Trabajo Práctico de Introducción a la Estadística y Ciencia de Datos

Bruno Muschietti, Augusto Guarnaccio y Victoria Aguirre

2do cuatrimestre 2024

Índice

1 Resumen

Este informe documenta todo el trabajo realizado para construcción de nuestro propio test de Wilcoxon en R. Por eso a lo largo del mismo podrán encontrarse respuestas a distintas cuestiones de programación como también demostraciones de resultados teóricos necesarios para la final elaboración del mismo. No incluiremos código a menos que sea necesario puesto que este va a encontrarse en un archivo de extensión .R junto con el mismo informe.

2 Introducción

El objetivo de este trabajo es no solo entender una nueva herramienta para testear hipótesis acerca de los datos como lo es el Test de los rangos signados o Test de Wilcoxon sino también lograr construir objetos en R para comprender su funcionamiento y ampliar las herramientas con las que contamos a la hora de trabajar con datos.

El estudio del test mencionado previamente es de nuestro interés por dos motivos. El primero es por pertenecer a la familia de los llamados tests no paramétricos, cruciales cuando no podemos afirmar que el Data Generating Process viene dado por una distribución particular que depende de ciertos parámetros sino que viene de una distribución desconocida pero con ciertas características.

El segundo motivo de interés recae en que es la herramienta que implementaremos como un objeto en R puesto que las cuentas que debemos realizar alrededor del estadístico se facilitan si nos valemos de un enfoque computacional. Por ejemplo, para el computo de la distribución del estadístico aparece la necesidad de realizar cálculos recursivos (el por qué de esto se verá más adelante). Cálculos que son sencillos de resolver computacionalmente pero complejos si los hacemos a mano. Además, lo interesante de poder entender como se implementa como un objeto es que nos brinda la posibilidad de desarrollar nuestros propios objetos en contextos donde lo brindado por R no sea lo que se ajuste a nuestras pretenciones. Por lo tanto, es un test que muestra la fuerte relación entre la estadística y la computación; y como esta última complementa a la primera.

3 Resultados

A lo largo de toda esta sección incluiremos las respuestas a las preguntas que pueden encontrarse diseminadas alrededor del trabajo práctico como también algunos resultados teóricos (junto con sus demostraciones) cruciales para construir nuestro test y comprenderlo mejor.

3.1 OOP en R

3.1.1 Entendiendo las clases de objetos

Antes de implementar el test como el objeto correspondiente en R, un objeto de clase `htest`, tenemos que entender cómo R maneja los objetos, las clases y los métodos. Para eso veamos el siguiente ejemplo en R.

¿Qué clase tienen los siguientes vectores: `c(T, F)`, `c(T, F, 1)` y `c(T, F, 1, "1")`? ¿Qué cree que está sucediendo?

```{r}
class(c(T,F))
class(c(T,F,1))
class(c(T,F,1,"1"))
```

```
[1] "logical"
[1] "numeric"
[1] "character"
```

En R ocurre que los distintos tipos de datos básicos tienen una jerarquía entre sí que va desde el menos general, que es el `logical`, hasta el más general, que es el `character`. Por lo tanto, cuando lee el vector le asigna la clase del elemento más general que esté conteniendo y considera que todos los otros valores comparten ese tipo. Por lo tanto, a la hora de construir un objeto es muy importante que si queremos definir distintos valores dentro del mismo y queremos que tengan una misma clase se las asignemos nosotros mismos pues ya vemos como maneja R esta asignación.

Ahora, ¿cómo puede R manejar todas estas clases y darnos objetos de distintas clases? La respuesta recae en las funciones que tiene implementadas. Estas lo que hacen es tomar objetos de ciertas clases y devolvernos un objeto de la clase deseada. Como bien se menciona en el enunciado en R la mayoría de estas funciones son métodos *genéricos* pero hay otras que no como por ejemplo `density`.

¿Qué clase tiene `density`? ¿Y `density(1:500)`? ¿Dónde está la diferencia?

```r
class(density)
class(density(1:500))
```

```
[1] "function"
[1] "density"
```

`Density` tiene clase `function` mientras que `density(1:500)` tiene clase `density`. La diferencia está en que `density` es una función que recibe ciertos parámetros de entrada y crea un objeto de clase `density` y que `density(1:500)` es justamente el objeto que crea la función. Como es un objeto pasa a tener atributos y métodos que son particulares a la clase `density`.

Con esta diferencia en mente cabe preguntarnos qué cosas nos permite hacer `density` con el objeto que crea y comparar qué es lo que ocurre con otra función *genérica* para ver si la diferencia es tan marcada.

¿A cuántas clases sabe despachar el genérico `print`? ¿Con cuántos métodos cuenta `density`, además de `plot`?

Primero fijemonos en la cantidad de clases a las que puede despachar `print`.

```r
sloop::s3_methods_generic("print")
```

A tibble: 263 × 4

generic <chr>	class <chr>	visible <lgl>
print	acf	FALSE
print	activeConcordance	FALSE
print	AES	FALSE
print	anova	FALSE
print	aov	FALSE
print	aovlist	FALSE
print	ar	FALSE
print	Arima	FALSE
print	arima0	FALSE
print	AsIs	TRUE

1-10 of 263 rows | 1-3 of 4 columns Previous 1 2 3 4 5 6 ... 27 Next

El genérico `print` sabe despachar a 263 clases en s3. Si miramos con

atención el data frame sabemos que tiene 263 filas y 4 columnas, una de ellas bajo el nombre `class`. Por lo tanto, utilizando la información brindada por la dimensión sabemos que en s3 sabe despachar a 263 métodos.

Veamos qué ocurre con `density`.

```{r}
sloop::s3_methods_class("density")
```

A tibble: 2 × 4

generic <chr>	class <chr>	visible <lgl>	source <chr>
plot	density	FALSE	registered S3method
print	density	FALSE	registered S3method

2 rows

Vemos que además de `plot` cuenta con `print`. No obstante, sabemos que `density` tiene más métodos ¿Qué fue lo que pasó? ¿Serán de S4?

```{r}
methods(class="density")
```

```
[1] coerce      initialize  plot        print       show        slotsFromS3
see '?methods' for accessing help and source code
```

```{r}
sloop::s4_methods_class("density")
```

A tibble: 4 × 4

generic <chr>	class <chr>	visible <lgl>	source <chr>
coerce	density	TRUE	
initialize	density	TRUE	
show	density	TRUE	
slotsFromS3	density	TRUE	

4 rows

Efectivamente los métodos que no aparecían pertenecían a S4. Por lo tanto, en S3 solo están implementadas las operaciones más básicas.

3.1.2 Entendiendo la clase `htest`

Ya expuesto el funcionamiento general y manejo de las clases y objetos de *R*, adentrémonos en la clase `htest` puesto que lo que haremos será construir un objeto de esa clase. Para ello, veamos a qué estructura de datos

se le asigna esa clase.

¿Que devuelve `class(unclass(test t))`?¿Por que?

Cuando ejecutamos `class(unclass(test t))` sobre un objeto de clase `htest` obtenemos una lista de 10 elementos pues `unclass()` lo que hace es devolver una copia del argumento pero con su atributo removido y `test t` contiene el resultado de hacer una llamada con la función `t.test`. Función que devuelve una lista con la clase `htest` que tiene la información sensible al test como: el estadístico, el pvalor, el intervalo de confianza y demás. Entre toda esta información sensible cuenta con 10 elementos. Por lo tanto, a la hora de implementar el Test de Wilcoxon nuestro output deberá ser una lista con la clase `htest`.

3.2 Resultados y demostraciones

3.2.1 Simetría de D = X - Y (Pregunta 5)

Como justamente dice el enunciado este resultado es fundamental para entender por qué importa el Test de Wilcoxon y como provee una herramienta para realizar tests para muestras apareadas.

Bajo H_0: $F_X = F_Y$ (los tratamientos son indistinguibles) y asumiendo que la asignación de cada individuo al tratamiento se realiza de forma aleatoria, la distribución conjunta $F_{X,Y}$ del vector aleatorio (X, Y) es la misma que la del vector (Y, X); es decir, $F_{X,Y} = F_{Y,X}$. Probar que entonces la distribución de $D = X - Y$ es simétrica alrededor del cero

Demostración

Queremos ver que $D = X - Y$ es simétrica alrededor del 0. Para eso basta con ver que $f_D(t) = f_D(-t)\ \forall t > 0$. Para despejar f_D invocamos el Teorema de Cambio de Variables con $g(x, y) = (x-y, x)$, $g : \underbrace{\mathbb{R} \times \mathbb{R}}_{G_1} \to \underbrace{\mathbb{R} \times \mathbb{R}}_{G_2}$. Con G_1, G_2 abiertos, $g \in C^1$, $P((x, y) \in \mathbb{R}^2) = 1$ (pues estamos trabajando con variables continuas aunque también vale para discretas). Entonces, ahora

debemos despejar $g^{-1}(d,v)$:

$$\implies (x-y,x)=(d,v) \implies \begin{cases} v=x \\ d=x-y \end{cases} \implies \begin{cases} x=v \\ y=v-d \end{cases} \implies g^{-1}(d,v)=(v,v-d) \tag{1}$$

$$|detD_{g^{-1}}(d,v)| = \begin{vmatrix} 0 & 1 \\ -1 & 1 \end{vmatrix} = 1 \tag{2}$$

Entonces, invocando el teorema, $f_{dv}(d,v) = f_{XY}(v,v-d)\mathbf{1}_{G_2}(d,v)$. Como X,Y son continuas, $\mathbf{1}_{G_2}(d,v)$ vale siempre 1. Como $v=x$ podemos directamente reemplazar por x: $f_{dv}(d,v) = f_{XY}(x,x-d)$. Así obtenemos:

$f_d(d) = \int_{-\infty}^{+\infty} f_{XY}(x,x-d)dx$

$f_d(-d) = \int_{-\infty}^{+\infty} f_{XY}(x,x-(-d))dx = \int_{-\infty}^{+\infty} f_{XY}(x,x+d))dx$

¿Vale la igualdad? Por hipótesis sabemos que $F_{XY} = F_{YX}$ y esto implica que $f_{XY} = f_{YX}$. Además, D = X - Y lo que implica que X = D + Y y que Y = X - D. Entonces, $f_d(d) = \int_{-\infty}^{+\infty} f_{XY}(x,x-d)dx = \int_{-\infty}^{+\infty} f_{YX}(x-d,x)dx$ $= \int_{-\infty}^{+\infty} f_{YX}(y,y+d)d(y+d) = \int_{-\infty}^{+\infty} f_{XY}(y,y+d)dy$, pues d es una constante.

Nos queda una integral respecto de y y tenemos una variable y dando vueltas. No obstante, y no es más que una "dummy variable" o una variable de integración por lo que si la reemplazamos por x no apareja cambio alguno. Así queda: $f_d(d) = \int_{-\infty}^{+\infty} f_{XY}(y,y+d)dy = \int_{-\infty}^{+\infty} f_{XY}(x,x+d)dx = f_d(-d)$.

3.2.2 Resultados para analizar la distribución de T^+ (Preguntas 6, 7, 8 y 9).

Para conocer la distribución de T^+ será necesario probar primero estos resultados.

Muestre que los siguientes estadísticos:

$$T^+ = \sum_{i=1}^{n} \mathbf{1}\{X_i > 0\} R_i \quad T^- = \sum_{i=1}^{n} \mathbf{1}\{X_i < 0\} R_i$$

son equivalentes a T (i.e., muestre que a partir de cualquiera de los 3 y conociendo n, se pueden computar exactamente los otros dos). Sugerencia: calcule $T^+ + T^-$.

Demostración

Para mostrar que a partir de cualquiera de los 3 y conociendo n podemos computar los otros 2, usaremos las siguientes equivalencias:

$$\begin{cases} T = T^+ - T^- \\ T^+ = T + T^- \\ T^- = T^+ - T \\ T^+ + T^- = \dfrac{n(n+1)}{2} \end{cases}$$

$$T^+ + T^- = \sum_{i=1}^{n} \mathbf{1}\{X_i > 0\} R_i + \sum_{i=1}^{n} \mathbf{1}\{X_i < 0\} R_i = \sum_{i=1}^{n} R_i = \sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

Podemos afirmar que es la suma de los rangos pues $P(X_i = 0) = 0$ y, por lo tanto, se justifica la igualdad. Entonces, ya con estas igualdades en mano podemos ver los diferentes casos:

- Supongamos que conocemos T y n por lo que queremos despejar T^+ y T^-.

Podemos definir el sistema de ecuaciones:

$$\begin{cases} T = T^+ - T^- \\ \dfrac{n(n+1)}{2} = T^+ + T^- \end{cases}$$

Como es un sistema de 2 ecuaciones y 2 incógnitas, el sistema puede ser SCD o SCI. Para ver esto consideramos la matriz del sistema y calculamos el determinante.

$$det \begin{vmatrix} 1 & -1 \\ 1 & 1 \end{vmatrix} = 2$$

Como es diferente de cero la matriz es no singular es decir que el sistema tiene solución única. Podemos despejar lo buscado.

Para los otros casos el procedimiento es análogo

- Con T^+ y n conocidos

  Definimos el sistema:

$$\begin{cases} T^+ = T + T^- \\ \dfrac{n(n+1)}{2} - T^+ = T^- \end{cases}$$

  Con el determinante:

$$det \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1$$

  El sistema es SCD por lo que existe solución única.
  Y por último,

- Con T^- y n conocidos

  Definimos el sistema:

$$\begin{cases} T^- = T^+ - T \\ \dfrac{n(n+1)}{2} - T^- = T^+ \end{cases}$$

  Con el determinante:

$$det \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1$$

Nuevamente, el sistema es SCD por lo que tiene solución única.

En conclusión, demostramos que a partir de cualquiera de los tres estadisticos podemos obtener los otros dos, por ende, T, T^+ y T^- son equivalentes.

Demuestre que, bajo H_0, $|X_i|$ es independiente de $signo(X_i)$.

Demostración

Demostremos que $|X_i|$ es independiente del signo(X_i): Para esto usemos la definición: son independientes si la probabilidad conjunta se factoriza. Comenzando por la conjunta:

$$P(|X_i| \leq t,\, \text{sg}(X_i) = 1) = P(-t \leq X_i \leq t, X_i > 0) \underbrace{=}_{\text{Intersección}} P(0 \leq X_i \leq t)$$

Por otro lado:

$$P(|X_i| \leq t) \cdot P(\text{sg}(X_i) = 1) = P(-t \leq X_i \leq t) \cdot P(X_i > 0)$$

Usando que, bajo H_0, X_i tiene distribución simétrica respecto del 0, tenemos que:

$$\begin{cases} P(-t \leq X_i \leq t) = 2 \cdot P(0 \leq X_i \leq t) \\ P(X_i > 0) = P(X_i < 0) = \frac{1}{2} \end{cases}$$

$$\implies \quad P(|X_i| \leq t) \cdot P(\text{sg}(X_i) = 1) = 2 \cdot P(0 \leq X_i \leq t) \cdot \tfrac{1}{2} = P(0 \leq X_i \leq t)$$

Por lo tanto:

$$P(|X_i| \leq t,\, \text{sg}(X_i) = 1) = P(0 \leq X_i \leq t) = P(|X_i| \leq t) \cdot P(\text{sg}(X_i) = 1)$$

Como este procedimiento es análogo para el caso de $\text{sg}(X_i) = -1$ y $P(\text{sg}(X_i) = 0) = 0$, demostramos que la probabilidad conjunta de $|X_i|$ y $\text{sg}(X_i)$ se factoriza. Es decir, $|X_i|$ es independiente del signo(X_i).

A partir del resultado anterior, pruebe que, bajo H_0 , los vectores de rangos $\mathbf{R} = (R_1, \ldots, R_n)$ y antirrangos $\mathbf{D} = (D_1, \ldots, D_n)$, correspondientes a $|\mathbf{X}|$, son independientes del vector de signos $\mathbf{S} = (signo(X_1), \ldots, signo(X_n))$ de la muestra original $\mathbf{X}$.

Demostración

Del resultado anterior tenemos que $|X_i|$ es independiente de $S(X_i)$, además sabemos que $|X_j|$ es independiente de $S(X_l)$ $\forall$ j $\neq$ l puesto que las X_i son independientes entre sí y aquí le estamos aplicando dos funciones a dos variables aleatorias distintas e independientes. Por lo tanto, se mantiene la independencia.

Entonces, podemos decir que los vectores $(|X_1|, |X_2|, ..., |X_n|)$ y $(s(X_1), s(X_2), ..., s(X_n))$ son independientes entre sí. Con esto en mente la demostración se simplifica bastante pues $\mathbf{R} = (R(|X_1|), R(|X_2|), ..., R(|X_n|))$ es equivalente a aplicar una función $G(X)$ sobre $|\mathbf{X}|$. Y nosotros sabemos que si $|\mathbf{X}|$ y $\mathbf{S}$ son independientes, $G(|\mathbf{X}|)$ y $\mathbf{S}$ también. Por lo tanto, $\mathbf{R}$ es independiente de $\mathbf{S}$.

Lo análogo ocurre con $\mathbf{D} = (D(|X_1|), D(|X_2|), ..., D(|X_n|))$. Como $\mathbf{D}$ es una $H(|\mathbf{X}|)$ y $|\mathbf{X}|$ y $\mathbf{S}$ son independientes, $H(|\mathbf{X}|)$ y $\mathbf{S}$ también. Por lo tanto, $\mathbf{D}$ es independiente de $\mathbf{S}$.

Pruebe que, bajo $H_0 : \theta = 0$, $F \in \Omega_s$, las variables aleatorias $W_j = \mathbf{1}\{X_{D_j} > 0\}$ distribuyen según $W_1, \ldots, W_n \sim^{iid}$ Bernoulli $\left(\frac{1}{2}\right)$.

Demostración

Sea $S_n = \{(d_1, d_n), 1 \leq d_i \neq d_j \leq n\}$ el espacio de las permutaciones de $\{1, ..., n\}$ y sea $\underline{D} = (D_1, ..., D_n)$ el vector de antirangos. Queremos demostrar que $W_j = \mathbf{1}\{X_{D_j} > 0\}$ son independientes e identicamente distribuidas según $Be(\frac{1}{2})$. Para ello, utilizando la Ley de Probabilidad Total, escribimos:

$$P(W_1 = a_1, W_2 = a_2, ..., W_n = a_n) = \sum_{\underline{d} \in S_n} P(\underline{W} = \underline{a} | \underline{D} = \underline{d}) P(\underline{D} = \underline{d}) =$$

$$\sum_{\underline{d} \in S_n} P(I(X_{D_j} > 0) = a_1, ..., I(X_{D_n} > 0) = a_n | D_1 = d_1, ..., D_n = d_n) P(\underline{D} =$$

$\underline{d}) =$

$$\sum_{\underline{d}\in S_n} P((I(X_{d_1}>0)=a_1,...,I(X_{d_n}>0)=a_n|D=\underline{d})P(\underline{D}=\underline{d}) \underbrace{=}_{\underline{D} \text{ y } signo(X_i) \text{ son indep}}$$

$$\sum_{\underline{d}\in S_n} \underbrace{P(I(X_{d_1}>0)=a_1,...,I(X_{d_n}>0)=a_n)}_{\prod_{i=1}^n P(I(X_{d_1}>0)=a_i)=\prod_{i=1}^n \frac{1}{2}} P(\underline{D}=\underline{d}) =$$

$$\sum_{\underline{d}\in S_n} (\tfrac{1}{2})^n P(\underline{D}=\underline{d}) = \tfrac{1}{2^n} \underbrace{\sum_{\underline{d}\in S_n} P(\underline{D}=\underline{d})}_{=1} = \tfrac{1}{2^n}$$

$\Rightarrow W_1, W_2, ..., W_n$ son variables independientes y cada una tiene una distribución $Be(\frac{1}{2})$

3.2.3 Distribución exacta de T^+ (Preguntas 10 y 11)

Como puede verse en los resultados probados más arriba, T^+ es una suma ponderada de Bernoullis por lo que no podemos decir que su distribución es una binomial. Por lo tanto, veamos cómo serían los valores que puede tomar la puntual del estadístico.

Reproduzca la tabla de Observación 6 para $n = 5$

t	$S_{n,t}$	$\#S_{n,t}$	$p_5(t)$
0	$\{\emptyset\}$	1	$\frac{1}{32}$
1	$\{\{1\}\}$	1	$\frac{1}{32}$
2	$\{\{2\}\}$	1	$\frac{1}{32}$
3	$\{\{3\},\{1,2\}\}$	2	$\frac{2}{32}$
4	$\{\{4\},\{1,3\}\}$	2	$\frac{2}{32}$
5	$\{\{5\},\{1,4\},\{2,3\}\}$	3	$\frac{3}{32}$
6	$\{\{1,5\},\{2,4\},\{1,2,3\}\}$	3	$\frac{3}{32}$
7	$\{\{2,5\},\{3,4\},\{1,2,4\}\}$	3	$\frac{3}{32}$
8	$\{\{3,5\},\{1,2,5\},\{1,3,4\}\}$	3	$\frac{3}{32}$
9	$\{\{4,5\},\{1,3,5\},\{2,3,4\}\}$	3	$\frac{3}{32}$
10	$\{\{1,4,5\},\{2,3,5\},\{1,2,3,4\}\}$	3	$\frac{3}{32}$
11	$\{\{2,4,5\},\{1,2,3,5\}\}$	2	$\frac{2}{32}$
12	$\{\{3,4,5\},\{1,2,4,5\}\}$	2	$\frac{2}{32}$
13	$\{\{1,3,4,5\}\}$	1	$\frac{1}{32}$
14	$\{\{2,3,4,5\}\}$	1	$\frac{1}{32}$
15	$\{\{1,2,3,4,5\}\}$	1	$\frac{1}{32}$

Si prestamos atención a la tabla notamos que hay una fuerte simetría en los valores respecto de la mediana. Intentar probarla de forma exhaustiva o por inducción no es posible dado a que la cantidad de valores escala muy rápido así que veamos otra forma.

Muestre que T^+ es simétrica alrededor de $\frac{n(n+1)}{4}$.

Demostración

Para probar esto utilizaremos un argumento muy parecido a la demostración de simetría realizada previamente. En el fondo vamos a querer ver que $p_n(\frac{n(n+1)}{4}+t) = p_n(\frac{n(n+1)}{4}-t)$. No obstante, esto sería lo mismo que probar que $p_n(t) = p_n(\frac{n(n+1)}{2}-t)$, que simplifica la notación.

Utilizando la definición de arriba, probar $p_n(t) = p_n(\frac{n(n+1)}{2}-t)$ equivale a probar $\frac{\#S_{n,t}}{2^n} = \frac{\#S_{n,\frac{n(n+1)}{2}-t}}{2^n}$, que se reduce a ver que $\#S_{n,t} = \#S_{n,\frac{n(n+1)}{2}-t}$.

Si hallamos la biyección entre ambos conjuntos queda probado y no

hay más que hacer. Si $\omega \in \#S_{n,t} \implies T^+ = t$ y como $T^+ + T^- = \frac{n(n+1)}{2} \implies T^- = \frac{n(n+1)}{2} - t$. Esto es lo que nos dice es que $\forall \omega \in \#S_{n,t}$, ω^c es tal que la suma de sus rangos es $\frac{n(n+1)}{2} - t$. Por lo tanto, para cada conjunto tal que la suma de sus rangos es t podemos asociarlo a un conjunto tal que la suma de sus rangos es $\frac{n(n+1)}{2} - t$.

Ahora, si consideramos a los $\omega \in \#S_{n,\frac{n(n+1)}{2}-t} \implies T^+ = \frac{n(n+1)}{2} - t$ y como $T^+ + T^- = \frac{n(n+1)}{2} \implies T^- = t$. Esto es lo que nos dice es que $\forall \omega \in \#S_{n,\frac{n(n+1)}{2}-t}$, ω^c es tal que la suma de sus rangos es t.

Hallamos tal biyección. De forma un tanto hablada pero la hallamos y queda probada la simetría.

3.2.4 Resultado para analizar la distribución asintótica de T^+ (Pregunta 15).

Para analizar la distibución asintótica de T^+ primero es necesario calcular su esperanza y varianza

Bajo H_0, ¿Cuánto vale $\mathbb{E}[T^+]$? Y $\mathbb{V}[T^+]$?

Bajo H_0 notamos que

$$T^+ = \sum_{i=1}^{n} iW_i \qquad \text{con } W_1, \ldots, W_n \sim \text{iid Bernoulli}\left(\frac{1}{2}\right)$$

Por lo tanto, utilizando propiedades de la esperanza

$$\mathbb{E}[T^+] = \mathbb{E}[\sum_{i=1}^{n} iW_i] = \sum_{i=1}^{n} i\mathbb{E}[W_i] = \frac{1}{2}\sum_{i=1}^{n} i = \frac{1}{2}\frac{n(n+1)}{2} = \frac{n(n+1)}{4}$$

$$\mathbb{V}[T^+] = \mathbb{V}[\sum_{i=1}^{n} iW_i] = \sum_{i=1}^{n} i^2\mathbb{V}[W_i] + \sum_{1\le i<j}^{n} ij \underbrace{\text{Cov}(W_i, W_j)}_{=0 \text{ pues ind. bajo } H_0} = \frac{1}{4}\sum_{i=1}^{n} i^2 = \frac{1}{4}\frac{n(n+1)(2n+1)}{6}$$

3.2.5 Distribución asintótica de T^+

Como se puede ver en su esperanza y varianza, T^+ es una suma ponderada de Bernoullis por lo que no sigue una distribución binomial. Es por esto, que no podremos utilizar el Teorema Central del Limite como lo conocemos sino que trendremos que requerir de una de sus variaciones.

Dé la distribución asintótica de T^+

Para esto usaremos el Teorema Central del Limite de Lindeberg el cual nos dice lo siguiente:

Sean $X_1, \ldots, X_n$ iid con $\mathbb{E}[X_i] = 0$ y $\mathbb{V}[X_i] = \sigma^2 < \infty$. Definimos $S = \sum_{i=1}^{n} \frac{a_i X_i}{\sqrt{n}}$

$$\text{Si} \quad \frac{\max_i |a_i|}{\sqrt{\sum_{i=1}^n a_i^2}} \to 0$$

$$\text{Entonces} \quad \frac{S}{\sqrt{\mathbb{V}[S]}} \xrightarrow{D} N(0,1) \quad \text{con } \mathbb{V}[S] = \frac{\sigma^2(\sum_{i=1}^n a_i^2)}{n}$$

Comenzamos construyendonos las X_i:

$$T^+ = \sum_{i=1}^{n} iW_i \implies T^+ - \mathbb{E}[T^+] = \sum_{i=1}^{n} iW_i - \sum_{i=1}^{n} i\frac{1}{2} = \sum_{i=1}^{n} i \underbrace{(W_i - \frac{1}{2})}_{X_i}$$

$$\text{Notemos que} \begin{cases} \mathbb{E}[X_i] = & \mathbb{E}[W_i - \frac{1}{2}] = \frac{1}{2} - \frac{1}{2} = 0 \\ \mathbb{V}[X_i] = & \mathbb{E}[W_i - \frac{1}{2}] = \mathbb{V}[W_i] = \frac{1}{4} < \infty \end{cases}$$

Es decir, conseguimos las variables X_i segun piden las hipótesis del teorema.

Veamos que $i = a_i$ cumple la condición.

$$\frac{\max_{1\le i\le n} |a_i|}{\sqrt{\sum_{i=1}^n a_i^2}} = \frac{\max_{1\le i\le n|i|} }{\sqrt{\sum_{i=1}^n i^2}} = \frac{n}{\sqrt{\frac{n(n+1)(2n+1)}{6}}} \le \frac{n}{\sqrt{\frac{nn2n}{6}}} = \frac{n}{n\sqrt{\frac{2n}{6}}} = \frac{1}{\sqrt{\frac{2n}{6}}} \overset{n\to\infty}{\to} 0$$

Logramos construirnos una S tal que se cumpla lo pedido:

$$T^+ - \mathbb{E}[T^+] = \sum_{i=1}^{n} i(W_i - \frac{1}{2}) = \sqrt{n}S \implies \frac{T^+ - \mathbb{E}[T^+]}{\sqrt{n}} = S$$

Por último, calculemos $\mathbb{V}[S]$:

$$\sqrt{\mathbb{V}[S]} = \sqrt{\frac{\sigma^2(\sum_{i=1}^{n} a_i^2)}{n}} = \sqrt{\frac{\frac{1}{4}(\sum_{i=1}^{n} i^2)}{n}} = \sqrt{\frac{\mathbb{V}[S]}{n}}$$

Ahora sí, estamos en condiciones de calcular la distribución asintótica usando el teorema.

$$\frac{S}{\sqrt{\mathbb{V}[S]}} = \frac{\sum_{i=1}^{n} \frac{a_i X_i}{\sqrt{n}}}{\sqrt{\frac{\sigma^2(\sum_{i=1}^{n} a_i^2)}{n}}} = \frac{\sum_{i=1}^{n} a_i X_i}{\sqrt{\sigma^2(\sum_{i=1}^{n} a_i^2)}} = \frac{\sum_{i=1}^{n} i(W_i - \frac{1}{2})}{\sqrt{\frac{1}{4}(\sum_{i=1}^{n} i^2)}} = \frac{T^+ - \mathbb{E}[T^+]}{\sqrt{\mathbb{V}[T^+]}} \xrightarrow{D} N(0,1)$$

Por lo tanto, $T^+ \xrightarrow{D} N(\mathbb{E}[T^+], \mathbb{V}[T^+])$.

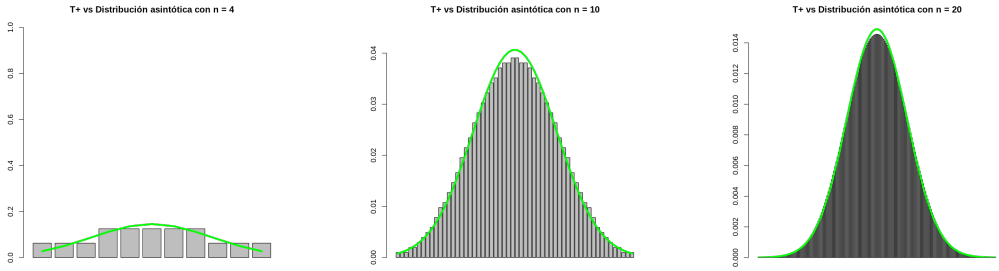
Hagamos una comparación entre las puntuales de T^+ y su distribución asintótica para diferentes valores de n.

Fije $n_1 = 4$, $n_2 = 10$, $n_3 = 20$. Para cada n, realice un gráfico de barras con la probabilidad puntual exacta de T_+ (válgase de dTmas) y superpóngale una línea con la densidad asintótica esperada. ¿Coinciden razonablemente? ¿En toda la distribución, en el centro, en las colas? ¿A partir de qué n? ¿Se le ocurre alguna corrección sencilla para los n pequeños?

Coinciden razonablemente aunque a medida que n_i crece los graficos comienzan a superponerse mejor y vemos que en n_3 es casi exacta. Completamente esperado dado que estamos trabajando con distribuciones asintóticas.

Para n_1 la superposición es razonablemente buena aunque en las colas la curva queda debajo de las barras.

Para n_2 la superposición es buena pero la curva queda centrada apenas más arriba que las barras.



Para el caso de n_1 no hay a priori una corrección evidente pues son pocos los datos entonces es esperable una estimación pobre. Aunque quizás podría evaluarse la densidad en algunos puntos de interés y unir esos puntos con una recta.

Para el caso de n_2 podría multiplicarse la densidad por algún escalar cercano a 1 y así reducir la distancia del máximo a las curvas sin perder el buen fiteo que tiene.

3.2.6 Performance del test de Wilcoxon

Los siguientes datos fueron generados por $D = N(1,1), n = 12$: Compute Φ_w, el test de Wilcoxon de rango signado de nivel menor o igual a $\alpha = 0.05$ para las hipótesis: $H_0 : \theta = 0$ versus $H_1 : \theta > 0$.
A continuación, fije $\theta_1 = 1, m = 10.000$ y estime por bootstrap la potencia.

Como el estimador de la potencia por bootstrap está dado por

$$\pi_{\hat{\Phi}_w}(\theta_1) = \frac{1}{m}\sum_{i=1}^{n} 1_{\{T^+(Y_i)>k^*\}}$$

y m es conocido y T^+ sabemos calcularlo, solo queda hallar k^*.

Para despejarlo hay que asegurarse de que cumpla con el nivel de test pedido, es decir,

$$P_{H_0}(T^+ > k^*) = 0.05$$

Como T^+ es discreto es posible que no exista valor de k tal que se cumpla esa igualdad por lo que hay que encontrar

$$k^* = argmax_k\{F_{T^+}(k) : F_{T^+}(k) \leq 0,95\}$$

Como T^+ tiene soporte entre 0 y $\frac{n(n+1)}{2}$ alcanza calcular la acumulada para cada valor del intervalo y recorrerlo hasta encontrar k^*.

Con los datos del problema, $n = 12$ y $\frac{n(n+1)}{2} = 78$, entonces, siguiendo el razonamiento desarrollado hallamos $k^* = 59$.

Ya calculada la región de rechazo del test podemos computar el estimador por bootstrap de la potencia, obteniendo así:

$$\hat{\pi_{\phi_w}}(\theta_1) = 0.9407$$

Compute para las mismas hipótesis y condiciones de Pregunta 18, ϕ_n, un test para el valor de la media (y mediana) de $v.a.i.i.d.$ con varianza conocida, según D.
Calcule (analíticamente, sin estimar) la potencia $\pi_{\phi_n}(\theta_1)$. Compute además Φ_s, el test del signo para las mismas hipótesis y estime por bootstrap la potencia en θ_1.
Compare y contraste los resultados obtenidos. ¿Es el test t efectivamente el más potente? ¿Por cuánto?

Sabemos que el test n (con varianza conocida) usa como estadístico:

$$T = \frac{(\overline{X} - \theta_0)}{\frac{\sqrt{\mathbb{V}x}}{\sqrt{n}}} \sim N(0,1) \implies \phi_t = \mathbf{1}\{T > z_{1-\alpha}\}/ \quad P(Z \leq z_{1-\alpha}) = 1-\alpha \quad Z \sim N(0,1)$$

Si queremos calcular la potencia para $\theta_1 = 1$:

$$\pi_{\Phi_t}(\theta_1) = P_{\theta_1}(\frac{(\overline{X} - \theta_0)}{\frac{\sqrt{\mathbb{V}x}}{\sqrt{n}}} > z_{1-\alpha}) \underbrace{=}_{\text{normalizo para } \theta_1} P_{\theta_1}(\frac{(\overline{X} - \theta_1}{\frac{\sqrt{\mathbb{V}x}}{\sqrt{n}}} > z_{1-\alpha} + \frac{(\theta_0 - \theta_1)\sqrt{n}}{\sqrt{\mathbb{V}x}})$$

Si reemplazamos con los datos provistos llegamos a:

$$P_{\theta_1}(\frac{(\overline{X} - \theta_1}{\frac{\sqrt{\mathbb{V}x}}{\sqrt{n}}} > z_{1-\alpha} + \frac{(\theta_0 - \theta_1)\sqrt{n}}{\sqrt{\mathbb{V}x}}) = P_{\theta_1}(\frac{(\overline{X} - \theta_1)}{\frac{\sqrt{\mathbb{V}x}}{\sqrt{12}}} > z_{0.95} + \frac{(-1)\sqrt{12}}{\sqrt{\mathbb{V}x}})$$

Bajo la hipótesis de $\theta = \theta_1$, $\quad \frac{(\overline{X}-\theta_1)}{\frac{\sqrt{\mathbb{V}x}}{\sqrt{12}}} \sim N(0,1) = Z$

$$P(Z > 1.65 + \frac{(-1)\sqrt{12}}{\sqrt{\mathbb{V}x}}) = \pi_{\hat{\phi_n}}(\theta_1) = 0.9656$$

Luego, podemos computar de manera sencilla un test del signo basandonos en las siguientes variables:

$$\text{Sea} \quad T_i = \mathbf{1}\{X_i > \theta_0\} \sim Be(\lambda)$$

Proponemos el estadístico:

$$\overline{T_n} = \frac{\sum_{i=1}^n T_i}{n} \quad \Longrightarrow \quad \sum_{i=1}^n T_i \sim Bi(n, \lambda)$$

Observamos que bajo $H_0 \quad \lambda = \frac{1}{2}$, por lo que testeamos para las hipótesis

$$\begin{cases} H_0 : \lambda = \frac{1}{2} \\ H_1 : \lambda > \frac{1}{2} \end{cases} \iff \begin{cases} H_0 : \theta = Med_F(X) = 0 \\ H_1 : \theta = Med_F(X) > 0 \end{cases}$$

Ahora bien, para estimar por bootstrap la potencia, debemos hallar la región de rechazo.

$$P_{H_0}(\overline{T_{12}} > k) = 0.05 \iff P_{H_0}(\textstyle\sum_{i=1}^{12} T_i > 12k) = 0.05$$
$$\iff P_{H_0}(X > 12k) = 0.05 \quad \text{con } X \sim Bi(12, \tfrac{1}{2})$$

Como $Bi(n, \frac{1}{2})$ es discreta, nuevamente puede que el k no exista por lo que utilizamos el mismo enfoque que en el punto anterior. Calculamos la acumulada para todo el soporte de la binomial y encontramos el valor que la maximice y siga siendo menor o igual que 0.95

$$\text{El cuantil es } q = 8 \implies k = \frac{8}{12} \implies k \approx 0.6667$$
$$\implies \phi_s = \mathbf{1}\{\overline{T_i} > 0.6667\}$$

Ya con esto podemos computar la potencia via bootstrap

$$\hat{\pi_{\phi_s}}(\theta_1) = 0.8886$$

Estos resultados nos permiten ver que $\hat{\pi_{\phi_s}}(\theta_1) \leq \hat{\pi_{\phi_w}}(\theta_1) \leq \hat{\pi_{\phi_n}}(\theta_1)$. Esto se condice con lo que sabemos pues $\phi_n(t)$ es el test UMP para estas hipótesis. No obstante, nos permite afirmar que dada una alternativa fija es preferible el test de Wilcoxon sobre el test del signo.

A su vez, cuando calculamos las distancias entre las potencias vemos que $d(\hat{\pi_{\phi_n}}(\theta_1), \hat{\pi_{\phi_w}}(\theta_1)) = 0.0249$ no es tanta. Mientras que $d(\hat{\pi_{\phi_n}}(\theta_1), \hat{\pi_{\phi_s}}(\theta_1)) = 0.0772$ es casi el triple. Esto que podemos visualizar es importante pues nos deja entrever que el test de Wilcoxon tiene una potencia bastante buena para testear las hipótesis propuestas; potencia que hasta se asemeja a la del test UMP, por lo que entre el test del signo y el de Wilcoxon sería preferible elegir este último.

4 Conclusión

Como pudimos ver en los dos ejercicios anteriores, la potencia del test de Wilcoxon dada una alternativa fija no dista tanto de la del test normal UMP (0.0249). Si bien este resultado puede quizás parecer trivial o no decir mucho por su cuenta, es un resultado que en el fondo aporta información muy valiosa. Pues si bien el test normal es el test UMP para cualquier par de hipótesis unilaterales, para poder aplicarlo debemos asumir que nuestra muestra tiene una distribución $F(x, \theta) \in \mathcal{F} = \{F(x, \theta) : \theta \in \Theta\}$, es decir, debemos ajustar un modelo parámetrico a la muestra. Ajuste no trivial dado que se deben realizar una serie de supuestos acerca de nuestros datos.

Por lo tanto, si estos supuestos no pueden ser realizados, la herramienta deja de ser útil y debemos recurrir a un test no paramétrico como lo es el del signo o el de Wilcoxon. Por ende, el hecho de que las potencias del test de Wilcoxon y del test UMP no disten tanto lo que nos dice, en el fondo, es que el test de Wilcoxon es una gran herramienta que poco debe envidiarle al test UMP cuando tenemos una alternativa fija. Es un test donde no debemos realizar ningún supuesto salvo el de simetría y que tiene una performance casi tan buena como el de la normal.

No obstante, con el test del signo no ocurre lo mismo pues su potencia tiene una distancia considerable a la del test UMP (0.0772). Por lo tanto, no cualquier test paramétrico resulta en una gran herramienta y esto aporta muchísimo valor al resultado que nos dio Wilcoxon[1] posicionandolo como un recurso a tener en consideración siempre que se trabaje con datos.

[1]En este caso en particular, justo Wilcoxon se desprende del test del signo entonces podemos entender que parte de la diferencia en los resultados se da porque el test del signo es "miope" frente a la posición relativa respecto de la mediana mientras que Wilcoxon no.