李 聖 鍵 敎授指導

碩士學位 請求論文

비모수적 순위 검정 통계량의 불편성에 관한 연구

2007

誠信女子大學校 大學院

統計學科

李 佳 英

비모수적 순위 검정 통계량의
불편성에 관한 연구

李 聖 鍵 敎授指導

이 論文을 碩士學位 論文으로 提出함

*2007*年 *5*月

誠信女子大學校 大學院

統計學科

李 佳 英

認　准　書

李佳英의 碩士學位 論文으로 認准함.

審査委員________________________印

審査委員________________________印

審査委員________________________印

誠信女子大學校 大學院

논 문 개 요

Lehmann(1959,1986)은 위치모수에 대한 단측 윌콕슨 순위합 검정이 편의가 없음을 보였으나 양측 검정에 편의(bias)가 존재하는 것에 대한 의문을 제기하였다. 그 후 Sugiura(1965)는 연속확률함수인 지수함수로부터 두 개의 확률표본을 추출하여 두 확률분포의 위치모수가 동일한 것인가에 대해 윌콕슨 순위합 검정을 실시하여 표본의 크기가 각각 다를 때 검정력 함수는 양측 검정이 편의가 있음을 보였다. 그러나 표본의 크기가 같을 때에는 편의가 있음을 밝히지 못하였다. 이에 Sugiura, et al.(2006)은 표본크기가 같을 때 양측 윌콕슨 순위합 검정이 편의가 있음을 보여주었다. 본 논문은 두 표본크기가 2인 경우 편의가 있음을 보여주는 추가적인 반례를 제시하고 척도모수에 관한 비모수 검정 방법인 Ansari-Bradley 검정이 편의가 있음을 반례(counter example)를 제시하여 증명하였다.

목　　　차

제1장 서 론

모수적인 통계방법은 모집단의 가정을 전제로 한다. 그러나 연구자들이 통계적인 기법을 사용할 때 정규성을 가정하고 연구하는 경우가 많다. 그러나 모수적 통계기법 결과는 주어진 가정을 만족해야 유용하다. 그래서 모집단의 가정에 의존하지 않는 비모수 통계학이 등장하게 된다. 비모수 통계학은 확률표본의 독립성 등의 제한적인 가정만을 요구하기 때문에 여러 분야에서 다양하게 응용될 수 있다. 또한 비모수통계학은 가정 없이 통계적 추론을 하므로 상당히 일반적이고, 양적 관측값으로 표현할 수 없는 경우 순위를 사용함으로서 실제적이며, 특이점이 있는 경우 모수적 방법보다 검정력의 손실을 피할 수 있는 장점이 있다(이승우, 2003).

그런데 Lehmann(1959,1986)은 위치모수에 대한 단측 윌콕슨 순위합 검정이 편의가 없음을 보였으나 위치모수에 대한 양측 윌콕슨 순위합 검정의 편의가 존재하는 것에 대한 의문을 제기하였다. 그 후 Sugiura(1965)는 지수함수로부터 두 개의 확률표본을 추출하여 두 확률분포의 위치모수가 동일한 것인가에 대해 윌콕슨 순위합 검정을 실시하였다. 그 결과 표본의 크기가 각각 다를 때 검정력 함수는 편의가 있음을 보였다. 그러나 표본의 크기가 같을 때에는 편의가 있음을 밝히지 못하였다. 이에 Sugiura, et al.(2006)은 표본크기가 같을 때 양측 윌콕슨 순위합 검정에 편의가 있음을 보여주었다. 그러나 이 연구는 표본크기가 2인 경우에 국한된 결과로써 추가적인 연구가 필요하다 할 수 있다. 이에 본 논문은 위치모수 검정과 함께 척도모수 검정 통계량에 대한 편의의 유무를 살펴보고자 한다.

제2장 비모수적 방법

일반적으로 정규성의 가정을 만족하지 않을 때 비모수적인 방법을 이용하여 자료를 분석한다. 비모수적인 방법에서 위치 모수를 검정하기 위한 대표적인 방법에는 윌콕슨 순위합 검정이 있고 척도 모수를 검정하기 위한 대표적인 방법에는 앤사리-브래들리 검정이 있다.

2.1 윌콕슨 순위합 검정 (Wilcoxon rank sum test)

Wilcoxon(1945)과 Mann-Whitney(1947)에 의하여 제안된 두 표본 위치 문제를 위한 대표적인 비모수 검정기법을 맨-휘트니(Mann-Whitney)검정이라 한다. 이 검정기법은 일반적으로 윌콕슨 순위합 검정, 맨-휘트니 윌콕슨 검정, 윌콕슨-맨-휘트니 검정이라 칭하기도 한다. 맨-휘트니 순위합 검정은 두 모집단 간의 모평균의 차이 유무를 검정하기 위한 대표적인 비모수적 검정기법으로 두 모집단의 정규성, 분산의 동일성 및 확률표본의 독립성의 가정을 전제로 하는 모수적 검정기법인 두표본 t검정의 대안으로 간주할 수 있다. 즉 맨-휘트니 순위합 검정은 두 모집단의 정규성 및 분산의 동일성 가정을 요구하지 않으며, 단지 확률표본의 독립성만을 필요로 하는 상당히 뛰어난 검정기법이다(최영훈, 2004). 윌콕슨의 부호 순위 검정은 관측값이 귀무가설 하에서 위치모수의 값 θ_0보다 크고 작음뿐만 아니라 관측 값의 상대적인 크기도 고려하여 검정을 실시한다.

2.1.1 검정 통계량

두 모집단으로부터 각각의 확률표본 X, Y를 추출한다. X와 Y의 혼합표본의 순위를 R_j라고 한다. 윌콕슨 순위합 통계량을 다음과 같이 나타낸다.

$$W=\sum_{i=1}^{n} R_j$$

통계량 W는 Y표본에 부여된 순위의 합을 나타낸다.

2.1.2 검정법

θ_X와 θ_Y는 각각 X와 Y의 위치모수이고, $\Delta=\theta_X-\theta_Y$이다.
이 때, 귀무가설은 $H_o:\Delta=0$이다.

유의수준 α에서
(i) $H_1:\Delta>0$일 때 $W\geq w(\alpha,\ m,\ n)$이면 H_0 기각
(ii) $H_1:\Delta<0$일 때 $W\leq w(1-\alpha,\ m,\ n)$이면 H_0 기각
(iii) $H_1:\Delta\neq 0$일 때 $W\geq w(\frac{\alpha}{2},\ m,\ n)$ 또는 $W\leq w(1-\frac{\alpha}{2},\ m,\ n)$
이면 H_0 기각한다.
여기서 m과 n은 각각 X와 Y의 표본크기이다.

귀무가설 $H_o:\Delta=0$하에서 윌콕슨 순위합 검정 통계량 W는 그 평균 $\frac{n(m+n+1)}{2}$에 대하여 대칭인 분포를 갖는다.

따라서 $w(\alpha, m, n)-\frac{n(m+n+1)}{2}=\frac{n(m+n+1)}{2}-w(1-\alpha,\ m,\ n)$임 을 알 수

있고 이 식으로부터 다음의 관계식이 성립한다.

$$w(1-\alpha,\ m,\ n)\ =\ n(m+n+1)-w(\alpha,\ m,\ n).$$

0이 아닌 상수 값 Δ_0에 대해 가설 $H_0 : \Delta=\Delta_0$을 검정하고자 하는 경우를 고려해 보자. 이때는 수정된 Y의 관측 값 $Y_j'=Y_j-\Delta_0,\ j=1,\cdots,n$ 을 구한 다음 X와 Y_j'의 관측 값을 이용하여 윌콕슨 순위합 검정의 절차에 따라 검정을 실시한다. 만약 혼합표본에서 X_i의 순위를 S_i라 하고 X관측 값들의 순위합을 $W'=\sum_{i=1}^{n} S_i$로 표현하면 $W'=\frac{(m+n)(m+n+1)}{2}-W$ 임을 알 수 있다.

2.2 앤사리-브래들리 검정(Ansari-Bradley test)

모집단의 위치모수의 추론과 함께 척도모수의 추론을 생각해 볼 수 있다. 두 모집단의 위치 모수차가 같은 경우에는 척도모수가 혼합표본의 순위에 절대적인 영향을 주게 된다. 이에 여러 종류의 분포무관 검정법이 있으며 대표적으로 앤사리-브래들리 검정(Ansari-Bradley test)이 있다.

2.2.1 검정 통계량

X와 Y의 표본의 크기가 각각 m과 n이라 할 때, $N=m+n$이라 하자. N개의 관측 값을 작은 값부터 크기순으로 나열하고 최소값과 최댓값에 순위 1을 주고 두 번째로 작은 값과 큰 값에 순위 2를 세 번째로 작은 값과 큰 값에 순위 3을 준다. S_i은 X_i에 주어진 순위라 하면 앤사리-브래들리 통계량은 다음과 같이 정의한다.

$$T_{AB} = \sum_{i=1}^{m} S_i.$$

2.2.2 검정법

θ_X와 θ_Y는 각각 X와 Y의 위치모수이고 σ_X와 σ_Y는 X와 Y의 표준편차라고 할 때, $\gamma=\sigma_Y/\sigma_X$라고 하자. 귀무가설은 $H_0:\gamma=1$이다.

유의수준 α에서, $\gamma = \sigma_Y/\sigma_X$ 에 대하여

(i) $H_1:\gamma^2>1$일 때 $T_{AB}\geq t_{AB}(\alpha,\ m,\ n)$이면 H_0은 기각

(ii) $H_1:\gamma^2<1$일 때 $T_{AB}\leq t_{AB}(1-\alpha,\ m,\ n)$이면 H_0은 기각

(iii) $H_1:\gamma^2\neq 1$일 때 $T_{AB}\geq t_{AB}\ (\frac{\alpha}{2},\ m,\ n)$ 또는 $T_{AB}\leq t_{AB}(1-\frac{\alpha}{2},\ m,\ n)$이면 H_0은 기각이다. $t_{AB}(\alpha,\ m,\ n)$ 는 표본의 크기가 $m,\ n$일 때 H_0하에서 T_{AB} 의 상위 $100\cdot\alpha$ 백분위수를 나타낸다. 즉 $P_0\{T_{AB}\geq t_{AB}(\alpha,\ m,\ n)\}=\alpha$를 만족하는 상수이다.

크기 순서대로 나열된 혼합표본에서 Y_i들의 값이 넓게 퍼져 있으면

X_i들은 상대적으로 가운데에 위치하여 S_i는 큰 값을 가지는 경향이 있다. 앤사리-브래들리 검정법은 $\theta_Y - \theta_X$을 알고 있다고 가정한다. 만약 $\theta_Y - \theta_X$의 값이 알려지지 않은 경우에는 $\Delta = \theta_Y - \theta_X$의 추정량을 이용하는 방법이 제안되었다.

제3장 비모수 검정통계량의 편의

3.1 윌콕슨 검정 통계량의 편의

3.1.1 두 표본의 크기가 다를 때 윌콕슨 검정 통계량의 편의

Sugiura(1965)는 두 표본에서 양측 윌콕슨 순위합 검정이 표본크기가 각각 다른 경우 다음과 같은 편의가 있음을 보였다.
여기서 불편검정법(unbiased test)을 다음과 같이 정의한다.

$$\max\{\pi(\theta) : \theta \in \Omega_0\} \le \min\{\pi(\theta) : \theta \in \Omega - \Omega_0\}$$

$G(x)$를 $F(x-\Delta)$라고 할 때, $F(x)$의 표본을 $X_1, X_2, ..., X_m$으로 $G(x)$의 표본을 $Y_1, Y_2, ..., Y_n$이라 한다. 귀무가설 $H_o : \Delta = 0$과 대립가설은 $H_1 : \Delta \neq 0$이다.
가설의 검정하기 위해 다음과 같은 검정함수(test function)를 정의한다.

$$\phi_1(X_1, \ldots, X_m; Y_1, \ldots, Y_n) = \begin{cases} 1, X_{(m)} < Y_{(1)} \text{ 또는 } Y_{(n)} < X_{(1)}, \\ 0, \text{그렇지 않으면} \end{cases} \tag{1}$$

검정의 유의수준은 $\alpha = 2(m!\,n!)/(m+n)!$ 이고 이 검정력 함수(power function)는 다음과 같다.

$$\begin{aligned} \beta &= P\{X_1, \ldots, X_m < Y_1, \ldots, Y_n\} + P\{X_1, \ldots, X_m > Y_1, \ldots, Y_n\} \\ &= \int_{-\infty}^{\infty} F(x)^m d[1-(1-G(x))^n]dx + \int_{-\infty}^{\infty} G(x)^n d[1-(1-F(x))^m]dx \end{aligned} \tag{2}$$

$$=n\int_{-\infty}^{\infty}F(x)^m(1-G(x))^{n-1}dG(x)+m\int_{-\infty}^{\infty}G(x)^n(1-F(x))^{m-1}dF(x).$$

$F(x)$와 $G(x)$를 다음과 같이 정의 하자.

$$F(x)=\begin{cases}1-e^{-x}, & x\ge 0\\ 0, & x<0,\end{cases}$$

$$G(x)=F(x-\Delta)=\begin{cases}1-e^{-(x-\Delta)}, & x\ge 0\\ 0, & x<0.\end{cases} \tag{3}$$

$\Delta\ge 0$인 경우에 검정력 함수(power function)는 다음과 같다.

$$\begin{aligned}\beta_{m,n}(\Delta)&=n\int_{\Delta}^{\infty}(1-e^{-x})^m e^{-n(x-\Delta)}dx+m\int_{\Delta}^{\infty}(1-e^{-(x-\Delta)})^n e^{-mx}dx\\ &=\frac{m!n!}{(m+n)!}e^{-m\Delta}+n\int_0^1 x^{n-1}(1-e^{-\Delta}x)^m dx.\end{aligned} \tag{4}$$

$\Delta<0$인 경우에 검정력 함수(power function)는 다음과 같다.

$$\frac{m!n!}{(m+n)!}e^{n\Delta}+n\int_0^1 x^{m-1}(1-e^{\Delta}x)^n dx. \tag{5}$$

(4)과 (5)식을 통해 $\beta_{m,n}(\Delta)=\beta_{n,m}(-\Delta)$임을 알 수 있다.

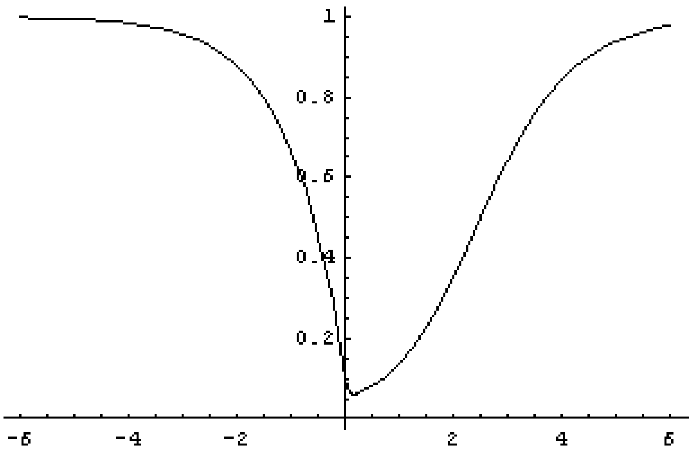
$m=19$, $n=1$, $\alpha=0.10$인 경우와 $m=n=3$, $\alpha=0.10$인 경우의 각각의 검정력 함수(power function)는 다음과 같다.

(i) $m=19,\ n=1,\ \alpha=0.10$

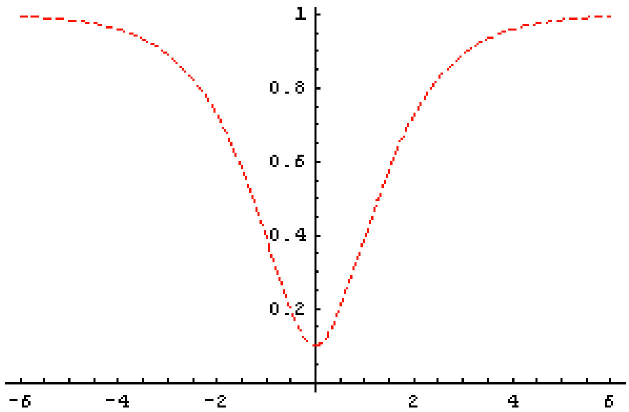
$$\beta_{19,1}=\begin{cases}\dfrac{1}{20}[e^{-19\Delta}+e^{\Delta}(1-(1-e^{-\Delta})^{20})], & \Delta\geq 0\\ 1-\dfrac{9}{10}e^{\Delta}, & \Delta<0.\end{cases} \tag{6}$$

(ii) $m=n=3,\ \alpha=0.10$

$$\beta_{3,3}(\Delta)=\begin{cases}1-\dfrac{9}{4}e^{-\Delta}+\dfrac{9}{5}e^{-2\Delta}-\dfrac{9}{20}e^{-3\Delta}, & \Delta\geq 0\\ 1-\dfrac{9}{4}e^{\Delta}+\dfrac{9}{5}e^{2\Delta}-\dfrac{9}{20}e^{3\Delta}, & \Delta<0.\end{cases} \tag{7}$$



[그림 1] $m=19, n=1$일 때 검정력 함수



[그림 2] $m=n=3$일 때 검정력 함수

$m=19,\ n=1$의 검정력 함수 [그림1]을 보면 검정력 함수가 $\Delta=0$에서 대칭이 아니므로 편의가 있음을 알 수 있다. 그러나 $m=n=3$인 검정력 함수 [그림2]는 $\Delta=0$에서 최소값이며, 좌우대칭이므로 편의가 없다.

3.1.2 두 표본 크기가 같을 때 윌콕슨 검정 통계량의 편의

Sugiura, et al.(2006)은 표본크기가 같은 경우에 다음과 같이 편의를 보였다. $F(x)$와 $G(x)$로부터 확률표본 $X_1, X_2, \ldots, X_n$ 와 $Y_1, Y_2, \ldots, Y_n$에 대해 H_0: $\Delta=0$ vs. $H_1 : \Delta \neq 0$을 검정한다. $X_{(1)} < \cdots < X_{(n)}$는 X_s의 순서통계량이며, $Y_{(1)} < \cdots < Y_{(n)}$ 또한 Y_s의 순서통계량이다. 양측 윌콕슨 검정의 검정함수(test function)를 각각 ϕ_1, ϕ_2로 정의한다.

$$\phi_1(X_1, \ldots, X_n; Y_1, \ldots, Y_n) = \begin{cases} 1, & X_{(n)} < Y_{(1)} \text{ 또는 } Y_{(n)} < X_{(1)} \\ 0, & \text{그렇지 않으면} \end{cases} \tag{8}$$

$$\phi_2(X_1,...,X_n;Y_1,...,Y_n)=\begin{cases}1\begin{cases}X_n < Y_{(1)}\text{또는 } Y_{(n)} < X_{(1)}\\ X_{(n-1)} < Y_{(1)} < X_{(n)} < Y_{(2)}\\ Y_{(n-1)} < X_{(1)} < Y_{(n)} < X_{(2)}\end{cases}\\ 0,\text{그렇지않으면.}\end{cases} \tag{9}$$

ϕ_1(test function)의 검정력 함수(power function)는 다음과 같다.

$$\begin{aligned}\beta_1(\Delta) &= P(X_{(n)} < Y_1,...,Y_n)+P(Y_1,...,Y_n < X_{(1)})\\ &= n\int_{-\infty}^{\infty}(1-G(x))^n F(x)^{n-1}f(x)dx\\ &+n\int_{-\infty}^{\infty}G(x)^n(1-F(x))^{n-1}f(x)dx.\end{aligned} \tag{10}$$

그리고 ϕ_2(test function)의 검정력 함수(power function)는 다음과 같다.

$$\begin{aligned}\beta_2(\Delta) &= \beta_1(\Delta)+nP(X_{(n-1)} < Y_1 < X_{(n)} < Y_2,....,Y_n)\\ &+nP(Y_1,....,Y_{n-1} < X_{(1)} < Y_n < X_{(2)})\end{aligned} \tag{11}$$

$$\begin{aligned}&=\beta_1(\Delta)+n^2(n-1)\iint_{-\infty<x<y<\infty}(G(y)-G(x))(1-G(y))^{n-1}\\ &\times F(x)^{n-2}f(x)f(y)dxdy+n^2(n-1)\\ &\times\iint_{-\infty<x<y<\infty}G(x)^{n-1}(G(y)-G(x))(1-F(y))^{n-2}f(x)f(y)dxdy.\end{aligned} \tag{12}$$

정리 1 검정력 함수(power function)는 Δ=0에 대하여 대칭이다 $\beta_1(\Delta)=\beta_1(-\Delta)$ 그리고 $\beta_2(\Delta)=\beta_2(-\Delta)$

(증명)

가정) (i) $X_1,\ldots,X_n$ 와 $Z_1,\ldots,Z_n$은 모두 독립이다.

(ii) $Z_1,\ldots,Z_n$는 $X_1,\ldots,X_n$와 같은 분포를 가진다.

(iii) $(X_{(1)},Z_{(n)})$ 와 $(Z_{(1)},X_{(n)})$ 는 교환가능하며, 마찬가지로 ($X_{(1)}$, $X_{(2)},Z_{(n-1)}$, $Z_{(n)}$) 와 ($Z_{(1)}$, $Z_{(2)}$, $X_{(n-1)}$, $X_{(n)}$)은 교환가능 하다.

모든 $i=1,\ldots,n$ 에 대해 $Y_i-\Delta$=Z_i라 두면 $Z_{(i)}$는 순서통계량이다.

식(10)식에 $Y_i-\triangle$=Z_i 을 대입하면 다음과 같이 표현된다.

$$\begin{aligned}\beta_1(\Delta)&=P(X_{(n)}<Z_{(1)}+\Delta)+P(Z_{(n)}+\Delta<X_{(1)})\\&=P(X_{(n)}<Z_{(1)}+\Delta)+P(X_{(n)}+\Delta<Z_{(1)}).\end{aligned} \tag{13}$$

마찬가지로 식(11)에 $Y_i-\Delta$=Z_i 을 대입하면 다음과 같이 표현된다.

$$\begin{aligned}\beta_1(\Delta)&=P(X_{(n-1)}<Z_{(1)}+\Delta<X_{(n)}<Z_{(2)}+\Delta)\\&\quad+P(X_{(n-1)}<Z_{(1)}-\Delta<X_{(n)}<Z_{(2)}-\Delta).\end{aligned} \tag{14}$$

$\beta_1(\Delta)$은 식(13)과 같고, $\beta_1(-\Delta)$은 다음과 같이 표현된다.

$$\beta_1(-\Delta)=P(X_{(n)}<Z_{(1)}-\Delta)+P(Z_{(n)}-\Delta<X_{(1)})$$

$$= P(Z_{(n)} < X_{(1)} - \Delta) + P(X_{(n)} - \Delta < Z_{(1)})$$

$$= P(X_{(n)} + \Delta < Z_{(1)}) + P(X_{(n)} < Z_{(1)} + \Delta). \tag{15}$$

$$\therefore \beta_1(\Delta) = \beta_1(-\Delta)$$

비슷한 방법으로 $\beta_2(\Delta) = \beta_2(-\Delta)$을 증명하면 다음과 같다. $\beta_2(\Delta)$는 (14)와 같고, $\beta_2(-\Delta)$는 다음과 같이 표현된다.

$$\beta_2(-\Delta) = \beta_1(-\Delta) + P(X_{(n-1)} < Z_{(1)} - \triangle < X_{(n)} < Z_{(2)} - \triangle)$$

$$+ P(X_{(n-1)} < Z_{(1)} + \triangle < X_{(n)} < Z_{(2)} + \triangle)$$

$$\therefore \beta_1(\Delta) = \beta_1(-\Delta). \qquad \square \tag{16}$$

정리2. 검정력 함수(power function) $\beta_1(\Delta)$의 도함수는 적분기호 하에 미분법에 의해 얻어질 수 있다고 가정하자. 그 때 검정 ϕ_1(test function)은 분포들의 위치 모수족에 대해 편의가 없다. 가정으로부터 우리는 적분기호 하에 미분할 수 있다.

$$\frac{d}{d\Delta}\beta_1(\Delta) = n^2 \int_{-\infty}^{\infty} f(x)g(x)[(1-(G(x))F(x)]^{n-1} - [G(x)(1-F(x))]^{n-1}dx$$

$$= n^2 \int_{-\infty}^{\infty} f(x)g(x)(F(x)-(G(x))$$

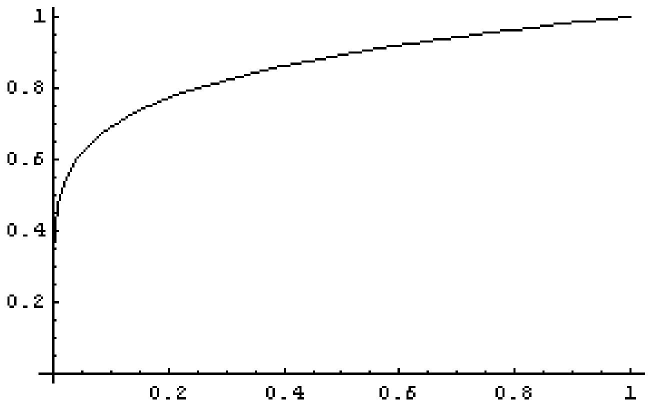
$$\times[(1-G(x))F(x)]^{n-2} + \dots + G(x)(1-F(x))]^{n-2}dx. \tag{17}$$

여기서 $\Delta > 0$이면 $F(x) - G(x) = F(x) - F(x-\Delta)$는 음이 아니며 $\Delta < 0$이면

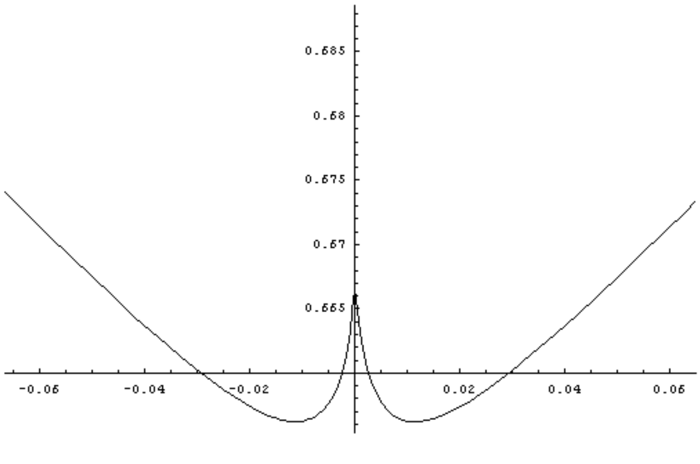
양이 아니다. □

$F(x)$와 $f(x)$를 다음과 같이 정의하자.

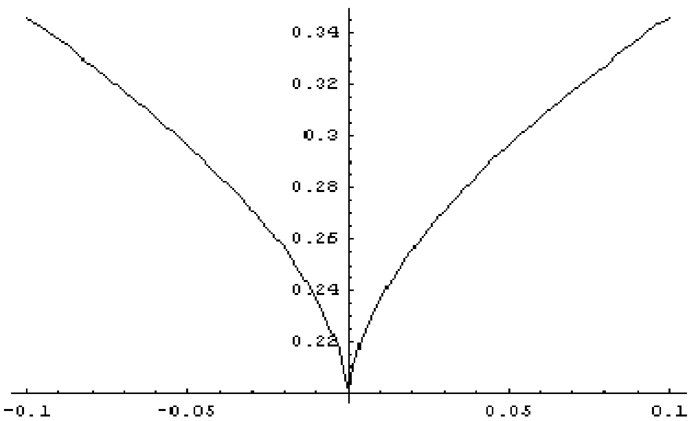
$$F(x)=\begin{cases}0, & x<0\\ x^{\rho}, & 0\le x\le 1\\ 1, & x>1\end{cases} \text{ 이고 } f(x)=\begin{cases}\rho x^{\rho-1}, & \text{만약} 0<x<1\\ 0, & \text{그렇지않으면}\end{cases}. \tag{18}$$



[그림 3] $F(x)$의 그래프



[그림 4] 표본크기가 2일 때 검정력 함수, $\rho = 0.16$



[그림 5] 표본크기가 3일 때 검정력 함수, $\rho = 0.16$

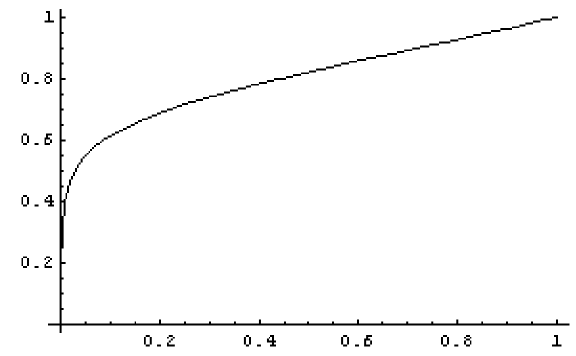
$n=2$일 때 검정력 함수인 [그림 4]는 유의수준이 2/3이다. $\Delta=0.01$일 때 검정력 함수 $\beta_2(0.01)=0.655287$이므로 편의가 있다. Sugiura, et al.(2006)논문에서 $\rho=0.7$이하이면 편의가 있고 $\rho=0.7$이상이면 편의가 없음을 알 수 있었다. 그리고 $n=3$일 때 검정력 함수인 [그림 5]는 $\Delta=0$에서 최소이고, $\Delta=0$에 대해 대칭이므로 편의가 없다.

3.1.3 윌콕슨 검정 통계량의 편의에 대한 추가적인 반례

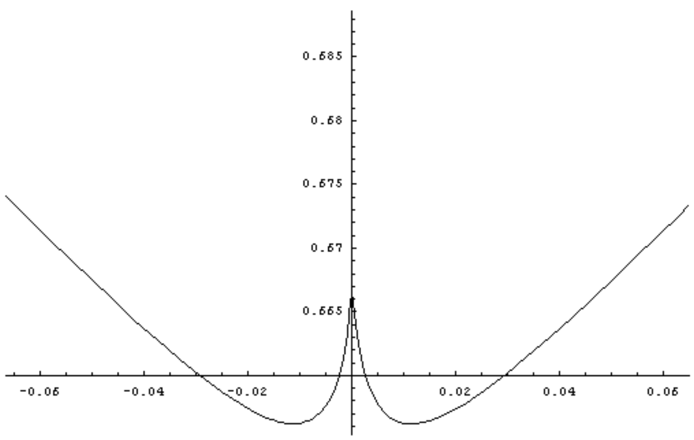
본 절에서는 함수가 다음과 같을 때 윌콕슨 검정 통계량이 편의가 있는 추가적인 반례를 제시하고자 한다.

$$F(x)=\begin{cases}0 & ,\ x<0 \\ \dfrac{\log_{10}(9999x+1)+2^x-1}{5} & ,\ 0\le x\le 1 \\ 1 & ,\ x>1\end{cases} .$$

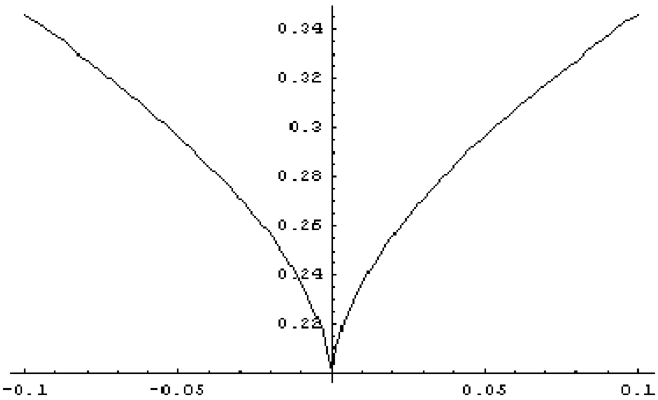
$$G(x)=F(X-\Delta) \tag{19}$$



[그림 6] $F(x)$의 그래프



[그림 7] 표본크기가 2일 때 검정력 함수



[그림 8] 표본크기가 3일 때 검정력 함수

$n=2$일 때 검정력 함수인 [그림 7]는 유의수준이 2/3이다. $\Delta=0.01$일 때 검정력 함수 $\beta_2(0.01)=0.655287$이므로 편의가 있다. 그리고 $n=3$일 때 검정력 함수인 [그림 8]은 $\Delta=0$에서 최소이고, $\Delta=0$에 대해 대칭이므로 편의가 없다. 표본크기가 3일 때 편의를 찾기 위해 다음과 같은 여러 가지 확률분포의 검정력 함수를 고려하였으나 편의가 있음을 밝히지 못했다.

$$F(x) = \log_2(x+1) \ , \ 0 \le x \le 1$$

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt, \quad -\infty \le x \le \infty$$

$$F(x) = 0.5\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{(t+5)^2}{2}} dt + 0.5\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{(t-5)^2}{2}} dt,$$

$$-\infty \le x \le \infty$$

$$F(x) = \frac{\log_{10}(99x+1)}{0.49875311720698257 \times (1+\log_{10}(999x+1)+0.01^x)}, \quad 0 \le x \le 1$$

$$F(x) = \frac{2\,Tan(0.3\pi x) + 10\log_{10}(99x+1)}{22.752763840942347}, \quad 0 \le x \le 1$$

$$F(x) = \frac{x^{200} + \log_{10}(999x+1)}{4}, \quad 0 \le x \le 1$$

$$F(x) = \frac{x^{20} + 0.9^x \times \log_{10}(999x+1)}{3.7}, \quad 0 \le x \le 1$$

$$F(x) = \frac{\log_{10}(999x+1) + \log_5(4x+1)}{4}, \quad 0 \le x \le 1$$

$$F(x) = \frac{\log_2(x+1) + 2^x - 1}{2}, \quad 0 \le x \le 1$$

3.2 앤사리-브래들리 검정 통계량의 편의

위치모수에 대한 윌콕슨 순위합 검정의 편의 유무를 조사하였고, 척도 모수에 대한 비모수 검정 통계량의 편의 유무를 조사하고자 대표적인 앤사리-브래들리검정을 다음과 같이 고려하였다.

연속확률함수인 $F(x)$와 $G(x)$로부터 확률표본 $X_1, X_2, \cdots, X_m$ 와 $Y_1, Y_2, \ldots, Y_n$ 을 추출하고 $\gamma=\sigma_Y/\sigma_X$라 할 때, $H_0: \gamma=1$, $H_1: \gamma \neq 1$을 검정하자. $X_{(1)} < \cdots < X_{(m)}$는 X_s의 순서통계량이며, $Y_{(1)} < \cdots < Y_{(n)}$ 또한 Y_s의 순서통계량이다. 앤사리-브래들리의 검정함수(test function)인 ϕ_1을 다음과 같이 정의한다.

$$\phi_1(X_1, \ldots, X_m; Y_1, \ldots, Y_n) = \begin{cases} 1 \begin{cases} \text{만약} X_{(\frac{m}{2}-1)} < Y_1 \cdots Y_n < X_{(\frac{m}{2})} \\ \text{또는} Y_{(\frac{n}{2}-1)} < X_1 \cdots X_m < Y_{(\frac{n}{2})} \end{cases} \\ 0, \text{그렇지않으면} \end{cases} \tag{20}$$

유의수준은 $\alpha=2(m!\,n!)/(m+n)!$ 이며 검정력 함수는 다음과 같다.

$$\beta_1(\gamma) = \frac{m!}{(\frac{m}{2}-1)!(\frac{m}{2}-1)!} \times \int_{-\infty}^{\infty} \int_{-\infty}^{y} (G(y)-G(x))^m F(x)^{\frac{m}{2}-1}$$

$$\times (1-F(y))^{\frac{m}{2}-1} \times f(x) f(y) dx dy$$

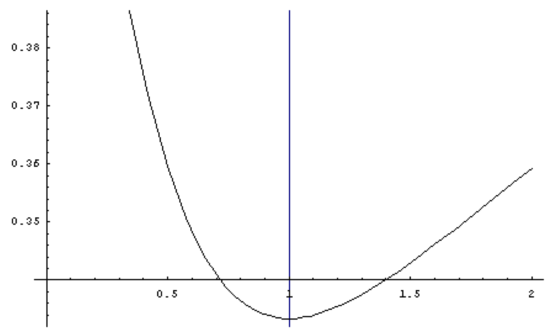
$$+ \frac{n!}{(\frac{n}{2}-1)!(\frac{n}{2}-1)!} \times \int_{-\infty}^{\infty} \int_{-\infty}^{y} (F(y)-F(x))^n G(x)^{\frac{n}{2}-1}$$

$$\times (1-G(y))^{\frac{n}{2}-1} g(x)g(y)dxdy. \tag{21}$$

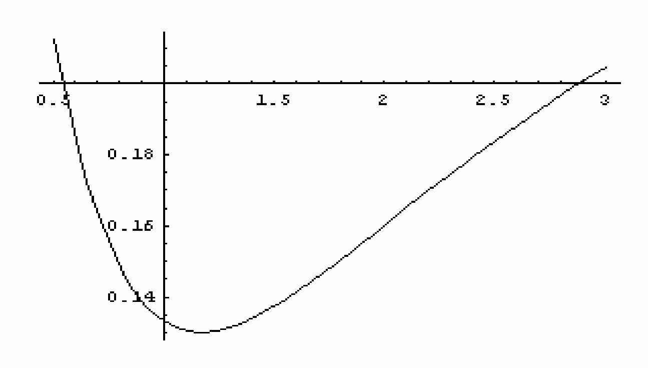
$F(x)$와 $G(x)$는 다음과 같이 정의한다.

$$F_X(x) = N(0,\ 1) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt\ ,\ -\infty \le x \le \infty,$$

$$G_X(x) = N(0,\ \gamma),\quad -\infty \le x \le \infty. \tag{22}$$



[그림 9]$m=n=2$ 의 검정력 함수



[그림 10]$m=2,\ n=4$의 검정력 함수

$m=n=2$일 때 검정력 함수인 [그림 9]는 $\beta_1(1)=0.3333$이고 최소값이므로 편의가 없음을 알 수 있다. 그리고 $m=2,\ n=4$의 검정력 함수인 [그림 10]에서 보면 $\beta_1(1)=0.133333$이고, $\beta_1(1.5)=0.130201$값으로 $\gamma=1$에서 최소값이 아니므로 편의가 있음을 알 수 있다. 즉, 앤사리-브래들리 검정은 편의가 있음이 위의 반례를 제시함으로써 증명되었다.

제4장 결 론

Lehmann(1959,1986)은 위치모수에 대한 단측 윌콕슨 순위합 검정이 편의가 없음을 보여주고, 위치모수에 대한 양측 윌콕슨 순위합 검정의 편의가 존재하는 것에 대한 의문을 제기했다. 그 후 Sugiura(1965)는 표본크기가 다른 경우 편의가 있음을 보여주었다. 그러나 표본의 크기가 같을 때에는 편의가 있음을 밝히지 못했다. 이에 Sugiura et al.(2006)의 논문에서 표본크기가 같을 때 양측 윌콕슨 순위합 검정의 편의가 있음을 보여주었다. 그러나 이것은 표본크기가 2일 때 국한된 결과였다.

본 논문에서는 윌콕슨 순위합 검정에 대한 추가적으로 반례를 제시하였다. 또한 척도모수에 대한 비모수 검정법이 편의가 있음을 반례를 통해 보여주었다. 그러나 표본의 크기가 3이상일 때의 편의가 있는 반례를 제시하지 못하였으며 본 연구의 한계점이라 할 것이다.

참 고 문 헌

[1] 이승우(2003). *『통계학의 비모수 추정에 관한 역사적 고찰』*. 한국 수학사 학회.

[2] 최영훈(2004). *『순위통계량』,서울, 자유아카데미.*

[3] Ansari, A. R. and Bradley, R. A.(1960). Rank-sum tests for dispersions. *Ann. Math. Statist*. 31, 1174-89.

[4] Lehmann, E. L.(1959). *Testing statistical hypotheses*. New York : Wiley.

[5] Lehmann, E. L.(1986). *Testing statistical hypotheses* (2nd ed). New York: Wiley.

[6] Mann, H. and Whitney, D. R.(1947). On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist. 18, 50-60.*

[7] Sugiura, N.(1965). An example of the two-sided Wilcoxon test which is not unbiased. *Ann. Math. Statist*. *17, 261-263*.

[8] Sugiura, N., Murakami, H., Lee, S.(2006). Biased and unbiased two-sided Wilcoxon tests for equal sample sizes. *Ann. Math. Statist*. *58, 93-100.*

[8] Wilcoxon, F.(1945). Individual comparisons by ranking methods. *Biometrics* 1, 80-3.

ABSTRACT

A study on the unbiasedness of non-parametric rank test

Lee, Ka Young
Department of Statistics
The Graduate School
Sungshin Women's University

Lehmann(1959,1986) showed that one-sided wilcoxon rank test is unbiased against at location parameter family of distributions. But he raised the question of two-sided wilcoxon rank test is biased. Then Sugiura(1965) showed that two-sided wilcoxon rank test for different sizes is biased against a location parameter family of distributions by giving a counter example. Sugiura, etal.(2006) showed that the two-sided wilcoxon rank test for equal sample sizes is biased against a location parameter family of distributions by giving a counter example. But it is restricted that sample size is 2. In this paper, we have found another counterexample of Sugiura, etal.(2006). For scale parameters, we have considered Ansari-Bradley test and showed that Ansari-Bradley test for different sample sizes is unbiased against a scale parameter family of distributions by giving a counter example.