNU Common Lisp (GCL) GCL 2.6.12
Distributed under the GNU Public License. See the file COPYING.
Dedicated to the memory of William Schelter.
The function bug_report() provides bug reporting information.

```

$$\begin{array}{r}
2 n_A n_B (2 n_A n_B - n_B - n_A) \\
\hline
(n_B + n_A - 1) (n_B + n_A)
\end{array}$$

```

## Ejemplo en R
install.packages("snpar")
set.seed(13) # para p-valor polémico
n <- 10
(X <- sample (c("A","B"), n, TRUE)) # bajo H0
(rachas <- rle(X))
(R <- length(rachas$values)) # número de rachas
(Xn <- +(X=="A")) # runs.test exige vector numérico
snpar::runs.test(Xn)
snpar::runs.test(Xn, exact=TRUE) # p-valor cambia
## EJECUCIÓN:
R version 4.1.1 (2021-08-10) -- "Kick Things"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

```

R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribución.

R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.

Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.

Escriba 'q()' para salir de R.

```
> set.seed(13) # para p-valor polémico
> n <- 10
> (X <- sample (c("A","B"), n, TRUE)) # bajo H0
```

```
[1] "B" "A" "B" "A" "B" "A" "B" "B" "B" "B"
> (rachas <- rle(X))
```

Run Length Encoding

```
lengths: int [1:7] 1 1 1 1 1 1 4
values : chr [1:7] "B" "A" "B" "A" "B" "A" "B"
> (R <- length(rachas$values)) # número de rachas
```

```
[1] 7
> (Xn <- +(X=="A")) # runs.test exige vector numérico
```

```
[1] 0 1 0 1 0 1 0 0 0 0
> snpar::runs.test(Xn)
```

Approximate runs test

```
data: Xn
Runs = 7, p-value = 0.1408
alternative hypothesis: two.sided
```

```
> snpar::runs.test(Xn, exact=TRUE) # p-valor cambia
```

Exact runs test

```
data: Xn
Runs = 7, p-value < 2.2e-16
alternative hypothesis: two.sided
```