

Distribución del estadístico W^+ de Wilcoxon. Eficiencia de estimadores de centro de simetría

Diego F. Fonseca. V. & Roodolfo. Quintero *

Resumen. En el presente documento abordamos dos situaciones de especial interés en el campo de la estadística no paramétrica, uno es indagar sobre aproximación normal a la distribución W^+ por medio de simulaciones de muestras en donde se evidenciara dicho comportamiento. El otro problema indaga sobre la eficiencia de estimadores de centro de simetría, para tal fin estimaremos su potencia y de acuerdo a dicho valor determinaremos la eficiencia.

Palabras clave: Estadístico, potencia, aproximación.

1 Introducción

La primera situación que abordaremos es sobre aproximación normal a la distribución W^+ , para tal fin realizaremos simulaciones de muestras en donde se evidenciara dicho comportamiento. El otro problema indaga sobre la eficiencia de estimadores de centro de simetría, en ese sentido estimaremos su potencia su potencia por medio de simulaciones. A continuación introducimos en detalle cada problema.

1.1 Problema 1: Aproximación normal a la distribución W^+

En [1] se intenta establecer una versión análoga del teorema del límite central pero en este caso para estadísticos dentro de la teoría de estadísticos no paramétricos (U -estadísticos), en dicho contexto toma especial relevancia el estadístico de rangos signados W^+ definido para variables aleatorias X_1, \dots, X_n i.i.d. como

$$W^+ = \sum_{i=1}^n \psi(X_i) R_i^+$$

donde $R = (R_1^+, \dots, R_n^+)$ es el vector de rangos de la muestra $|X_1|, \dots, |X_n|$. Como consecuencia de dicho intento se obtiene un resultado importante acerca de W^+ , específicamente se concluye que

$$\frac{\sqrt{(3n)(W^+ - EW^+)}}{\binom{n}{2}} \xrightarrow{(d)} N(0, 1), \text{ cuando } n \rightarrow \infty. \quad (1)$$

Por conveniencia en materia de notación llamamos

$$V := \frac{\sqrt{(3n)(W^+ - EW^+)}}{\binom{n}{2}}.$$

Aunque (1) esta plenamente demostrado en [1], en este artículo pretendemos ilustrar numérica y gráficamente dicha convergencia por medio del análisis de la diferencia (numérica) entre la función de distribución de V y la distribución normal estándar, el análisis consiste en tratar de hacer evidente que dicha diferencia tiende a cero a medida que n aumenta teniendo como noción de distancia

$$\max_{r \in \mathcal{R}} |F_v(r) - F_a(r)| \quad (2)$$

* Universidad de los Andes, Bogotá, Colombia.

donde \mathcal{R} es el conjunto de los posibles valores que toma el estadístico V , F_V es la distribución de V y F_a es la distribución normal estándar. En ese sentido, se establece que una distancia optima entre ambas distribuciones es cuando la expresión en (2) es inferior a 0.005, donde "optimo" hace referencia a que para todos los valores donde dicha distancia es optima la distribución de V se puede considerar idéntica a una normal estándar, esto es importante para efectos prácticos ya que la distribución de V computacionalmente podría no ser practica cuando deseamos manejar en numero grande de datos mientras que la distribución normal estándar esta ampliamente estudiada e implementada en la mayoría de software estadísticos y matemáticos siendo eficiente en el manejo de un numero grande de datos.

En aras de allanar el camino hacia nuestro propósito es importante establecer una expresión para \mathcal{R} y F_v , en primer lugar, como W_+ solo toma los valores

$$\left\{0, 1, \dots, \frac{n(n+1)}{2}\right\}$$

entonces se sigue

$$\begin{aligned} \mathcal{R} &= \left\{ \frac{\sqrt{(3n)(k-EW^+)}}{\binom{n}{2}} \mid k = 0, 1, \dots, \frac{n(n+1)}{2} \right\} \\ &= \left\{ \frac{\sqrt{(3n)(-EW^+)}}{\binom{n}{2}}, \frac{\sqrt{(3n)(1-EW^+)}}{\binom{n}{2}}, \dots, \frac{\sqrt{(3n)(\frac{n(n+1)}{2}-EW^+)}}{\binom{n}{2}} \right\}. \end{aligned}$$

Como $EW_+ = \frac{n(n+1)}{4}$ es una constante ya que no depende de k , entonces \mathcal{R} como se escribio en la ultima igualdad esta ordenado de menor a mayor.

Para el caso de F_v es importante el siguiente corolario demostrado en [1]

Corolario: Sea W^+ el estadístico de rangos signados para verificar $H_0 = \theta = \theta_0$. Para una muestra de tamaño n , la $H_0 : \theta = \theta_0$ distribución de W^+ es

$$f_{W^+}(k) = P(W^+ = k) = \begin{cases} \frac{c_n(k)}{2^n} & k = 0, 1, \dots, \frac{n(n+1)}{2} \\ 0 & \text{en otro caso} \end{cases}$$

(en [1] es el corolario 1.2.12) Donde $c_n(k)$ =numero de subconjuntos de enteros de $\{1, \dots, n\}$ para los cuales la suma es igual a k . Entonces para $t \leq 0$ se tiene

$$F_{W^+}(t) = P(W^+ \leq t) = \sum_{k \leq t} \frac{c_n(k)}{2^n} \quad (3)$$

Por lo tanto, para $r \in \mathcal{R}$ se concluye

$$\begin{aligned} F_v(r) &= P\left(\frac{\sqrt{(3n)(W^+-EW^+)}}{\binom{n}{2}} \leq r\right) \\ &= P\left(W^+ \leq \binom{n}{2} \frac{r}{\sqrt{3n}} + EW^+\right) \\ &= F_{W^+}\left(\binom{n}{2} \frac{r}{\sqrt{3n}} + EW^+\right). \end{aligned}$$

Como \mathcal{R} esta ordenado de menor a mayor denotamos

$$r_i = \frac{\sqrt{(3n)(i-EW^+)}}{\binom{n}{2}} \quad \text{para } i = 0, 1, \dots, \frac{n(n+1)}{2}. \quad (4)$$

Esto significa que $\mathcal{R} = \left\{ r_0, r_1, \dots, r_{\frac{n(n+1)}{2}} \right\}$, en ese sentido se sigue

$$\binom{n}{2} \frac{r_i}{\sqrt{3n}} + EW^+ = i.$$

Por lo tanto, se obtiene

$$F_v(r_i) = F_{W^+}(i) = \sum_{k \leq i} \frac{c_n(i)}{2^n} \quad \text{para } i = 0, 1, \dots, \frac{n(n+1)}{2}. \quad (5)$$

La importancia de esta ultima expresión radica en que permite escribir (2) de la siguiente forma

$$\begin{aligned} \max_{r \in \mathcal{R}} |F_v(r) - F_a(r)| &= \max_{r=0, \dots, \frac{n(n+1)}{2}} \left| F_{W^+}(i) - F_a \left(\frac{\sqrt{(3n)(i - EW^+)}}{\binom{n}{2}} \right) \right| \\ &= \max_{r=0, \dots, \frac{n(n+1)}{2}} \left| \frac{c_n(i)}{2^n} - F_a \left(\frac{\sqrt{(3n)(i - EW^+)}}{\binom{n}{2}} \right) \right| \end{aligned} \quad (6)$$

Esta ultima sera útil para efectos de programación que es tema de la siguiente sección, en donde entraremos en los detalles de la implementación de las simulaciones realizadas para diferentes valores de n , en donde se intenta explicar las líneas de código que permiten las simulaciones en el programa estadístico R para luego en una sección posterior interpretar los resultados.

1.2 Problema 2: Eficiencia de estadísticos para pruebas sobre el centro de simetría

Los estadísticos son usados para determinar valores de parámetros asociados a una muestra, en algunos casos para inferir aspectos de la muestra como la simetría, en este problema consideramos datos Z_1, \dots, Z_n i.i.id. de una distribución continua simétrica en θ , nos enfocamos en tres estadísticos que son los siguientes:

- t student, dado por

$$t = \frac{\sqrt{n}(\bar{Z} - \theta_0)}{s}$$

siendo s la desviación estándar.

- El estadístico de signos $B = \sum_{i=1}^n \Psi(Z_i - \theta_0)$.

- El estadístico de rangos signados, $W^+ = \sum_{i=1}^n \Psi(Z_i - \theta_0) R_i^+$, donde los R_i^+ son los rangos absolutos de los $Z_i - \theta$.

El objetivo es analizar la eficiencia de estos estadísticos mediante su potencia, es decir, la siguiente probabilidad

$$\text{Potencia} = P(\text{rechazar } H_0 | H_a \text{ es verdad}).$$

Para dicho análisis, tomamos sin pérdida de generalidad, $\theta_0 = \pi$ y para la alternativa $\theta_a = \pi + 0.1, \pi + 0.25, \pi + 0.5, \pi + 1$. Como tamaños muestrales tomamos $n = 20, 30, 50$ y 100 . Consideramos las siguientes distribuciones:

- $N(\theta, 1)$.
- $\text{Unif}(\theta - 1, \theta + 1)$.

► Distribución de Laplace, con densidad

$$f(x) = \frac{1}{2} \exp(-|x - \theta|) \quad \text{para } x \in \mathbb{R}.$$

Trataremos de estimar la eficiencia de cada estadístico en cada una de las distribuciones mencionadas, note que para el caso del estadístico t -estudent dicho calculo podría hacerse explícito, en diversos textos lo hacen, pero para el caso de los estadísticos restantes la situación nos es tan optimista, de modo que se hace relevante estimar su eficiencia y descubrir bajo que condiciones esa eficiencia mejora, precisamente, en esa dirección descubriremos en secciones posteriores que la eficiencia depende del tamaño de la muestra y del valor de θ_a .

La estrategia para descubrir esas dependencias se basa en los siguientes pasos. Para cada distribución simétrica, tamaño muestral, valor de la alternativa y estadístico hacemos los siguiente:

1. Generamos una muestra de la respectiva distribución simétrica, con centro en θ_a .
2. Aplicamos el estadístico para evaluar la hipótesis nula H_a : el centro es $\theta = \pi$, contra la alternativa $H_a : \theta > \pi$, y decidimos si el valor es significativo a nivel de 5%.
3. Repetimos los pasos anteriores un total de 500 veces, almacenando la potencia como el porcentaje de veces que se rechaza H_0 .

El ultimo paso nos da un buen estimativo de la potencia, es natural pensar en la potencia como se interpreta en el paso 3, y que se asocie con la confianza en el estadístico, entre mas rechazo mas confianza podríamos tener en el estadístico ya que esto evidencia que no esta en contra de la realidad dado que conocemos de antemano que la distribución no es simétrica en θ_0 . En la sección 2 desarrollaremos la forma de llevar esta intención a cavo, y en la sección 3 presentaremos e interpretaremos los resultados de las simulaciones realizadas.,

2 Implementación de las simulaciones realizadas

En esta sección abordaremos los detalles de la implementación de las simulaciones realizadas, en este caso tales simulaciones fueron realizadas en el software de análisis estadístico R, concretamente describiremos parte del código desarrollado para cada uno de los problemas mencionados anteriormente y que ocupan el tema central de este artículo.

2.1 Implementación: Aproximación normal a la distribución W^+

Como se menciona en la introducción, el objetivo es determinar el tamaño de la muestra, en este caso se denota por n , tal que

$$\max_{r \in \mathcal{R}} |F_v(r) - F_a(r)| < 0.005.$$

En ese sentido, el primer paso es crear el conjunto \mathcal{R} , para tal fin se crea la función `r_posibles`, esta presenta \mathcal{R} como un vector de tamaño $\frac{n(n+1)}{2} + 1$ cuya componente i -ésima es r_i donde este ultimo fue definido en (4), la rutina de esta función es la siguiente

```
r_posibles <- function(n){
  Wmas <- 0:(n*(n+1)/2)
  r <- sqrt(3*n)*(Wmas - n*(n+1)/4)/choose(n,2)
}
```

El siguiente paso es crear la distribución F_v , en nuestro programa la llamamos distribución verdadera, es decir, `F_verdadera`, la razón de dicho nombre radica en que debido a (1) en la practica es común asumir que V tiene distribución normal estándar hasta el punto de que se llega a asumir que esa es la distribución de V olvidando que esta tiene una distribución que en principio no es normal

estándar pero que si es parecida para n grande, que es lo que queremos ilustrar. En ese sentido, **F_verdadera** evaluada n es un vector de tamaño $\frac{n(n+1)}{2} + 1$ cuya componente i -ésima es $F_V(r_i)$, recordemos que por (5) este termino se escribe en términos de $c_n(k)$, la función c_n fue previamente programada por el profesor Adolfo Quiroz quien la llamo **cnk** la cual para cada n devuelve un vector de tamaño $\frac{n(n+1)}{2} + 1$ cuya componente i -ésima es $c_n(i)$. En ese orden de ideas, la rutina que crea **F_verdadera** es

```
F_verdadera <- function(n){
  N <- n*(n+1)/2 +1
  m <- matrix(1,N,N)
  m[upper.tri(m)] <- 0
  Fv <- 1/2^n*(m %*% t(t(cnk(n))))
}
```

Considerando $N = \frac{n(n+1)}{2} + 1$, gracias a (5) la idea de la rutina anterior se fundamenta en el hecho de que

$$\begin{pmatrix} F_V(r_0) \\ F_V(r_1) \\ F_V(r_2) \\ \vdots \\ F_V(r_{N-1}) \\ F_V(r_N) \end{pmatrix} = \frac{1}{2^n} \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & 1 & \cdots & 1 & 0 \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{pmatrix} \begin{pmatrix} c_n(0) \\ c_n(1) \\ c_n(2) \\ \vdots \\ c_n(N-1) \\ c_n(N) \end{pmatrix}.$$

Esto nos permite evitar usar lineas de iteración como **for**, en su lugar creamos una matriz triangular inferior de unos a partir de una matriz de unos que denotamos como **m**.

Por ultimo, antes de determinar el valor $\max_{r \in \mathcal{R}} |F_v(r) - F_a(r)|$ creamos un vector con los valores que toma F_a evaluada en \mathcal{R} , es decir, $F_a(\mathcal{R})$, dicho vector es

```
pnorm(r_posibles(n))
```

En ese sentido, como $\max_{r \in \mathcal{R}} |F_v(r) - F_a(r)|$ depende de n , creamos la función **maximo** que representa dicho valor y que depende de n la cual tiene la siguiente rutina

```
maximo <- function(n){
  max <- max(abs(pnorm(r_posibles(n))-F_verdadera(n)))
}
```

Crear esta ultima función era el objetivo de la implementación de la simulación, simular es darle valores a n y evaluarlos en la función **maximo**, los resultados e interpretación los discutiremos en mayor detalle en la siguiente sección.

2.2 Implementación: Eficiencia de estadísticos para pruebas sobre el centro de simetría

El enfoque de la implementación del código en este punto se centró primeramente en el análisis del estadístico *t*-student, y posteriormente, hacer un análisis comparativo de los tres estadísticos para distribuciones puntuales y valores de n (el número de elementos de la muestra aleatoria) específicos.

Así que para responder a la pregunta sobre la eficiencia del estadístico *t*-student diseñamos un código que nos permitiera contrastar para distribuciones distintas, cuál era la potencia del estadístico según el número de elementos que tomáramos como muestra y la hipótesis alternativa a considerar.

```
t_student <- function(theta_a){
  potencias <- matrix(0,5,4)
  for(n in c(20,30,50,100)){
```

```

#a es un vector contador de las veces que se rechaza H0
a <- c(0,0,0,0,0)
for(i in 1:500){
  #Generamos 5 muestras aleatorias con distribucion indicada 500 veces
  z1 <- rnorm(n,theta_a,1)
  z2 <- runif(n,theta_a-1,theta_a+1)
  z3 <- rlaplace(n,theta_a,1)
  z4 <- rt(n,df = 2,ncp = theta_a)
  z5 <- rcauchy(n,theta_a,1)
  z <- rbind(z1,z2,z3,z4,z5)
  t <- c(0,0,0,0,0)
  p_valor <- rep(0,5)
  #Calculamos el p-valor para cada muestra aleatoria y lo guardamos en un
  vector
  for(j in 1:5) {
    #t[j] calcula el valor observado por el estadistico.
    t[j] <- sqrt(n)*(mean(z[j,])-pi)/sd(z[j,])
    p_valor[j] <- 1-pt(t[j],n-1)
    if (p_valor[j] < 0.05) a[j] <- a[j]+1
  }
}
#Guardamos la potencia del estadistico para cada distribucion
for (i in 1:5){potencias[i,which(c(20,30,50,100)==n)] = a[i]/500}
}
#Le damos nombres a las filas y columnas de la matriz potencias y la
imprimimos.
rownames(potencias) <- c("Normal","Uniforme","Laplace","t-student","Cauchy"
)
colnames(potencias) <- c("n=20","n=30","n=50","n=100")
potencias
}

```

Esta función toma como parámetro un valor posible de θ_a y regresa una matriz con las potencias obtenidas para muestras aleatorias de tamaño $n = 20, 30, 50$ y 100 respectivamente provenientes de cinco distribuciones distintas. Las filas corresponden a las distribuciones y las columnas al tamaño de la muestra.

Posteriormente, con el fin de comparar los tres estadísticos en cuestión, creamos de nuevo una matriz que para una distribución y valores de n fijos nos muestra la potencia de los estadísticos dependiendo de los posibles valores para la hipótesis alternativa. Damos como ejemplo la distribución normal, los demás son análogos.

```

Distribucion_normal <- function(n){
  potencias <- matrix(0,3,4)
  #A continuacion se calcula para valores posibles de theta_a la potencia de
  cada
  estadistico.
  for(theta_a in c(pi+0.1,pi+0.25,pi+0.5,pi+1)){
    # a sera un vector que contendra el numero de veces que se rechaza la
    hipotesis nula para cada estadistico
    a <- c(0,0,0)
    for(i in 1:500){
      z <- rnorm(n,theta_a,1)
      #Calculamos el estadistico t
      t <- sqrt(n)*(mean(z)-pi)/sd(z)

```

```

p_valor1 <- 1-pt(t,n-1)
#- Si el p-valor es menor que 0.05 rechazamos la hipotesis nula
if (p_valor1 < 0.05) a[1] <- a[1]+1

#- Calcula el p-valor para el estadístico de Wilcoxon
x <- wilcox.test(z,mu=pi, alternative = "greater")
p_valor3 <- x[["p.value"]]
#- Si el p.valor es menor que 0.05 rechazamos la hipotesis nula
if (p_valor3 < 0.05) a[3] <- a[3]+1

#Calcula el p-valor para el estadístico de signos
y <- sum(z-pi>0)
p_valor2 <- 1-pbinom(y-1,n,0.5)
if (p_valor2 < 0.05) a[2] <- a[2]+1
}
#Insertamos el numero de veces que rechazamos la hipotesis nula segun el
estadístico considerado
for (i in 1:3){potencias[i,which(c(pi+0.1,pi+0.25,pi+0.5,pi+1)==theta_a)]
= a[i]/500}
}
#ponemos nombres a las filas y columnas de la matriz e imprimimos sus
valores
rownames(potencias) <- c("t-student", "Signos", "Wilcoxon")
colnames(potencias) <- c("pi+0.1", "pi+0.25", "pi+0.5", "pi+1")
potencias

```

Teniendo ya estos dos tipos de contrastes quisimos ver aún mejor los comportamientos a considerar. Por ello hicimos varias gráficas que nos mostraran justamente lo que recién generamos.

3 Interpretación de los resultados

En esta sección presentaremos e interpretaremos los resultados obtenidos luego de realizar varias simulaciones con los métodos y rutinas implementadas.

3.1 Resultados: Aproximación normal a la distribución W^+

Los datos que presentamos a continuación son el resultado de aplicar la función `maximo` para varios valores de n , los valores que emplearemos son $n = 10, 20, 30, 50, 80, 100, 150$, para el valor $n = 120$ el calculo excede la capacidad de un computador común, recordemos que `maximo` representa el valor

$$\max_{r \in \mathcal{R}} |F_v(r) - F_a(r)|.$$

En la Figura 3.1 se ilustra el valor de la expresión anterior para diferentes valores de n .

Esta ultima gráfica ilustra una clara tendencia de como el valor de (2) se aproxima a 0 a medida que n crece, dicha apreciación queda corroborada al encontrar que dicha máxima diferencia es inferior a 0.005 para $n = 108$, en efecto, tal comprobación es hecha en el programa estadístico R mediante la siguiente rutina

```

naux<-5
while(maximo(naux)>0.005){
naux<- naux+1
}

```

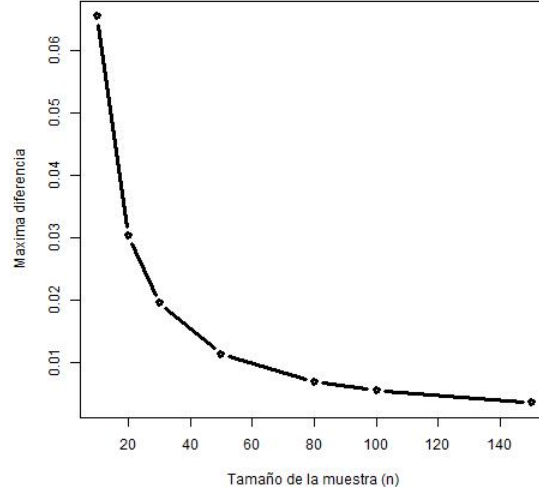


Fig. 3.1. $\max_{r \in \mathcal{R}} |F_v(r) - F_a(r)|$ para diferentes valores de n .

En la rutina anterior **naux** es un numero auxiliar que cuenta el numero de veces que el valor de (2) supera 0.005, dicho contador empieza desde 5 debido a que la función **cnk** tiene inconvenientes para n pequeños.

Este resultado, es decir, $n = 108$ significa que a partir de dicha cantidad de datos podemos asumir que el estadístico V tiene distribución normal estándar.

Estas observaciones son reafirmadas con las gráficas prob.prob de la Figura 3.2.

En resumidas palabras estas graficas representan puntos de la forma $(F_v(r), F_a(r))$ para todos los $r \in \mathcal{R}$, las figuras anteriores ilustran que para $n = 100$ y en adelante tales puntos pareciera que se ponen sobre la recta $y = x$, es decir, la identidad, esto reafirma la convergencia que queríamos ilustrar ya que dicho comportamiento solo se puede dar si ambas distribuciones se hacen cada vez mas indistinguibles, idénticas.

3.2 Resultados: Eficiencia de estadísticos para pruebas sobre el centro de simetría

Antes de introducir los resultados obtenidos es importantes establecer unas convenciones por medio de las cuales se podrán entender los graficos que es la forma como se presentan los resultados en el presente documento. En la Figura 3.3 se presenta los resultados obtenidos para la muestra t -student, de la misma forma se presentan los resultados para los demás estadísticos, esta grafica muestra en su eje horizontal los valores que toma θ_a y en el eje vertical los valores de la potencia estimada para dichos valores de θ_a , los colores de las lineas manejan la siguiente convención:

- **Línea negra** = Resultado de aplicar el estadístico (t , B o W^+) a una muestra de n datos con distribución normal $N(\theta, 1)$.
- **Línea azul** = Resultado de aplicar el estadístico (t , B o W^+) a una muestra de n datos con distribución uniforme $\text{Unif}(\theta - 1, \theta + 1)$.
- **Línea verde** = Resultado de aplicar el estadístico (t , B o W^+) a una muestra de n datos con distribución doble exponencial o de Laplace con parámetros de localización y de escala π y 0, respectivamente.

$$f(x; \theta, 1) = \frac{1}{2} \exp(-|x - \theta|)$$

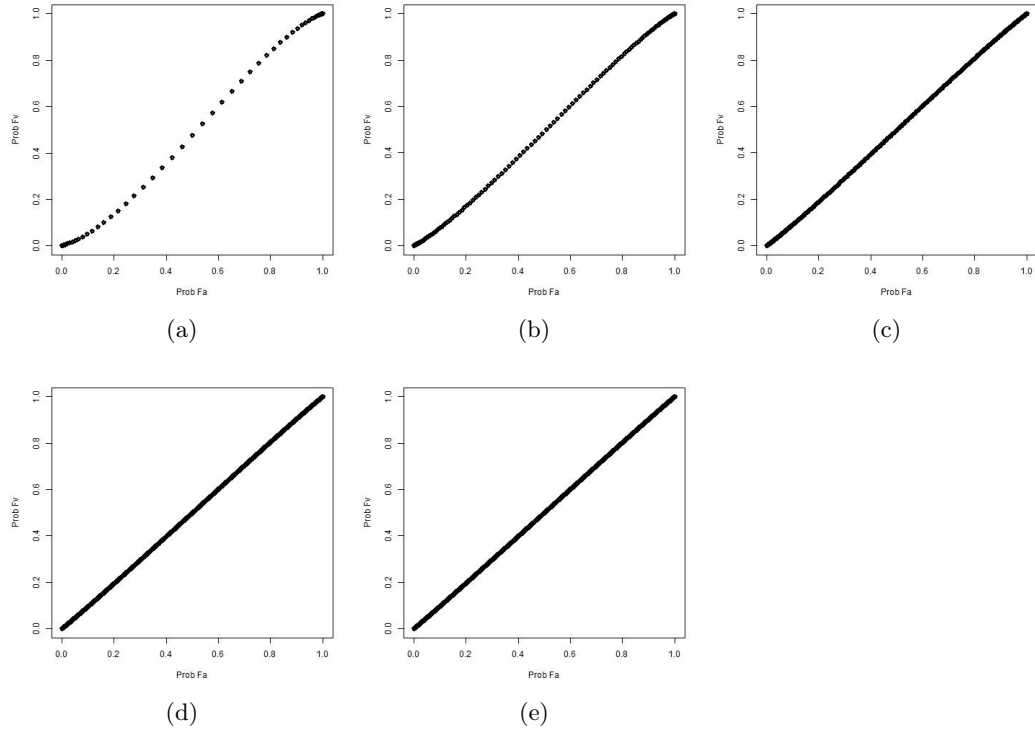


Fig. 3.2. Grafica prob.prob. (a) Para $n = 10$. (b) Para $n = 20$. , (c) Para $n = 50$, (d) Para $n = 100$, (e) Para $n = 150$.

- **Línea roja** = Resultado de aplicar el estadístico (t , B o W^+) a una muestra de n datos con distribución $F(x - \theta)$ donde F es la distribución t de student con dos grados de libertad y centro θ .
- **Línea magenta** = Resultado de aplicar el estadístico (t , B o W^+) a una muestra de n datos con distribución de Cauchy

$$f(x; \theta, 1) = \frac{1}{\pi} \left[\frac{1}{(x - \theta)^2 + 1} \right].$$

Donde de nuevo θ es el parámetro de localización y el parámetro de escala es 1.

Esta forma de presentar los resultados nos permite hacer visible las tendencias y dependencias que tiene el parámetro θ_a y la potencia del estadístico.

Ahora, entrando en materia de resultados es conveniente analizar y presentar los resultados estadístico por estadístico.

t -student. Debemos reconocer que el primer interés que nos atañe es estudiar principalmente el estadístico t -student y para cuáles distribuciones es más potente (en caso que lo sea), quizá ese interés se deba a su amplio uso y divulgación. En ese sentido, para distintos tamaños muestrales se obtuvo un comportamiento distinto, para analizarlo nos basamos en la gráficas de la Figura 3.3(ver al final del documento) siguiendo las convenciones expuestas anteriormente. Para efectos prácticos consideramos los valores representativos: $n = 20, 30, 50$ y $n = 100$.

Estas gráficas muestran que entre mas alejado θ_a de θ_0 mayor es la potencia del estadístico exceptuando el caso de la distribución de Cauchy, donde pareciera que la potencia del estadístico se estancara en un valor. Respecto al tamaño de las muestras se observa que a medida que es mas grande n la potencia alcanza valores suficientemente grandes para valores de θ_a no tan alejados de θ_0 , de nuevo, exceptuando la distribución de Cauchy que no sigue esta tendencia. A la pregunta ¿Para cuales distribuciones es mas efectivo el t de Student? la respuesta que podemos dar, de acuerdo a la gráficas, depende de lo que se considera más efectivo, lo ideal es que el estadístico requiera un tamaño de la muestra pequeño y que rechace H_0 para θ_0 a una distancia muy pequeña de θ_a , en ese sentido, el t de Student es mas eficiente precisamente para la distribución t de Student, seguido de cerca por la distribución uniforme.

Table 1. Estadístico t-student para $\theta = \pi + 0.1$

$\pi + 0.1$	$n = 20$	$n = 30$	$n = 50$	$n = 100$
Normal	0.122	0.124	0.162	0.262
Uniforme	0.198	0.232	0.326	0.530
Laplace	0.088	0.112	.126	0.174
t-student	0.682	0.880	0.970	0.988
Cauchy	.050	0.054	0.048	0.058

Table 2. Estadístico t-student para $\theta = \pi + 1$

$\pi + 1$	$n = 20$	$n = 30$	$n = 50$	$n = 100$
Normal	0.998	1.000	1.000	1.000
Uniforme	1.000	1.000	1.000	1.000
Laplace	0.908	0.972	1.000	1.000
t-student	0.950	0.968	0.990	0.990
Cauchy	.290	0.274	0.264	0.286

Estadístico de signos B . Uno de los objetivos es compara la eficiencia de los estadísticos no parametricos respecto a la eficiencia del estadístico t-student, en ese orden de ideas, debemos interpretar los resultados del estadístico B para así poder resaltar esas diferencias (si las hay) y para cuáles distribuciones es más potente (en caso que lo sea). En ese sentido, para distintos tamaños muestrales se obtuvo un comportamiento distinto, para analizarlo nos basamos en la gráficas de la Figura 3.4(ver al final del documento) siguiendo las convenciones expuestas anteriormente. Para efectos prácticos consideramos los valores representativos: $n = 20, 50$ y $n = 100$.

Estas gráficas muestran que entre mas alejado θ_a de θ_0 mayor es la potencia del estadístico, note que en este caso ya la distribución de Cauchy sigue la tendencia, lo que no ocurría con el estadístico t -student, esto se puede considerar como un punto a favor de B respecto a t . Respecto al tamaño de las muestras se observa que a medida que es mas grande n la potencia alcanza valores suficientemente grandes para valores de θ_a no tan alejados de θ_0 , pareciera que esto es común en los estadísticos, no obstante, el estadístico B para $n = 100$ muestra una mayor concentración en potencias altas para $\theta_a = \pi + 0.25$ para $n = 100$ que la el estadístico t , aunque en el caso de B la potencia respecto a la distribución de Cauchy se queda un poco rezagada, en ese punto, es mas eficiente B .

Estadístico de rangos signados W^+ . El otro estadístico importante no paramétrico es el de rangos signados W^+ , de nuevo nos preguntamos por su eficiencia respecto al estadístico t -student y adicionalmente respecto a B , en ese orden de ideas, debemos interpretar los resultados del estadístico W^+ para así poder resaltar esas diferencias (si las hay) y para cuáles distribuciones es más potente (en caso que lo sea). En ese sentido, como ocurrió con los anteriores estadísticos, para distintos tamaños muestrales se obtuvo un comportamiento distinto, para analizarlo nos basamos en la gráficas de la Figura 3.5(ver al final del documento) siguiendo las convenciones expuestas anteriormente. Para efectos prácticos consideramos los representativos: $n = 20, 50$ y $n = 100$.

Al igual que en los anteriores estadísticos, estas gráficas muestran que entre mas alejado θ_a de θ_0 mayor es la potencia del estadístico, al igual que con B la distribución de Cauchy sigue la tendencia, lo que no ocurría con el estadístico t -student, de nuevo esto se puede considerar como un punto a favor de W^+ respecto a t , no obstante se observa que B tiene una mayor potencia para $n = 100$ que W_+ en el caso de la distribución de Cauchy, sin embargo, en ambos casos la potencia de ambos estadísticos en dicha distribución no va al mismo ritmo que la potencia para las otras distribuciones. Ahora, para $n = 50$ la grafica denota que W^+ es el estadístico de mejor desempeño. Respecto al tamaño de las muestras se observa que a medida que es mas grande n la potencia alcanza valores suficientemente grandes para valores de θ_a no tan alejados de θ_0 , no obstante, la potencia del estadístico B para $n = 50$ y $\theta_a = \pi + 0.25$ para todas las distribuciones es mayor que las respectivas potencias W_+ , más aun, el estadístico t también evidencia una ligera ventaja exceptuando el caso de la distribución de Cauchy, esto muestra un desventaja respecto a los estadísticos t y B ya que esto significa que t y B necesita muestras de menor tamaño para llegar a una potencia significativa que por lo menos supera 0.5 que el tamaño de la muestra que requiere W^+ para llegar a tales niveles de potencia.

Comparaciones entre los estadísticos.

Para reafirmar las observaciones hechas antes y sustentar cualquier conclusión acerca de la eficacia de los estadísticos y su inevitable comparación, pensamos que una manera tal vez más clara de contrastar la potencia de los estadísticos es aislarlos según la distribución de los datos que usamos para ponerlos a prueba.

Consideremos las siguientes tablas para $n = 50$ y las distribuciones normal, exponencial de Laplace y la distribución t con centro en θ y dos grados de libertad.

Table 3. Comparación de estadísticos para la distribución normal

	$\pi + 0.1$	$\pi + 0.25$	$\pi + 0.5$	$\pi + 1$
t-student	0.162	0.524	0.956	1
Signos	0.096	0.322	0.782	1
Wilcoxon	0.164	0.514	0.942	1

Table 4. Comparación de estadísticos para la distribución exponencial de Laplace

	$\pi + 0.1$	$\pi + 0.25$	$\pi + 0.5$	$\pi + 1$
t-student	0.134	0.358	0.792	1.000
Signos	0.128	0.388	0.832	0.998
Wilcoxon	0.148	0.434	0.876	1.000

Table 5. Comparación de estadísticos para la distribución t-student con dos grados de libertad centrada en θ

	$\pi + 0.1$	$\pi + 0.25$	$\pi + 0.5$	$\pi + 1$
t-student	0.972	0.974	0.992	0.988
Signos	0.510	0.624	0.876	0.996
Wilcoxon	0.922	0.970	0.998	1.000

Table 6. Comparación de estadísticos para la distribución de Cauchy

	$\pi + 0.1$	$\pi + 0.25$	$\pi + 0.5$	$\pi + 1$
t-student	0.040	0.102	0.134	0.276
Signos	0.078	0.252	0.682	0.960
Wilcoxon	0.108	0.256	0.604	0.926

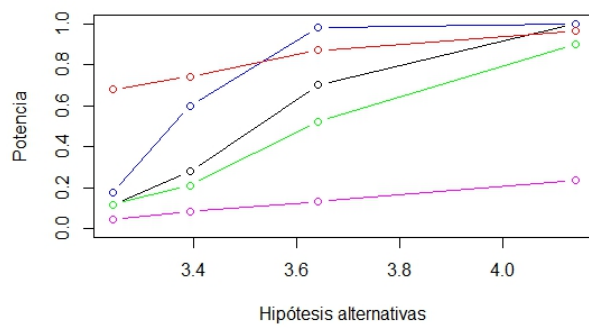
4 Conclusiones

De acuerdo a las simulaciones realizadas referente al primer problema se puede inferir que para muestras de datos suficientemente grandes, superiores a 108, la distribución de V es casi idéntica a la normal estándar, lo que permite asumir que toda población que se rija por la distribución de V es normalmente distribuida con media 0 y varianza 1.

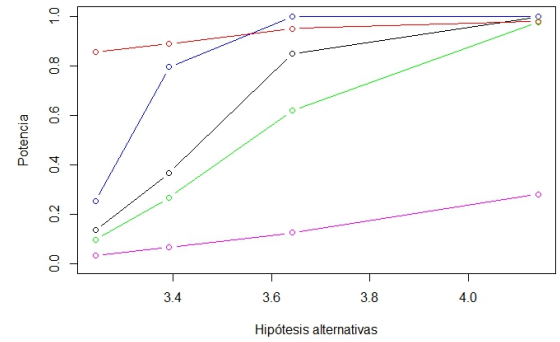
Por otro lado, para el segundo problema, la decisión de decantarse por un estadístico involucra varios factores, algunos son subjetivos pues dependen de las preferencias del investigador y sus necesidades, básicamente la controversia se concentra en la tensión entre tamaño de la muestra y la región de rechazo, es decir, entre n y el valor de θ_a , decidirse por un estadístico para el cual un valor alto de eficiencia en términos de potencia dependa del tamaño de la muestra implica costos de muestreo importante ya que con frecuencia la recolección de muestras es complejo mas cuando se requieren de un gran numero, por otro lado, si se decide que dicha dependencia sea mas ligada al tamaño de θ_a , esto implica que la región de rechazo de H_0 asumiendo que H_a es verdad es mas grande, es decir, para que se rechace H_0 se debe tener que θ_0 debe estar suficientemente lejos de θ_a , que puede ser contraproducente ya que se podrían aceptar θ_0 que contradicen la realidad de la distribución de la población. En ese sentido vemos pertinente determinar que una buena elección se debe fundamentar en un equilibrio entre el tamaño de la muestra y el tamaño de θ_a . En ese orden de ideas, bajo ese criterio, ¿como es el desempeño de los estadísticos no parametricos? entendiendo que una buena aproximación a la situación ideal es en donde un estadístico no requiere una muestra de tamaño grande ni un θ_a tan alejado de θ_0 para determinar cuando rechazar H_0 , nuestra posición es fundamentada en los datos expuestos en las gráficas de cada estadístico ilustradas en la sección anterior, aunque hay que recalcar que el estadístico W^+ demuestra una mayor eficiencia que B , particularmente, esto se evidencia en el caso de $n = 50$ y $\theta_a = \pi + 0.5$ que sería una situación aceptable ya que n no es tan grande y θ_a no estaría tan alejado de θ_0 , esto ultimo responde a la pregunta ¿cual de estos estadísticos es preferible?, es decir, todo conduce a W^+ . No obstante, existe una carga de subjetividad en este asunto, por ejemplo si para un investigado el tamaño de la muestra es una cuestión que signifique un obstáculo en su investigación y la población a la cual extrae la muestra se distribuye t -Student, entonces la Tablas 5 demuestra que en dicho contexto el test t es la mejor opción.

Referencias

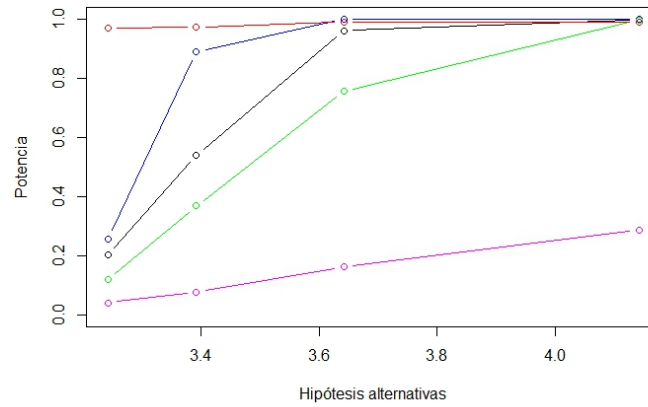
1. R. Randles & D. Wolfe, *Introduction to Theory of Nonparametric Statistic*, Krieger Publishing Company, 3-114 (1979)



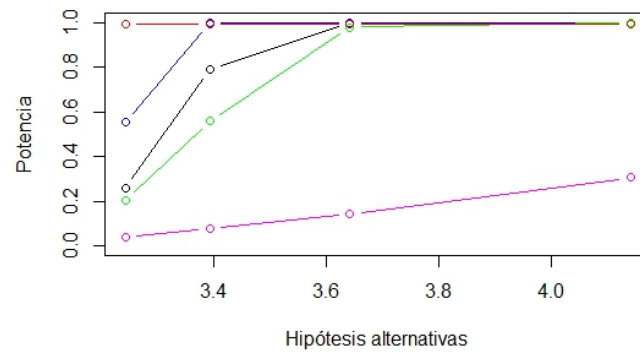
(a)



(b)



(c)



(d)

Fig. 3.3. Graficas de potencia para el estadístico t -student, el eje horizontal es valor de la hipotesis alterna y el vertical la potencia. (a) Para $n = 20$. (b) Para $n = 30$. , (c) Para $n = 50$, (d) Para $n = 100$.

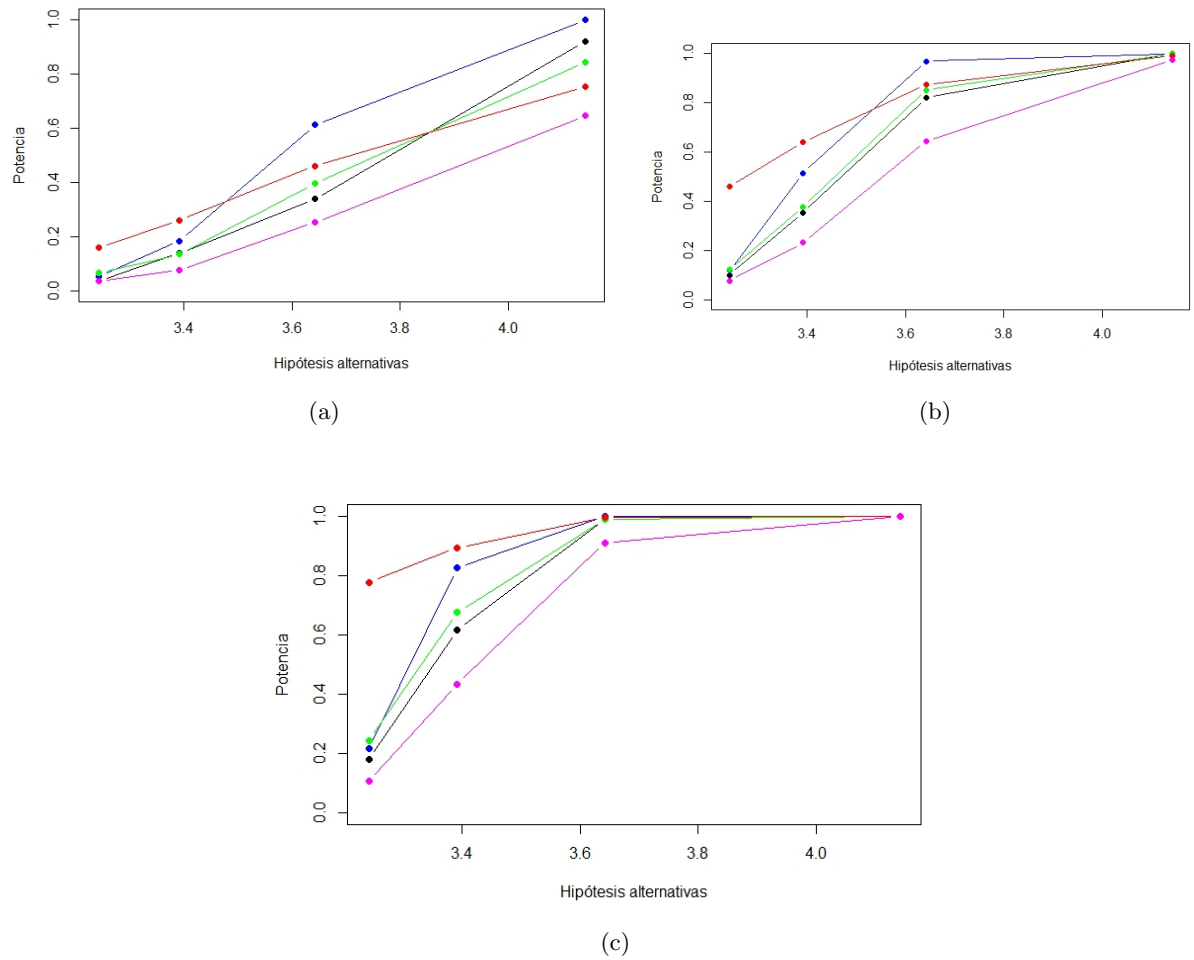
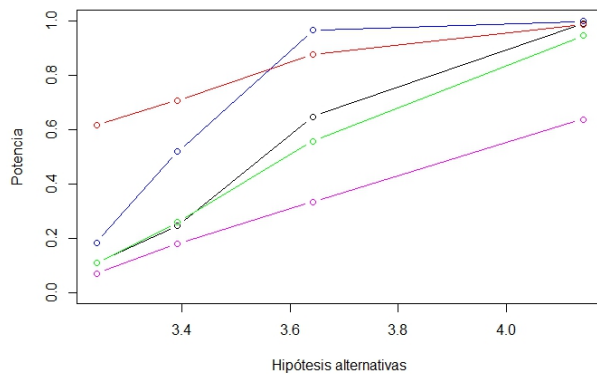
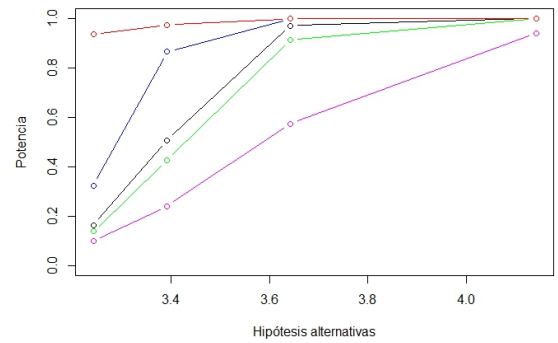


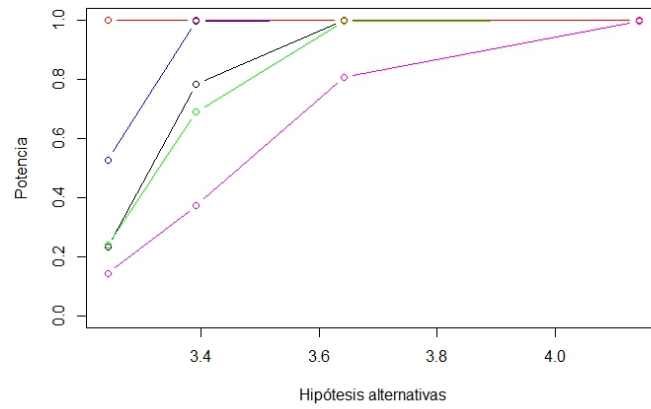
Fig. 3.4. Graficas de potencia para el estadístico B , el eje horizontal es valor de la hipotesis alternativa y el vertical la potencia. (a) Para $n = 20$. (b) Para $n = 50$. , (c) Para $n = 100$.



(a)



(b)



(c)

Fig. 3.5. Graficas de potencia para el estadístico W^+ , el eje horizontal es valor de la hipótesis alternativa y el vertical la potencia. (a) Para $n = 20$. (b) Para $n = 50$. , (c) Para $n = 100$.