San Antonio de Los Altos 5 de Marzo de 2018

USO LA INFORMACIÓN DE SHANNON EN LA COMPARACIÓN DE LOS TEST NO PARAMÉTRICOS: TEST DE LOS SIGNOS Y WILCOXON

Andrés E Reyes Polanco.
Prof. Asociado UCV

RESUMEN

El siguiente artículo trata de algunas aplicaciones de la Teoría de la Información en el campo de la estadística, tales como: el caso de los problemas de estimación de densidades, árboles de decisión, para después plantear la aplicación en la comparación de los test no paramétricos: de los signos y de Wilcoxon para muestras relacionadas, esto es: dada una hipótesis nula de locación de muestras pareadas y dado estos dos estadísticos no paramétrico que permiten realzar el contraste de la hipótesis de locación, comparar la entropía de los mismos y la cantidad de información aportada, considerando los conceptos de Shannon. La idea de esta comparación entre test no paramétricos vía la función de entropía proviene de mi primer trabajo de ascenso que presenté en la Universidad Central de Venezuela. En el primer punto, se explica las propiedades básicas de esta función y posteriormente algunas de sus aplicaciones en Estadística y en el segundo, la aplicación en la comparación de los dos test no paramétrico propuestos.

Palabras claves: entropía, información de Shannon, test no paramétrico: de los signos, Wilcoxon.

USE THE SHANNON INFORMATION IN THE COMPARISON OF THE NON PARAMETRIC TEST: TEST OF THE SIGNS AND WILCOXON

Andrés E Reyes Polanco.
Associate Professor UCV

SUMMARY

The following article deals with some applications of the Theory of Information in the field of statistics, such as: the case of problems of density estimation, decision trees, and then raise the application in the comparison of non-parametric tests : of the signs and of Wilcoxon for related samples, that is: given a null hypothesis of locating paired samples and given these two non-parametric statistics that allow to enhance the contrast of the location hypothesis, compare the entropy of the same and the quantity of information provided, considering the concepts of Shannon. The idea of this comparison between nonparametric tests via the entropy function comes from my first job of promotion that I presented at the Central University of Venezuela. In the first point, it explains the basic properties of this function and later some of its applications in Statistics and in the second, the application in the comparison of the two proposed non-parametric tests.

Keywords: entropy, Shannon information, nonparametric test: of the signs, Wilcoxon.

INTRODUCCIÓN

El siguiente artículo trata de algunas aplicaciones de la Teoría de la Información a diferentes temas de Estadística, su uso en esta disciplina fue introducido por Kullback en 1951 con su obra: Information Theory and Statistics, reeditado por Dover Publications (1997) desarrollando el concepto de divergencia que tiene hoy en día una amplia aplicación en diferentes problemas de Estadística, tales como el estudio de tablas de contingencia Kullback, S; Keegel, C (1991), contraste entre distribución teórica y su aproximación, estimación de densidad Kullback, S; (1997), etc. En nuestro caso tratamos el uso del concepto de entropía de Shannon

(1948) brevemente y algunas aplicaciones como estimación de la densidad de una variable aleatoria continua, árbol de decisión, comparación entre test no paramétrico que se emplean en el contraste de hipótesis para muestras pareadas, una idea que planteé en mi primer trabajo de ascenso en la Universidad Central de Venezuela presentado en 1979. Hoy lo presento nuevamente aquella idea, pero con mayor madures, por tanto más desarrollada. Para hacer las comparaciones, partiremos de las distribuciones exactas bajo la hipótesis nula de cada uno de los estadísticos de contraste y luego calculamos la función de entropía de Shannon y comparamos los resultados. En nuestro caso, estas distribuciones son discretas y siempre se puede definir la función de entropía. La idea original de la entropía que Shannon presentó era para minimizar el tiempo de transmisión de los mensajes, maximizando la función de entropía. Posteriormente su aplicación en diferentes disciplinas se hizo muy corriente, por ejemplo; la entropía de Shannon se emplea como medida de diversidad en ecología Laura Pla (2006). En economía está entre otros casos el problema de estimación de la elasticidad presentado por Cook L et al (2015), en los sistemas de información en las empresas: Shwarz D (2014). Una buena referencia de las diferentes aplicaciones de la Teoría de la Información a diferentes campos del conocimiento se encuentra en Pardo Llorente (1993), en su artículo se estudia además, las medidas generalizadas de la entropía y de divergencia. En Estadística la aplicación va desde estimación de la densidad de una variable aleatoria como alternativa al método del núcleo (kernel), partiendo del concepto de máxima entropía: Raynal; J. A (2008), aplicación en series de tiempo en la estimación de los parámetros GARCH: Rockinger et al (2000) etc. Existe además, una amplia referencia del uso de la teoría de la información con el enfoque bayesiano. Entre los paquetes de Estadística está el SAS que en el módulo ETS contiene el procedimiento Entropy donde se da como ejemplos: la estimación de los errores del modelo lineal, diseño factorial. Este artículo contempla: I.-Definiciones básicos de entropía y algunas aplicaciones en Estadística; como la estimación de de la función de densidad de una variable aleatoria continua, construcción del árbol de decisión y contrate de hipótesis. II.-El uso del concepto de entropía para comparar los test no paramétricos de locación para muestras pareadas.

I.-Definiciones básicas de entropía de Shannon y algunas de sus aplicaciones en Estadística.

A continuación presentamos el concepto de entropía Shannon y las propiedades más importantes, sin demostrarlas y algunas de sus aplicaciones de este concepto en Estadística.

Definición 1: La función de entropía de Shannon.

Dado una variable aleatoria discreta: $X : E \to S_1 \subset E$ con función de probabilidad puntual: $P(X = x_i ; \Theta)$, entonces se define como función de entropía a:

$$H(X) = -\sum_{i=1}^{L} P(X = x_i ; \Theta) \log_k P(X = x_i ; \Theta)$$

Si X es una variable continua: $X : E \to S_1 \subset R$ con función de densidad: $f_X(x;\Theta)$, entonces se define como función de entropía a:

$$H(X) = -\int_{-\infty}^{\infty} f_X(x;\Theta) \log_k f_X(x;\Theta) dx$$

Esta función existe si $f_X(x;\Theta)$ es monótona (salvo en un intervalo finito) y además existe un número positivo α tal que la integral:

$$\int_R |x|^{\alpha} f_X(x;\Theta) dx$$

Converge. Álvarez; P,G (1981 pág. 63).

El logaritmo puede tomarse en base binaria (k=2); base neperiana (k=e) o decimal (k=10), en cada uno de los casos se dice que la entropía está medida en unidades binarias o bit, nep o decimal. En el caso continuo se utiliza el logaritmo natural.

Propiedades de la función de entropía.

Hay varias propiedades importantes de la función de entropía que enumeraremos algunas de ellas brevemente, considerando el caso de variables aleatorias discretas, siguiendo a Shannon (1948), Álvarez; P,G (1981), Evren et al (2015) tenemos:

1) $H(X)$ es una función continua en P

2) $H(X) \geq 0$ y es monótona no decreciente cuando se trata de variables aleatorias discretas. Cuando el sistema es perfectamente ordenado $H(X) = 0.$

3) Si la distribución es discreta uniforme, esto es, todos los eventos tienen igual probabilidad de $1/n$, entonces $H(X)$ alcanza su máximo y este es $MaxH(X) = \log_k n$

4) Dada dos variables aleatorias discretas: X e Y; $X : E \to S_1 \subset E$, $Y : E \to S_2 \subset E$ las cuales tienen como función de probabilidad conjunta: $P(X = x_i, Y = y_i; \Theta) = p_{ij}$ y como probabilidades marginales: $P(X = x_i; \Theta) = p_i.$ y $P(Y = y_i; \Theta) = p._j$, la cantidad de información sobre una variable aleatoria discreta X contenida en otra variable aleatoria discreta Y es: $I_Y(X) = H(X) - H_Y(X)$ donde $H_y(X)$ representa la entropía condicional de la variable aleatoria discreta X dado la otra variable aleatoria discreta Y, la cual se define por la relación: $H_Y(X) = \sum_{i=1}^{r} \sum_{j=1}^{s} p_{ij} \log_k p_{ij} / p._j = H(X,Y) - H(Y)$.

Por tanto, se obtiene la relación entre la información y las funciones de entropía: $I_Y(X) = H(X) + H(Y) - H(X,Y)$. Se demuestra que $I_Y(X) = H(X) - H_Y(X) \geq 0$. Si las variables X es Y son independientes entonces: $I_Y(X) = H(X) - H_Y(X)$ =0. Por tanto se verifica: $H(X,Y) = H(X) + H(Y)$

La cantidad de información de una variable aleatoria X contenida en otra Y, está acotada por su función de entropía $H(X)$, esto es:

$$I_Y(X) = H(X) - H_Y(X) \leq H(X)$$

Lo que implica que la cantidad de información contenida en ella misma es su entropía:

$$I_X(X) = H(X)$$

Esto significa, que una vez observada una muestra seleccionada de la variable aleatoria X, la función de entropía se interpreta como cantidad de información promedio contenida en la misma, puesto que al observar la muestra ya no existe incertidumbre. Entonces, si tenemos las observaciones muestrales de dos variables aleatorias X e Y, podemos comparar sus funciones de entropía para saber cuál de las dos aporta mayor información una vez observada la muestra. Si ambas variables están relacionadas, la cantidad de información que aporta la variable observada, digamos Y, sobre la otra no observada X viene dada por la diferencia entre la función de entropía de la variable no observada X menos la función de entropía condicional de X dada la observación de Y.

5) Consideremos $H(X)$ como la función de entropía de una población X distribuida como una multinomial y, consideremos una muestra aleatoria de X: $X_i; i = 1,2...n$ llamemos $\overset{*}{H}(X)$ estimador de $H(X)$:

$$\overset{*}{H}(X) = -\sum_{i}^{L} P^*(X = x_i) \log_k P^*(X = x_i)$$

Donde $P^*(X = x_i)$ es el estimador de $P(X = x_i)$

Para tamaño de muestra suficientemente grande se cumple: $E(\overset{*}{H}(X)) = H(X)$, es decir, es un estimador insesgado (no para muestras pequeñas) y

$$Var\overset{*}{H}(X) = ((1/n)\sum_{i}^{L} P^{*}(X = x_i)\log_k P^{*}(X = x_i)^2 - H(X)^2) + O(n^3)(k-1)/2n^2$$

Entonces:

$z = (\overset{*}{H}(X) - H(X))/\sqrt{V}ar(\overset{*}{H}(X) \rightarrow N(0,1)$ Converge en ley a la normal $N(0,1)$.

Luego, dadas las variables aleatorias X e Y con funciones de entropía $H(X)$ y $H(Y)$, tomando una muestra aleatoria de cada una de las variables, lo suficientemente grande, podemos formular como hipótesis nula la igualdad de las funciones: $H_0 : H(X) = H(Y)$ y como alternativa: $H_1 : H(X) \neq H(Y)$. En este contraste el estadístico es:

$$z = (\overset{*}{H}(X) - \overset{*}{H}(Y))/\sqrt{Var\overset{*}{H}(X) + Var(\overset{*}{Y})}$$

Que por lo expuesto anteriormente sigue una normal: $N(0,1)$. Evren et al (2015).

La hipótesis nula se rechaza si fijado un nivel de significación α se cumple que el valor observado del estadístico z cae fuera del intervalo $[z_{\alpha/2}, z_{1-\alpha/2}]$.

6) La función H es cóncava en p.

En el caso continuo, las propiedades de la función de entropía coinciden con el caso discreto, salvo la no negatividad . Por ejemplo, la función de entropía puede ser negativa, es el caso, si se considera una distribución uniforme $f_X(x;a) = 1/a$ se puede comprobar que para $a < 0$, la función de entropía es negativa. Shannon (1946 pág. 38) y Álvarez; P,G (1981 pág. 51-52)

Si X es una variable aleatoria distribuida normalmente con media μ y desviación estándar σ y Z es una variable aleatoria continua con la misma media y desviación estándar de X , entonces se cumple: $H(X) \geq H(Z)$.

Entre todas las funciones de distribuciones de variables aleatorias continuas cuyos valores están dentro de un intervalo cerrado $[a,b]$ la distribución uniforme tiene máxima entropía. Y entre todas las funciones de distribuciones continuas con igual media μ, la distribución exponencial presenta la máxima entropía.

En este caso de la distribución normal la función de entropía de X es:

$$H(X) = \log_{10} \sigma \sqrt{2\pi e}$$

La cantidad de información en el caso de una distribución normal bivariante, es : $I(X,Y) = H(X) + H(Y) - H(X,Y) = 1/2 \log_{10}(1-\rho^2)$. Lo que quiere decir que la información depende del cuadrado de la correlación entre las dos variables.

La interpretación de la función de entropía en el caso continuo también es diferente al caso discreto, para ello se puede consultar a Shannon (1948 pág. 37-38).

"There is one important difference between the continuous and discrete entropies. In the discrete case the entropy measures in an *absolute* way the randomness of the chance variable. In the continuous case the measurement is *relative to the coordinate system*…"
"The entropy for a continuous stochastic process has many properties analogous to that for discrete processes. In the discrete case the entropy was related to the logarithm of the *probability* of long sequences, and to the *number* of reasonably probable sequences of long length. In the continuous case it is related in a similar fashion to the logarithm of the *probability density* for a long series of samples, and the *volume* of reasonably high probability in the function space."

La función de divergencia o información relativa de Kullback se define como una medida de distancia entre dos distribuciones, donde una se puede tomar como la distribución teórica y la otra como una aproximación a la misma. Formalmente tenemos la siguiente definición:

Definición 2: Información relativa de Kulback-Leiber.

Dada dos variables aleatorias X e Y con funciones de probabilidades puntuales $P(X = x_i;\Theta)$ y $P(Y = y_i;\Theta)$, la medida de información relativa de Kulback-Leibler es:

$$D(X,Y) = \sum_{i=1}^{n} P(X = x_i;\Theta) \log_k \left(P(X = x_i;\Theta)/P(Y = y_i;\Theta)\right)$$

Se puede observar que la expresión anterior es el promedio ponderado de la diferencia logarítmica de la división entre las dos funciones de probabilidades puntuales de las variables X e Y, donde la ponderación está dada por la función de probabilidad $P(X = x_i;\Theta)$. De la misma forma podemos obtener la relación inversa:

$$D(Y,X) = \sum_{i=1}^{n} P(Y = y_i;\Theta)\log_k (P(Y = y_i;\Theta)/ P(X = x_i;\Theta))$$

En el caso que tenemos dos variables aleatorias continuas X e Y con funciones de densidades $f_X(x;\Theta)$ y $f_Y(y;\Theta)$ la información relativa es :

$$D(X;Y) = \int_{-\infty}^{\infty} f_X(x;\Theta)\log_k (f_X(x;\Theta)/ f_Y(y;\Theta))d\mu(x)$$

Se puede observar en ambos casos que esta función está relacionada con la función de entropía de Shannon, para el caso discreto:

$$D(X,Y) = \sum_{i=1}^{n} P(X = x_i;\Theta)\log_k P(X = x_i;\Theta) - \sum_{i=1}^{n} P(X = x_i;\Theta)\log_k P(Y = y_i;\Theta))$$

Luego, la función de entropía de Shannon y la divergencia de Kullback están relacionadas.

$$D(X,Y) = -H(X) - \sum_{i=1}^{n} P(X = x_i;\Theta)\log_k P(Y = y_i;\Theta)$$

En el caso continuo:

$$D(X;Y) = -H(X) - \int_{-\infty}^{\infty} f_X(x;\Theta)\log_k f_Y(y;\Theta)d\mu(x)$$

Algunas aplicaciones de la función de entropía de Shannon en Estadística.

Ahora mencionaremos tan solo tres aplicaciones de la función de entropía más conocidas en Estadística:

a) **Estimación de densidad de una variable aleatoria continua.** En estadística se emplea la función de entropía de Shannon entre otras cosas para encontrar estimadores de poblaciones de la que solo se tiene un conocimiento de algunos momentos y, si no se conocen, se pueden estimar con los datos muestrales. Por tanto es un método alternativo al método de estimación no paramétrico del núcleo (kerner). Supongamos que se desea estimar la función de densidad: $f_X(x;\Theta)$ de la variable aleatoria X. Entonces, la estimación de la función de densidad, utilizando la función de entropía, se obtiene al maximizar la función:

$$H(X) = -\int_{-\infty}^{\infty} f_X(x;\Theta) \log_k f_X(x;\Theta) dx$$

Sujeta a las condiciones:

$$\int_{-\infty}^{\infty} f_X(x;\Theta) dx = 1$$

$$\int_{-\infty}^{\infty} x^i f_X(x;\Theta) dx = \mu_i$$

Cuando se compara este método de estimación con el método de mínimos cuadrados, se encuentra que la varianza del estimador es menor, pero el estimador es sesgado, ver Cook L et al (2015), Macke J.H et al(2013). Para resolver el problema del máximo empleamos los multiplicadores de Lagrange:

$$L(X;\Theta) = -\int_{-\infty}^{\infty} f_X(x;\Theta) \log_k f_X(x;\Theta) dx + \lambda(\int_{-\infty}^{\infty} f_X(x;\Theta) - 1) + \sum_{i=1}^{k} \lambda_i (\int_{-\infty}^{\infty} x^i f_X(x;\Theta) dx - \mu_i)$$

En el caso de la divergencia el problema de optimización es minimizar la misma.

b) **Arboles de clasificación o de decisión**. Es un método que permite clasificar los valores observado de una muestra partiendo de unos datos previamente observados conocidos como ejemplo o patrón. La construcción de un árbol supone definir en primer lugar la raíz, que corresponde al atributo que más discrimina, bien sea como variable continua o como categoría, luego los atributos de los demás nodos, asignándoles los atributos de mayor a menor capacidad discriminatoria y finalmente las hojas que pueden ser booleanas. Las ramas en este árbol representan conjuntos de decisiones.

La función de entropía se emplea como criterio para seleccionar los atributos o variables más relevantes en un árbol de decisión. La razón fundamental es que ella es una buena medida para saber el grado de impureza de cada atributo, o lo que es lo mismo, la dificultad que tiene cada atributo para hacer una buena discriminación. La entropía toma el valor cero, si todos los ejemplos pertenecen a la misma clase en el atributo meta, es decir, el conjunto de ejemplo es totalmente homogéneo, en este caso se dice que es puro, por tanto bastaría ese solo atributo para construir el árbol. Entonces, en la medida que la entropía aumenta en el intervalo $[0, \log_a m]$ es mayor su impureza. La selección se jerarquiza en la medida que se gana información y por tanto disminuye la incertidumbre.

c) **Comparación entre test estadísticos**. Sea la hipótesis nula $H_0 : f_X(x;\Theta) \in F$ y alterna $H_1 : f_X(x;\Theta) \notin F$ siendo F una familia de distribuciones, y consideremos que existen dos estadísticos: θ_1^* y θ_2^* que permiten hacer el contraste de hipótesis, con sus respectivas distribuciones en el muestro. Si existen las respectivas funciones de entropías de ambos estadísticos para cualquier tamaño de muestra, entonces, es posible establecer una comparación entre ellas y determinar cual estadístico genera mayor información o incertidumbre mediante la comparación entre sus funciones de entropía.

II.-Comparación entre test no paramétricos.

En la práctica, el problema que generalmente interesa es: dadas dos variables aleatorias X e Y, en cuál de ellas, la cantidad de incertidumbre en una es mayor o menor que la otra, más que la indeterminación absoluta de cada una por separado.

Lo que se propone es, comparar la incertidumbre de dos sucesiones de estadísticos $\{\theta_n^1\}$ y $\{\theta_n^2\}$ siempre y cuando se cumpla las siguientes proposiciones:

1. Sean $\{\theta_n^{*1}\}$ y $\{\theta *_n^2\}$ dos sucesiones de estadísticos para un mismo contraste de hipótesis de muestras relacionadas, con distribución discreta en el muestro, entonces para cualquier tamaño de la muestra n existen las sucesiones de funciones de entropía $\{H(\theta_n^1)\}$ y $\{H(\theta_n^2)\}$

2. Ambas sucesiones de entropía son monótonas no decrecientes.

Para el caso de los estadísticos de los test paramétricos que permiten hacer contraste de posición o escalamiento, estos estadísticos suelen tener como su distribución en el muestreo distribuciones del tipo: t, F o χ^2 que no tienen funciones de entropía, salvo cuando su distribución es una normal. Por tanto, la comparación está limitada entre estadísticos de libre distribución.

Para continuar desarrollando la idea propuesta partimos del siguiente conjunto de proposiciones:

Proposición 1.

Sean X e Y dos variables aleatorias continuas y no negativas tales que sus funciones de distribución cumplen con la siguiente relación $F_X(x;\Theta) \leq F_Y(y;\Theta)$, si existen $E(X)$, y $E(Y)$, entonces $E(X) \geq E(Y)$

Demostración:

La esperanza matemática de una variable aleatoria continuas X puede definirse como: $\int_0^\infty P(X > t)dt + \int_{-\infty}^0 P(X \leq t)dt$, si la variable aleatoria es no negativa, entonces:

$E(X)=\int_0^\infty P(X>t)dt$, de acuerdo a esta definición tenemos las siguientes consecuencias: $F_X(x;\Theta)\leq F_Y(y;\Theta)\Rightarrow (1-P(X\ \ >x))\leq (1-P(Y>y))\Rightarrow E(X)\geq E(Y)$.

En el caso discreto tenemos: si $F_{X1}(x;\Theta)\leq F_{X2}(x;\Theta)$ entonces $P(X_1<X_2)\leq 1/2$ esto es, la mayoría de la primera población son mayores que los de la segunda, por tanto la mediana de la primera es mayor a la segunda, y si existen las esperanzas matemáticas entonces: $E(X\ )\geq E(Y)$

Proposición 2.

Consideremos el contraste de hipótesis, donde la hipótesis nula es: $H_0=\theta_0$ y la hipótesis alternativa $H_1=\theta_1$ y consideremos dos estadísticos no paramétricos para hacer el contraste: θ^{*1} y θ^{*2} que toman valores no negativos discretos, con distribuciones bajo la hipótesis nula $P(\theta^{*1}|H_0)$ y $P(\theta^{*2}|H_0)$ Si el primer estadístico es estocásticamente mayor al segundo, esto es: $P(\theta^{*1}|H_0)\leq P(\theta^{*2}|H_0)$, entonces: $H(\theta^{*1}|H_0)\geq H(\theta^{*2}|H_0)$

Demostración: Si $P(\theta^{*1}|H_0)\leq P(\theta^{*2}|H_0)$ entonces: por la proposición 1 tenemos

$$-\sum_{i=1} P(\theta^{*1}=x_i|H_0)\log_a P(\theta^{*1}=x_i|H_0)\geq -\sum_i P(\theta^{*2}=x_i|H_0)\log_a P(\theta^{*2}=x_i|H_0)$$

Luego: $H(\theta^{*1}|H_0)\geq H(\theta^{*2}|H_0)$.

Obsérvese que $\log P(\theta^{*i}|H_0)$ para i=1,2 es un valor negativo, por tanto no sería válido aplicar la proposición 1. Pero al cambiar de signo se hace positivo y la entropía es la esperanza matemática de $-\log P(\theta^{*i}|H_0)$. Si llamamos $Z_i=-\log P(\theta^{*i}|H_0)$ ahora $Z_i\geq 0$ para I=1,2 son positivas, y como por hipótesis $P(\theta^{*1}|H_0)\leq P(\theta^{*2}|H_0)$, aplicando la proposición 1 se tiene: $H(\theta^{*1}|H_0)\geq H(\theta^{*2}|H_0)$

Ahora, consideremos lo siguiente: intuitivamente se concluye que si se tiene dos sucesiones de eventos equiprobables donde uno de ellos tiene un número mayor de posibles observaciones, este tiene una mayor entropía. En efecto, consideremos las siguientes sucesiones de eventos: $\{e_i^1; i = 1,2...n\}$ y $\{e_i^2; i = 1,2...m\}$ con $n \geq m$; con probabilidades $\{p_i^1 = 1/n; i = 1,2...n\}$ y $\{p_i^2 = 1/n; i = 1,2...n\}$ entonces la entropía de la primera sucesión de eventos es mayor o igual a la entropía asociada a la segunda sucesión de eventos. En efectos, como ambas sucesiones de eventos son equiprobables, entonces: $H(e_i^1) = \log_a n$ y $H(e_i^2) = \log_a m$ como $n \geq m$, entonces: $H(e_i^1) \geq H(e_i^2)$.

Proposición 3.

Dada dos variables aleatorias discretas no negativas X y Y con diferentes distribuciones de probabilidad, el supremo de la entropía de una será mayor a la de la otra, si el número de valores que puede tomar una con respecto a la otra es mayor.

Demostración.

Consideremos las variables discretas X, Y con funciones de probabilidad puntual para X $P(X = x_0)$ con $x_0 = 1,2,...n$ y $P(Y = y_0)$ con $y_0 = 1,2,...m$ y $n \geq m$ entonces: $\max H(X) \leq \log_k n; \max H(Y) \leq \log_k m$, $\log_k n \geq \log_k m$ luego $\sup H(X) \geq \sup H(Y)$.

Comparación entre el test de los signos y el test de los signos de Wilcoxon.

Consideremos dos muestras relacionadas seleccionadas de una población bivariante (X,Y) continua y simétrica, con función de probabilidad: $F_{XY}(x,y;\Theta)$. Tomemos una muestra aleatoria de tamaño n $(X_i, Y_i); i = 1,2...n$ y consideremos los valores observados: $(x_i, y_i); i = 1,2,...n$ .Se puede asumir que los pares corresponden a unas unidades experimentales donde el primer valor corresponde antes del tratamiento y el segundo es la respuesta al tratamiento. En este caso, cada individuo o sujeto es su propio control. Por consiguiente, podemos definir una nueva variable aleatoria como. $Z = Y - X$. Sea M_d la mediana poblacional de las diferencias Z_i. Entonces, se propone como hipótesis nula que la mediana es nula, esto es:

$$\begin{aligned} H_0 &: M_d = 0 \\ H_1 &: M_d \neq 0 \end{aligned}.$$

Si consideramos el modelo: $Z_i = Y_i - X_i = \tau + \varepsilon_i$; donde τ es el efecto de tratamiento y ε_i es la perturbación aleatoria, el test es equivalente a decir que el efecto de tratamiento τ es nulo en el modelo: $Z_i = Y_i - X_i = \tau + \varepsilon_i$; esto es: $H_0 : \tau = 0$. Dos tests que son apropiados para hacer el contraste de hipótesis son el test de los signos y el test de los signos de los rangos de Wilcoxon. Estos son los test más usuales para realizar este contraste.

El estadístico en el test de los signos cuenta el número de diferencias positivas Z_i y lo denotamos como n^+. El estadístico de Wilcoxon es: $W = \sum_{i=1}^{n} R_i \delta_i$ donde R_i indica el rango dado al valor absoluto de Z, esto es: se ordenan $|Z_i|$ de menor a mayor y se le asigna el rango correspondiente y δ_i es una variable aleatoria que indica : $\delta_i = 1$ si Z_i >0 y $\delta_i = 0$ si Z_i <0.

Las funciones de probabilidades puntuales bajo la hipótesis nula de estos estadísticos son:

La distribución en el muestreo del test de los signos es:

$$P\left(n^+ = x_0 / H_0\right) = p^{x0}(1-p)^{n^*-x0} = \binom{n^*}{x_o} 0{,}5^n$$ y

La distribución en el muestreo del test de los signos de los rangos de Wilcoxon es:

$$P\left(W = W_0 / H_0\right) = f(W_0)/2^n$$

Donde $f(W_0)$, indica el número de ordenamientos $r_1 < r_2 < r_3 < ....r_n$ correspondientes a las observaciones de Z cuya suma es igual W_0.

Por tanto las funciones de entropías son:

$$H\left(n^+ / H_0\right) = -\sum_{i=1}^{n^*} \binom{n^*}{i} 0{,}5^n \log \binom{n^*}{i} 0{,}5^n$$

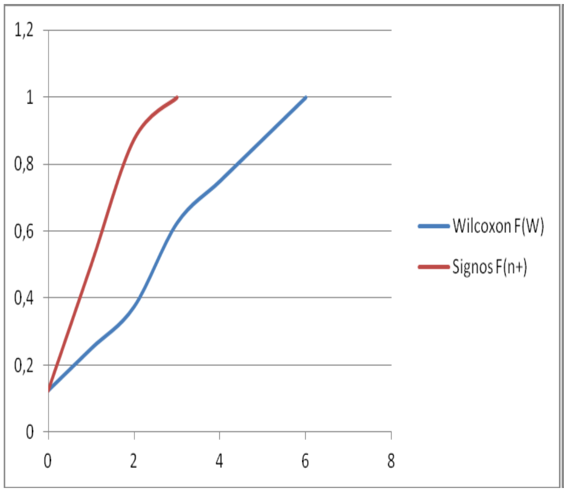
$$H\left(W / H_0\right) = -\sum_{W=0}^{n(n+1)/2} f(W_0)/2^n \log_k f(W_0)/2^n$$

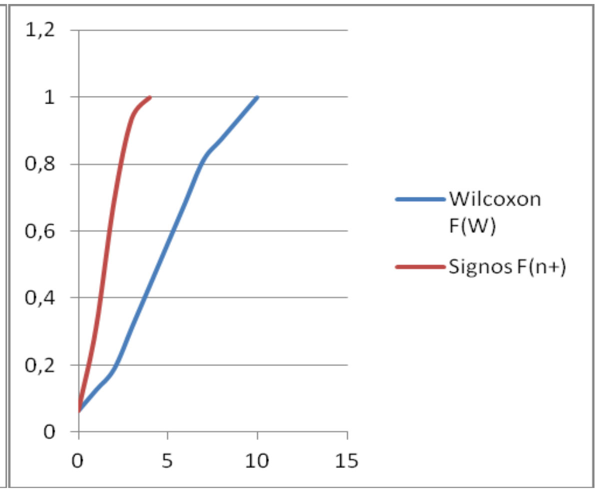
Para ilustrar la relación entre las dos funciones de entropía mostramos el siguiente resultado para tamaños de 3 y 4, donde se comparan primeramente las dos funciones de distribución:

n=3	Wilcoxon	Signos	n=4	Wilcoxon	Signos
X	F(W)	F(n+)	X	F(W)	F(n+)
0	0,125	0,125	0	0,0625	0,0625
1	0,25	0,5	1	0,125	0,3125
2	0,375	0,875	2	0,1875	0,6875
3	0,625	1	3	0,3125	0,9375
4	0,75		4	0,4375	1
5	0,875		5	0,5625	
6	1		6	0,6875	
			7	0,8125	
			8	0,875	
			9	0,9375	
			10	1	

Cuadro 1

Graficando las funciones de distribuciones para un mismo tamaño de muestra tenemos:





Comparación gráfica de las funciones de distribucióndе los test de los signos y Wilcoxon para tamaños de muestras n=3,4

La primera figura es la función de probabilidad bajo la hipótesis nula del estadístico de los signos y del Wilcoxon para un tamaño de n=3, y la segunda para n=4. Con estos gráficos se ilustra que $F_W(w;) \leq F_n(n^+)$. Si se calculan los valores esperados de cada distribución se encontrará que $E(W/H_0) \geq E(n^+/H_0)$ lo que coincide con lo expuesto en la proposición 1.

Ahora, para ilustrar la relación entre las dos funciones de entropía de los dos estadísticos, calculamos las respectivas funciones de entropía medidas en bit, partiendo las distribuciones bajo la hipótesis nula y considerando diferentes tamaños de muestras: de 3 hasta 7.

Cuadro 2.

n	3	4	5	6	7
H-Signo	1,81127812	2,03063906	2,19819241	2,33336203	2,44663975
H-Wilcoxon	2,75	3,375	3,85845859	4,25080724	4,58019367

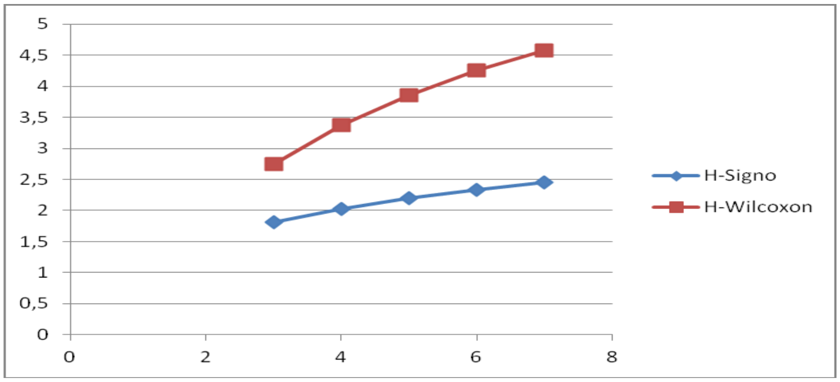


Figura 3. Comparación gráfica de las funciones de entropía de los test de los signos y Wilcoxon para tamaños de muestras n=3,…7

El cuadro 1 se puede observar que el número de valores que puede tomar la distribución de Wilcoxon es mayor a la distribución de los signos, en efecto el máximo valor que puede tomar Wilcoxon es: $n(n+1)/2$ mientras que de los signos es n, por la otra, en el cuadro 2 se observa que al comparar los valores de las de entropía de los dos estadísticos, la de Wilcoxon es mayor, cumpliéndose en este ejemplo, con las proposiciones 2 y 3.

Se puede demostrar que la relación $H(W) \geq H(n^+)$ siempre se cumple. Partiendo de sus respectivas distribuciones nulas

$$P(n^+ = x_0 | H_0) = \binom{n^*}{x_o} 0{,}5^n = \binom{n}{x_0} / 2^n \quad \text{y} \quad P(W = W | H_0) = f(W_0)/2^n$$

$$f(W)/2^n \log_a f(W_0)/2^n \leq \binom{n}{x} 0{,}5^n \log_a \binom{n}{x} 0{,}5^n$$

Considerando las proposiciones 1 y 2 obtenemos:

$$-\sum f(W)/2^n \log_a f(W_0)/2^n \geq -\sum \binom{n}{x} 0{,}5^n \log_a \binom{n}{x} 0{,}5^n$$

Luego $H(W/H_0) \geq H(n^+/H_0)$.

Observaciones finales.

Hemos propuesto la comparación entre dos test no paramétricos de locación para muestras relacionadas, en este caso bivariante, usando el concepto de entropía de Shannon. Sin embargo, es posible aplicar esta idea a otros test no paramétricos asociado a un mismo contraste de locación o de escala, donde las funciones de entropías existan. El problema que puede presentarse es la dificultad de cálculo que permita hacer esta comparación. En todo caso, siempre existe como herramienta la simulación Monte Carlo para realizar esta comparación.

REFERENCIA BIBLIOGRÁFICA.

Commenges D (2015)
Information Theory and Statistitics: an overview
Epidemiology Biostatistics Research Center. Bordeaux University

Cook L and Harslen PH (2015)
An Introduction to Entropy Estimation of Parameters in Economic Model.
Annual Conference on Global Economic Analysis Melbourne, June 17-19 2015

Cover, T. M; Joy A. Thomas (1991)
Elements of Information Theory-John Willey and Sons. Chapter 12.

Gibbons; J.D (1971)
Nonparametric Statitical Inference.
McGraw-Hill New York

Kullback, S; Keegel, C (1991)
Categorical Data Problems using information. Theoretic Approach.
Handbook of Statistics, Volume 4-Nonparametric Methods Editors: Krishnaiah, P.R; Sen.P.K. North –Holland. Amsterdam.

Kullback, S; (1997)
Information Theory and Statistics. Dover Publications.INC. New York

Macke; J,H. Murray; i. Latham; P. (2013)
Estimation Bias in Maximum Entropy Models.
Entropy 15, 3109-3219 August 2013

Liese; F and Vajda; I (2006)
On Divergences and Informations in Statitics and Information Theory
IEEE Transations on Information Theory Vol 52 N° 10 Octuber 2006.

Lucasz Debowsky (2013)
Information Theory and Stadistics
Institute Computation Science Polish Academy of Science

Pardo Llorente, P (1993)
Teoria de la Información Estadística.
Estadística Española Vol 35 N° 133 Págs.195 a 316)

Raynal; J. A (2008)
Comparación del Método del Principio de la Máxima Entropía en la Estimación del Parámetro de la Distribución de Valores Extremos Tipo I Tesis Doctoral
Universidad de la Américas, Puebla. Información Tecnológica Vol.19 N° 2 2008

Reyes A.E (1979)
Comentarios sobres algunos Métodos no Paramétricos. Trabajo de Ascenso para optar a la categoría de profesor Asistente. Universidad Central de Venezuela.

Shwarz D; M.G (2014).

Una Medida de la Incertidumbre basada en la Entropía para Cuantificar el Impacto de la Pérdida de Información de los Sistemas de Información.
Paideia XXI Vol. 4 N° 5 pag. 46-56

Shannon, C.S (1948)
A Mathematical Theory of Communication.
The Bell System Technical Jounal Vol 27 pag 379-423, 623-656 July. October 1948