EFICIENCIA RELATIVA ASINTÓTICA

Rubén Belmonte Coloma

Bajo el supuesto de tener una muestra aleatoria $X_1, X_2 \ldots\ldots X_n$ d e una variable continua con distribución acumulada F(X - θ) donde F(x)+F(-x) = 1 ∀x, es decir simétrica con respecto a θ. es posible estar interesados en pruebas tales como H_0: θ = 0 frente a la alternativa H_1: θ > 0. Los estadísticos B [1] de la prueba del signo y W^+ [2] del rango signado de Wilcoxon de carácter no paramétrico de distribución libre pueden resolver el problema. Sin embargo se desea saber cual es relativamente más eficiente.

En toda prueba estadística se debe asegurar que por un lado el procedimiento mantenga un nivel de significación por debajo de *α* preasignado y por otro que sea eficiente al detectar la hipótesis alternativa. Así si dos pruebas I, II tienen el mismo nivel de significancia y la prueba II tiene la función potencia superior a la función potencia de 1, para el parámetro que se está evaluando en la alternativa, entonces se prefiere la prueba II. Sin embargo en estadística no paramétrica la estructura de la distribución no es lo suficientemente rígida para obtener una teoría análoga a la que se usa en la teoría paramétrica en cuanto a generar un test unifórmenle más poderoso (lema de Neyman Pearson). En efecto una prueba no paramétrica rara vez es uniformemente poderosa comparada con otra.

En ausencia de una teoría óptima, se podría considerar una expresión para la función potencia de dos tests que compiten y usar propiedades relativas a ambos test. Pero, esta idea básica falla en muchas pruebas no paramctricos.

Al considerar algunos aspectos de las funciones de potencia de las pruebas del signo y Wilcoxon se ve que la función potencia para el test del signo su obtención es fácil, pero no necesariamente para el rango signado de la prueba de Wilcoxon.

Al considerar este ultimo se ve que depende del tamaño de muestra, los valores de la alternativa 0 y de la forma de la distribución y del nivel de significancia lo que demanda una tabulación muy extensa como la que se obtuvo para considerar la distribución uniforme, con *α*=0.05 con tres tamaños de muestra en una simulación de 1000 casos.

Cálculo empírico de la potencia en 100 corridas de datos *α*=0.05

Muestra	Estadístico	Distribución Uniforme (θ/σ)				
N = 10	W	99	133	177	474	681
	B	49	101	188	303	453
N = 15	W	51	163	283	642	852
	B	53	124	230	390	590
N = 20	W	50	205	479	768	935
	B	55	133	278	487	703

Como se puede apreciar generar una tabla empírica para la normal u otra distribución, considerar para otras alternativas de tamaño de muestra y finalmente considerar otros niveles de significación incrementaría el trabajo de la construcción de la tabla para tomar una decisión que puede no ser relevante para un análisis estadístico puntual.

Otra alternativa a considerar esta basada en las propiedades de distribuciones asintóticas Si bien el problema no es tan engorroso debido a que las distribuciones limites son continuas generalmente la comparación es fácil. Sin embargo una muestra de tamaño considerable también trae problemas. En el caso de la prueba del signo y en la del rango signado de Wilcoxon a mayor muestra la potencia cae muy rápidamente en I para cualquier valor de θ en la alternativa θ>0 lo que hace imposible comparar B y W para a fijo y θ>0.

La comparación asintótica que se generan con hipótesis alternativas inconsistentes es resuelta por Pitman (1948). Este investigador de la estadística no paramétrica para resolver el problema de la inconsistencia, propone considerar una secuencia de alternativas que convergen al valor de la hipótesis nula. Para escoger una

1 Sea $\Psi_i = \Psi(Xi - \theta)$ i=1,2,..,n donde Ψ(t)=1, si t>0 Ψ(t)=0, si t<0 $B = B(\Psi_i, \Psi_2, \ldots, \Psi_n) = \Sigma\Psi_i$

2 $W^* = \Sigma\Psi_i R_i^*$ donde R_i^* el rango absoluto de Z_i

secuencia útil para la prueba el comportamiento de la prueba no tiene que degenerarse. Desde este punto de vista, esta comparación no depende de la significancia y de esta manera se puede escoger asintoticamente uno de las pruebas que dependen solo de las formas de la distribución.

Se formalizan estos conceptos de la siguiente manera: Consideremos el problema de probar una hipótesis nula H_0:θ ε ω contra la alternativa H_1:θ ε Ω-ω donde θ es un parámetro conocido del espacio paramétrico Ω.

Sea {S_n} {T_n} sucesiones de estadísticos para probar H_0 fuente a H, donde el indica corresponde a las observaciones n y n ' es festivamente usados para la prueba sea _* mi elemento en el espacio alternativo_-w y denotemos por Cn y Dn' las regiones críticas de tamaño α, o < α <1 para las pruebas S_n y T_n respectivamente

$$P_\theta\ (\ S_n\ \varepsilon\ C_n\ ) = \alpha \qquad \forall \theta \varepsilon \overline{\omega}\ \ n$$

$$P_\theta\ (\ T_{n'}\ \varepsilon\ D_{n'}\ ) = \alpha \qquad \forall \theta \varepsilon \overline{\omega}\ \ n'$$

Sea β α<β<l arbitrario pero fijo. Se define N y N' el menor valor de n y n' respectivamente para el cuál $P_\theta\ (\ S_n\ \varepsilon\ C_n\ ) \geq \beta \quad \forall \theta \varepsilon \overline{\omega}\ \ n \quad P_\theta\ (\ T_{n'}\ \varepsilon\ D_{n'}\ ) \geq \beta \quad \forall \theta \varepsilon \overline{\omega}\ \ n'$

De esta manera N y N' son el número mínimo de observaciones par los cuales el nivel α basado en S_n $T_{n'}$ tiene una potencia de β bajo la alternativa θ*.

En realidad N es una función N (α, β, θ* y la distribución) y N' también N' (α, β, θ* y la distribución).

Se define entonces (Eficiencia Relativa para muestra finita) e (S T/ α, β, θ* y la distribución) = N' (α, β, θ* y la distribución) / N (α, β, θ* y la distribución) a la que llamaremos eficiencia relativa basada en el tamaño de la muestra n. Si e (S T/ α, β, θ* y la distribución)>1 S_n es más eficiente que T_n

Se debe comprender que e (S T/ α, β, θ* y la distribución) es local, y lograr esta conformación es muy dificultosa.

Una definición que se adecúa más a la construcción de la eficiencia se da formalmente de la siguiente manera:

Sean {S_n} y {T_n} dos sucesiones de los estadísticos de prueba para probar H_0: θ = θ_0, para algún valor especifico del p a r á m e t r o contra una clase de alternativas H_1 de tamaño α y sea {θ_1} una secuencia de alternativas H_0 tal que el $\lim\limits_{t \to \infty} \theta_i = \theta_0$ y además β $_{Sn}$ (θ_i) y β $_{Tn}$ (θ_i) son las potencias de las pruebas basadas en S_n y T_n respectivamente para la alternativa θ_i. Sea {n_i} y {n_i'} sucesiones crecientes de enteros positivos tales que las dos secuencias de pruebas tienen el mismo limite de nivel α y

$\alpha < \lim \beta_{Sn}(\theta_i) = \lim \beta_{Tn}(\theta_i) < 1$

Entonces la **eficiencia relativa asintótica (ARE)** de {S_n} relativo a {T_n} es

$$ARE(S,T) = \lim_{i \to \infty} n_i'/n_i$$

Referencias

Randles R.H. Wolfe D.A. (1991): Introduction to he theory of Nonparametric Statistis Krieger; Publishing Company Florida