

Examen computacional sobre Test de Hipótesis

Métodos estadísticos en física experimental

No son todos iguales

A. Rabinovich (LU:316/08)

*Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires,
Pabellón I, Ciudad Universitaria, 1428 Buenos Aires, Argentina.*

(Dated: July 19, 2017)

I. INTRODUCCIÓN

Sean dos hipótesis especificadas completamente por dos valores distintos de un parámetro θ en una función de distribución de probabilidad $f(x|\theta)$. Para ejemplificar, la hipótesis nula, H_0 , supone que $\theta = \theta_0$ mientras que la alternativa, H_1 , asume $\theta = \theta_1$.

Suponiendo que la hipótesis nula es cierta, es posible encontrar una región R en el espacio muestral W para la observación x tal que la probabilidad de que x pertenezca a R es igual a algún valor numérico asignado previamente. La región R es llamada la región de rechazo o región crítica para H_0 , mientras que $(W - R)$ es la región de aceptación de H_0 .

Las dos regiones están separadas por un valor x_c y se llama significancia a la probabilidad preasignada de que una observación x pertenezca a R . Ésta determina el nivel de significancia $100\alpha\%$.

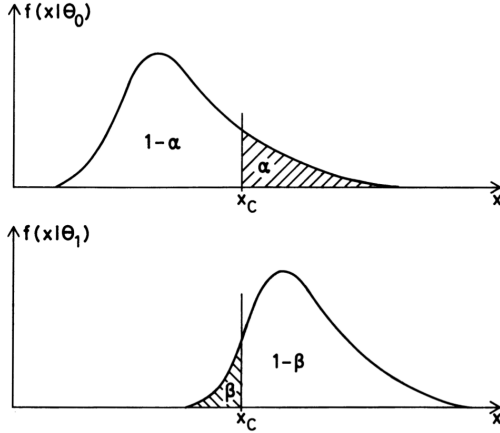


FIG. 1: Ejemplo de Error de Tipo I, Error de Tipo II y poder de test. Tomado de Froedsen[1].

De ésta definición se desprende que existe una probabilidad α de rechazar H_0 cuando era en realidad verdadera, error conocido como Error de Tipo I. Por otro lado, el Error de Tipo II se define como la probabilidad de aceptar H_0 cuando era en realidad falsa, y la probabilidad β de que ocurra depende de H_1 . Finalmente, el poder de un test se define como la probabilidad $(1 - \beta)$ de rechazar H_0 cuando era realmente falsa. El poder de un test depende de la cantidad de muestras, el nivel de significancia y el tamaño del efecto. Todas éstas definiciones se ejemplifican en la figura 1.[1]

II. EL PROBLEMA

Sean dos muestras independientes X_1, \dots, X_m e Y_1, \dots, Y_n con distribuciones gaussianas $N(\mu_x, \sigma^2)$ y $N(\mu_y, \sigma^2)$ con μ_x, μ_y, σ desconocidas.

Se quiere estudiar computacionalmente cual de los siguientes test es el de mayor potencia para testear a un nivel de significancia α la hipótesis $H_0: \mu_x \leq \mu_y$ contra $H_1: \mu_x > \mu_y$.

El primero consiste en rechazar H_0 cuando $U \geq T_{m+n-2, (1-\alpha)}$, donde:

$$U_1 = (\bar{X} - \bar{Y}) \sqrt{\frac{m+n-2}{(\frac{1}{m} + \frac{1}{n})(S_x^2 + S_y^2)}} \quad (1)$$

con $S_x^2 = \sum_i (x_i - \bar{X})^2$ y $S_y^2 = \sum_i (y_i - \bar{Y})^2$ y $T_{m+n-2, (1-\alpha)}$ es el cuantil $(1 - \alpha)$ de la distribución T con $n + m - 2$ grados de libertad.

El segundo consiste en rechazar H_0 cuando $U \geq T_{dof, (1-\alpha)}$, donde:

$$U_2 = \frac{(\bar{X} - \bar{Y})}{S} \quad (2)$$

$$\text{con } S^2 = \frac{S_x^2}{m(m-1)} + \frac{S_y^2}{n(n-1)} \text{ y}$$

$$dof \sim \frac{(\frac{S_x^2}{m(m-1)} + \frac{S_y^2}{n(n-1)})^2}{(\frac{S_x^2}{m(m-1)})^2/(m-1) + (\frac{S_y^2}{n(n-1)})^2/(n-1)}$$

III. SIMULACIONES

Se tomaron m muestras para $X \sim N(0.1, 1)$ y n muestras para $Y \sim N(0, 1)$, variando m y n desde 10 hasta 1100 cada una (es decir, en una iteración anidada dentro de la otra) y se calcularon los estadísticos U_1 y U_2 . Luego, para cada par m y n se obtuvo el p -value definido como $1 -$ la probabilidad acumulada de la T correspondiente hasta U . Ésto se repitió 100 veces y se calculó el promedio del p -value para cada par, como se muestra en la figura ??.

Se observa que los p -values obtenidos con ámbos tests son muy similares, y en ambos casos fue necesario tomar aproximadamente 900 muestras de cada distribución normal para obtener un p -value < 0.05 .

Se realizó el mismo proceso tomando muestras de

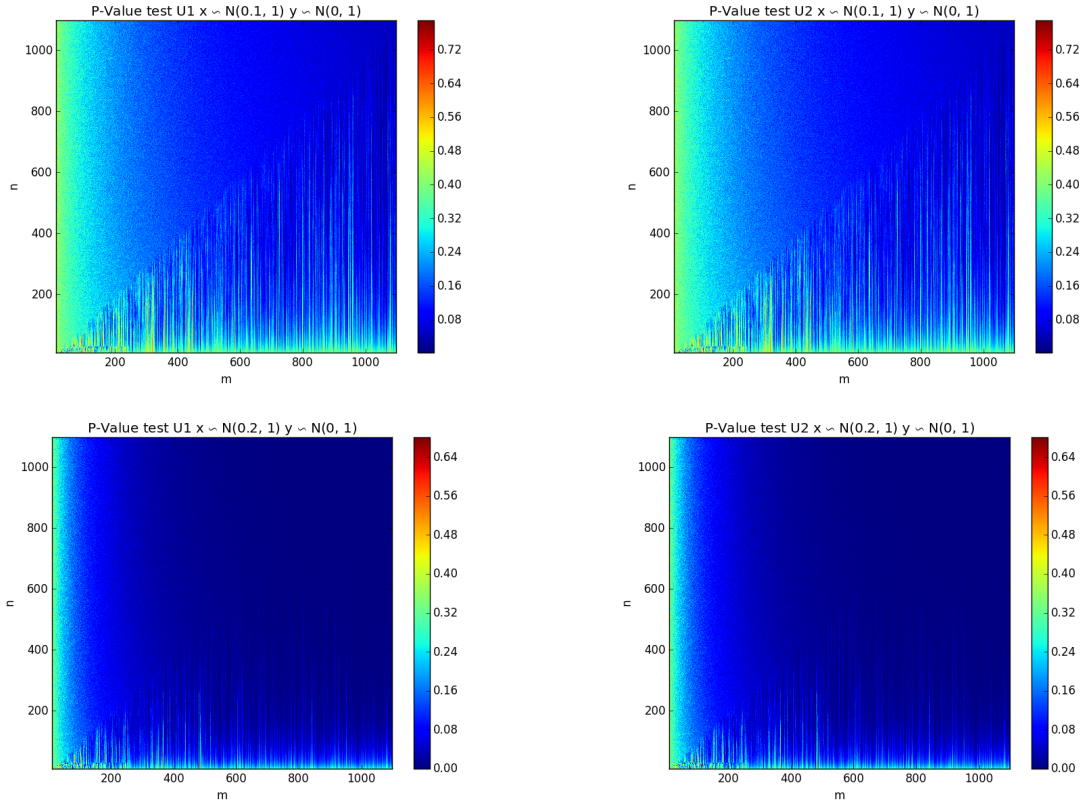


FIG. 2: P-Value para distintos tests, tamaños de muestra y efectos

$X \sim N(0.2, 1)$ y graficando el p -value, figura ??, y se encontró que con tomar alrededor de 300 muestras de cada distribución normal era suficiente. Ésto concuerda con la dependencia del test con el tamaño del efecto. A un efecto mayor (mayor diferencia entre las medias), se requiere una menor cantidad de muestras para obtener la significancia necesaria.

Por otro lado, se midió el poder de cada test tomando 2000 muestras de $U1$ y $U2$ para cada n y m , con $\mu_x > \mu_y$ y con $\mu_x < \mu_y$. En las figuras ?? y ?? se observan histogramas para $n = m = 100$, $n = m = 500$ y $n = m = 900$ para diferencia de medias de 0.1 y $n = m = 10$, $n = m = 100$ y $n = m = 300$ para diferencia de medias de 0.2 respectivamente. En rojo, x_c .

Se verifica que a medida que aumenta el tamaño de

las muestras, ambas distribuciones se separan. También se observa que a mayor efecto, menor cantidad de muestras alcanza para separar las distribuciones. La figura ?? muestra el poder de cada test y su dependencia con el tamaño de la muestra. Nuevamente, se observa que para tamaños de efecto mayores se requiere de una menor cantidad de muestras para conseguir aumentar el poder del test.

Se observa además que basta con aumentar el tamaño de las muestras de x para lograr aumentar la potencia del test.

IV. CONCLUSIONES

ACKNOWLEDGMENTS

A. Rabinovich es becario doctoral del CONICET.

[1] Frodesen, *Probability and statistics in particle physics* (UNIVERSITETSFORLAGET).

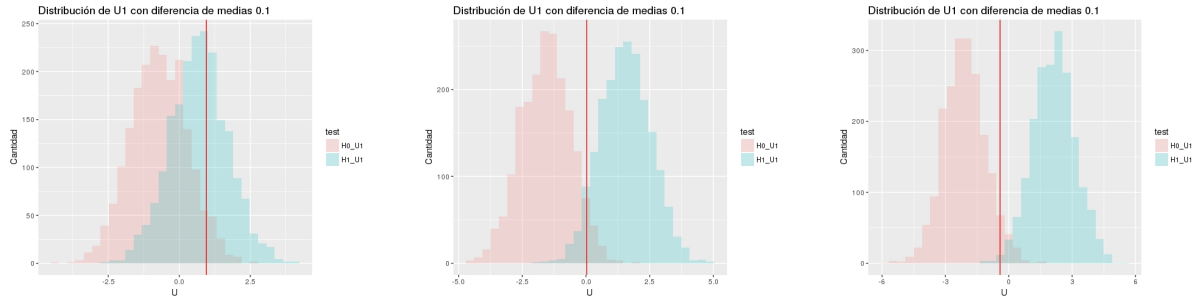


FIG. 3: Histogramas de H_0 y H_1 para test U_1 con diferencia de media de 0.1 para $n=m$ con $n=100$, 500 y 900 respectivamente. Se observa que las distribuciones se separan a medida que aumenta el tamaño de la muestra.

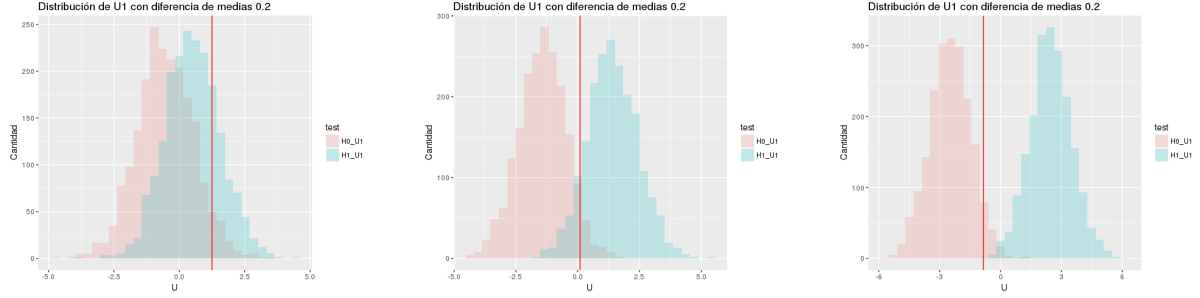


FIG. 4: Histogramas de H_0 y H_1 para test U_1 con diferencia de media de 0.2 para $n=m$ con $n=10$, 100 y 300 respectivamente. Se observa que las distribuciones se separan a medida que aumenta el tamaño de la muestra.

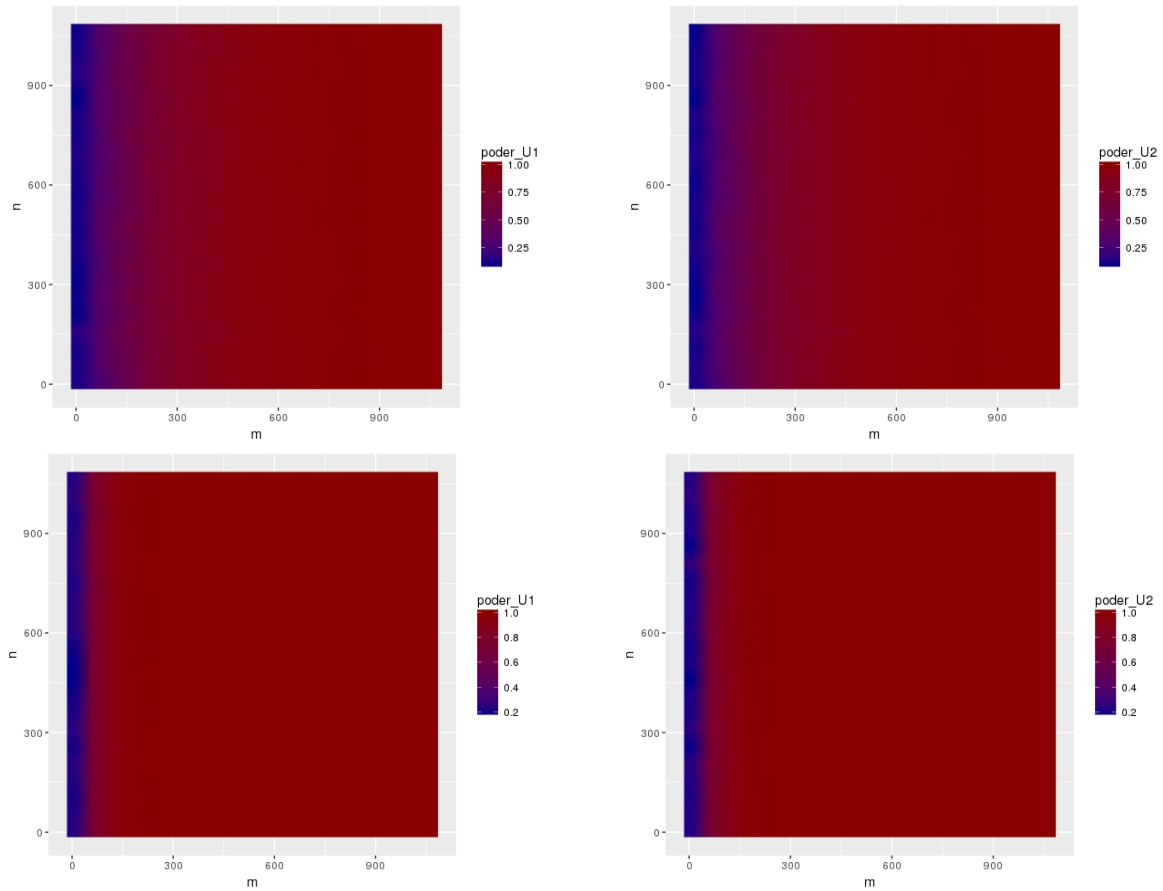


FIG. 5: Poder para distintos tests, tamaños de muestra y efectos