The Korean Communications
in Statistics Vol. 8, No. 1, 2001
pp. 159-171

A Modification of the W Test for Exponentiality

Nam Hyun Kim[1]

Abstract

Shapiro and Wilk (1972) developed a test for exponentiality with origin and scale unknown. The procedure consists of comparing the generalized least squares estimate of scale with the estimate of scale given by the sample variance. However the test statistic is inconsistent ; that is, the power of the test will not approach 1 as the sample size increases. Hence we give a test based on the ratio of two asymptotically efficient estimates of scale. We also have conducted a power study to compare the test procedures, using Monte Carlo samples from a wide range of alternatives. It is found that the suggested statistics have higher power for the alternatives with the coefficient of variation greater than or equal to 1.

Keywords : exponentiality, goodness of fit tests, asymptotically efficient estimates.

1. 서 론

특정한 분포를 이용한 통계적 모형이 주어진 자료에 적합한지를 알아보는 적합도 검정(goodness of fit tests)은 이론통계와 응용통계 양 분야에서 중요하게 다루어져왔다. 지수분포(exponential distribution)는 정규분포 다음으로 통계학에서 자주 쓰이는 분포라고 할 수 있을 것이다. 이는 생존분석, 신뢰성이론 등에서 중요하게 이용되어왔다.

지수분포는 분포함수

$$F(x;\alpha,\beta) = 1 - \exp\left(-\frac{x-\alpha}{\beta}\right),\ x > \alpha,\ \beta > 0,\ -\infty < \alpha < \infty \tag{1.1}$$

을 갖는 분포로 α는 위치모수(location parameter, origin)이고 β는 척도모수(scale parameter)이다. 분포함수 $F(x;\alpha,\beta)$을 나타내기 위해서 $\exp(\alpha,\beta)$를 쓰기로 한다. 또한 $\alpha=0$, $\beta=1$인 표준 지수분포 $F(x;0,1)$은 $F_0(x)$로 나타내고 $f_0(x)$를 F_0에 해당하는 확률밀도함수, F_0^{-1}은 F_0의 역함수라고 하자.

1) Assistant Professor, Department of Science, Hongik University, Seoul, 121-791, Korea.
E-mail : nhkim@wow.hongik.ac.kr

$X_1, \cdots, X_n$이 연속확률분포함수 $G(x)$에서의 확률표본이고 이 표본의 순서통계량을 $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$이라고 하자. $X_1, \cdots, X_n$이 지수분포의 모형에 적합한지를 검정하는 것은

$$H_0: \ G(x) = \exp(\alpha, \beta) \tag{1.2}$$

을 검정하는 것이다. 대립가설 H_A는

$$H_A: \ G(x) \neq \exp(\alpha, \beta) \tag{1.3}$$

이라고 하자. (1.2)의 귀무가설 H_0를 검정하기 위해 제안된 통계량은 무수히 많다. 이 경우 대부분 α는 기지 또는 0으로, β는 미지로 가정하고 있다. 예를 들면 Gail과 Gastwirth(1978a, 1978b), Jackson(1967), Stephens(1978) 등이 있다. 또한 D'Agostino와 Stephens(1986, 4, 5, 10장)에서 이에 대한 전반적인 설명과 참고문헌을 제시하고 있다.

본 논문에서는 모수 α와 β가 모두 미지일 때의 검정법을 고려한다. 이 경우의 검정법은 많이 고려되고 있지 않다. 한 가지 이유는 지수분포의 통계적인 성질(2절의 **Fact 1**)을 이용하여 α가 미지인 경우의 검정법 대신 $\alpha = 0$인 검정법을 이용할 수 있기 때문이다. 이 경우 위에서 언급한 대로 이용할 수 있는 검정법이 많이 소개되어 있다. 그러나 이러한 방법으로 모수 α를 제거하는 것이 가설 (1.2)를 검정하는데 있어서 α를 추정하는 것보다 더 좋은 검정법을 제공할 수 있는지는 확실하지 않다. 두 가지 방법을 비교하기 위해서 Spinelli와 Stephens(1987)은 실제적인 여러 가지 대립가설 하에서 모의실험(simulation)을 실시하였다. 또한 D'Agostino와 Stephens(1984, 4.16절, 10.14절)에서 미지의 모수가 검정에 미치는 영향과 α, β가 모두 미지인 경우의 검정법에 대해서 소개하고 있다. 그러나 모수에 대한 정보가 검정에 어떤 영향을 미치는지에 대해서는 좀 더 체계적인 연구가 필요하다.

지수분포의 검정에서 α와 β가 모두 미지인 경우 소개된 대표적인 통계량은 Shapiro와 Wilk (1972)의 W_E-통계량이다. 만일 가설 (1.2)의 H_0가 사실이라면, 모형

$$E(X_{(j)}) = \alpha + \beta \widetilde{v_{jn}} \tag{1.4}$$

가 성립할 것이다. 여기서 E는 기대값을 의미하고 $\widetilde{v_{jn}} = E(V_{(j)})$, $V_{(1)}, \ldots, V_{(n)}$은 $\exp(0,1)$에서의 순서통계량이다. 모형 (1.1)에서 α와 β의 일반화 최소자승 추정량(generalized least squares estimate) $\hat{\alpha}$, $\hat{\beta}$은

$$\hat{\alpha} = X_{(1)}, \quad \hat{\beta} = \frac{n(\overline{X} - X_{(1)})}{n-1} \tag{1.5}$$

이다. 여기서 $\overline{X}$는 표본평균을 의미한다. Shapiro와 Wilk(1972)의 W_E-통계량은 식(1.5)의 $\hat{\beta}$와 표본분산 $s^2 = S^2/(n-1) = \sum_{j=1}^{n}(X_j - \overline{X})^2/(n-1)$을 비교하는 것으로 β의 두 추정량의 비에 기초한 통계량이다. W_E-통계량의 형태는

$$W_E = \frac{n(\overline{X} - X_{(1)})^2}{(n-1)S^2} \tag{1.6}$$

이며 이는 양쪽 검정통계량이다. Shapiro와 Wilk(1972)에는 표본크기 n이 3부터 100일 때 W_E-

통계량의 분위수가 주어져 있다. Spinelli와 Stephens(1987)에서 지적한대로 W_E-통계량은 대부분의 대립가설에서 좋은 검정력을 갖는다. 그러나 W_E-통계량의 가장 심각한 단점은 이 검정법이 일치성(consistency)을 갖지 않는다는데 있다.

본 논문에서는 W_E-통계량의 이러한 단점을 보완하기 위한 수정된 형태의 W_E-통계량인 N_E-통계량을 소개하고 이의 통계적인 성질을 알아본다. 또한 모의실험을 통하여 이들 통계량의 검정력을 비교하여 그 특성을 알아보고자 한다.

2. 수정된 W_E-통계량

1절에서 언급한대로 식(1.6)의 W_E-통계량의 가장 큰 단점은 이 통계량에 기초한 검정법이 일치성을 갖지 않는다는 것이다. 다시 말해서 이 검정법의 검정력이 표본의 수가 증가해도 1로 가까이 가지 않는 분포가 존재한다는 말이다. 지수분포의 경우 모집단의 표준편차와 평균의 비인 변동계수(coefficient of variation)가 $C_V = \sigma/\mu = 1$이므로

$$nW_E \xrightarrow{p} 1/C_V^2 = 1, \quad n\to\infty \text{ 일 때}, \tag{2.1}$$

이다. 그러므로 $C_V=1$인 다른 분포에서도 nW_E는 1로 수렴할 것이다. 베타분포 $B(a,b)$에서 $a<1$, $b=a(a+1)/(1-a)$인 경우가 이에 해당한다. 예를 들면 (a,b)가 $(1/4, 5/12)$일 때를 고려할 수 있다(Spinelli와 Stephens(1987)). 이 분포는 검정력을 비교하는 <표 1>에 포함되어 있다.

D'Agostino와 Stephens(1986, 5장)에 따르면, 일반적으로 W_E와 같이 β의 두 추정량의 비로 구성된 통계량에 기초한 검정법의 일치성은 두 추정량의 통계적인 성질에 달려있다. 따라서 β의 유효추정량이나 점근유효추정량으로 구성된 통계량이 좀 더 좋은 검정법을 제시할 수 있으리라 짐작되고, 이에 대한 이론적인 연구가 필요하다고 생각된다. 이러한 사실을 확인할 수 한다면 W_E에 기초한 검정법이 일치성을 갖지 않는 이유는 직관적으로 분모의 표본분산이 지수분포의 경우 점근유효추정량(asymptotically efficient estimator)이 아니기 때문일 것이다. 실제로

$$\sqrt{n}(s^2-\beta^2) \xrightarrow{d} N(0, 8\beta^4) \tag{2.2}$$

이고 피셔정보(Fisher information)는 $I(\beta^2) = 1/4\beta^4$이다(Ferguson(1996, 7절, 19절)). 따라서 분모의 추정량을 β^2의 점근유효추정량으로 대치하면 W_E-통계량의 단점을 보완할 수 있을 것이라 기대된다. β^2의 추정량으로

$$L_n = \frac{1}{n-1}\sum_{j=2}^{n}(X_{(j)}-X_{(1)})^2/v_{jn}\ , \quad v_{jn} = F_0^{-1}\left(\frac{j}{n+1}\right) = -\log\left(1-\frac{j}{n+1}\right) \tag{2.3}$$

을 고려하면 L_n은 β^2의 점근유효추정량(아래의 **정리 4**)이므로 수정된 W_E-통계량으로 N_E-통계량,

$$N_E = \frac{n(\overline{X}-X_{(1)})^2}{(n-1)^2 L_n} = \frac{n(\overline{X}-X_{(1)})^2}{(n-1)\sum_{j=2}^{n}(X_{(j)}-X_{(1)})^2/v_{jn}} \tag{2.4}$$

을 제안한다. de Wet과 Venter(1973)는 일반적인 척도모수 분포모임(scale parameter family of distributions)에 대해서 식(2.4)와 유사한 형태의 통계량을 제안하고 이의 점근분포(asymptotic distribution)을 구하였다. 이에 따르면 일반적인 분포에서 $(n-1)N_E \le 1+o_p(1)$ 이므로 N_E-통계량은 적당한 c에 대해서 $N_E < c$일 때 H_0를 기각하는 것이 합리적이다. N_E-통계량은 de Wet과 Venter(1973)의 통계량을 β뿐만 아니라 α도 미지인 경우의 지수분포에 적용한 것이라고 생각할 수 있다.

$V_{(1)}, \ldots, V_{(n)}$을 $\exp(0,1)$에서의 순서통계량이라고 할 때, 식(2.3)의 L_n에서 v_{jn}은 $\widetilde{v_{jn}} = E(V_{(j)})$의 근사값이라고 할 수 있다. 일반적으로 순서통계량의 기대값은 계산하기 힘드나 지수분포의 경우에는 $\widetilde{v_{jn}} = \sum_{i=1}^{j} \frac{1}{n-i+1}$ 으로 수치적 적분 없이 계산이 가능하다. 따라서 L_n의 v_{jn}을 $\widetilde{v_{jn}}$로 대치하면 β^2의 추정량으로

$$\widetilde{L_n} = \frac{1}{n-1} \sum_{j=2}^{n} (X_{(j)} - X_{(1)})^2 / \widetilde{v_{jn}} \tag{2.5}$$

을 고려할 수 있고, 수정된 W_E-통계량으로 $\widetilde{N_E}$-통계량,

$$\widetilde{N_E} = \frac{n(\overline{X} - X_{(1)})^2}{(n-1)^2 \widetilde{L_n}} \tag{2.6}$$

을 제안할 수 있다. 이 절에서는 수정된 W_E-통계량인 N_E와 $\widetilde{N_E}$의 성질을 살펴보기로 하자.

Fact 1. $X_{(1)}, \cdots, X_{(n)}$이 $\exp(\alpha, \beta)$로 부터의 크기 n인 순서통계량이면, $Y_{(i)} = X_{(i+1)} - X_{(1)}$, $i=1, \ldots, n-1$은 $\exp(0, \beta)$로 부터의 크기 $n-1$인 순서통계량이다.

보조정리 1. 식(2.4)의 N_E와 식(2.6)의 $\widetilde{N_E}$는 위치, 척도 불변(origin and scale invariant)인 통계량이다.

증명. $X_{(i)} = \alpha + \beta V_{(i)}$, $i=1, \ldots, n$을 대입해 보면 자명하다.

□

식(2.3), (2.5)의 L_n, $\widetilde{L_n}$가 β^2의 점근유효추정량임을 보이기 위하여 Chernoff, Gastwirth와 Johns(1967)의 주요 정리를 소개한다.

$X_{(1)}, \ldots, X_{(n)}$이 분포함수 $G(x)$로부터의 순서통계량이고

$$T_n = \frac{1}{n} \sum_{j=1}^{n} c_{jn} h(X_{(j)}) \tag{2.7}$$

라고 할 때, Chernoff 외(1967)은 T_n이 점근적 정규분포를 갖기 위한 가정과 그에 따른 정리에 대해서 연구하였다. G^{-1}를 분포함수 G의 역함수라고 하고 $U_{(1)}, \ldots, U_{(n)}$을 균일분포 $U(0,1)$에서의 순서통계량이라고 하면, 분포이론으로부터

$$X_{(j)} = G^{-1}(U_{(j)})$$

로 나타낼 수 있고, 마찬가지로

$$V_{(j)} = -\log(1 - U_{(j)})$$

이다. 이로부터 함수 H와 $\tilde{H}$를

$$h(x) = H(u) = \tilde{H}(v) \tag{2.8}$$

으로 정의한다. 여기서 $x = G^{-1}(u)$, $v = -\log(1-u)$이다. 그리고

$$\mu_n = \frac{1}{n} \sum_{j=1}^{n} c_{jn} \tilde{H}(\widetilde{v_{jn}}) \tag{2.9}$$

$$a_{jn} = \frac{1}{n-j+1} \sum_{i=j}^{n} c_{in} \tilde{H}'(\widetilde{v_{in}}) \tag{2.10}$$

$$\sigma_n^2 = \frac{1}{n} \sum_{j=1}^{n} a_{jn}^2 \tag{2.11}$$

이다.

가정 A. $\tilde{H}(v)$는 $0 < v < \infty$에서 미분가능이다.

가정 B. $H(\cdot)$은 $(0,1)$에서 연속이고 모든 $x, y \in [a, , b] \subset (0,1)$에 대해서

$$|H(x) - H(y)| \leq M|x - y|$$

이 성립하는 M이 존재한다. 또한 H'가 존재하고 almost everywhere 연속이다.

가정 C.

$$\frac{1}{n} \sum_{j=1}^{n} c_{jn} H'\left(\frac{j}{n+1}\right)\left[\frac{j}{n+1}\left(1 - \frac{j}{n+1}\right)\right]^{1/2} = O(\sigma_n) \tag{2.12}$$

$$\frac{1}{n} \sum_{j=[n\delta]+1}^{[n(1-\delta)]} |c_{jn}| = O(\sigma_n), \qquad \delta > 0. \tag{2.13}$$

가정 D. 다음 (a), (b)를 만족하는 δ_0, $0 < \delta_0 < 1$ 가 존재한다.

(a) 충분히 큰 n에 대해서 $j \leq n\delta_0$일 때 $c_{jn} = 0$이거나, 또는 각 $K > 0$에 대해서 $0 < u_1, u_2 < \delta_0$이고 $K^{-1} < u_1/u_2 < K$일 때 $M^{-1} < H'(u_1)/H'(u_2) < M$을 만족하는 $M < \infty$이 존재한다.

(b) 충분히 큰 n에 대해서 $j \geq n(1-\delta_0)$일 때 $c_{jn} = 0$이거나, 또는 각 $K > 0$에 대해서 $1 - \delta_0 < u_1, u_2 < 1$이고 $K^{-1} < (1-u_1)/(1-u_2) < K$일 때 $M^{-1} < H'(u_1)/H'(u_2) < M$을 만족하는 $M < \infty$이 존재한다.

정리 1. (Chernoff 외(1967)의 정리 1, 보조정리 2와 3.) **가정 A, B, C, D**가 성립할 때,

$$\sqrt{n}(T_n - \mu_n)/\sigma_n \xrightarrow{d} N(0,1)$$

이 성립한다. 여기서 μ_n과 σ_n은 각각 식(2.9)와 식(2.11)이다.

□

$c_{jn}=J(j/(n+1))$일 때 즉,

$$T_n=\frac{1}{n}\sum_{j=1}^{n}J\left(\frac{j}{n+1}\right)h(X_{(j)}) \tag{2.14}$$

일 때,

$$\mu=\int_0^1 J(u)H(u)\,du \tag{2.15}$$

$$\alpha(u)=\frac{1}{1-u}\int_u^1 J(w)H'(w)(1-w)\,dw \tag{2.16}$$

$$\sigma^2=\int_0^1\alpha^2(u)\,du=\int_0^1\int_0^1 J(u)H'(u)J(w)H'(w)K(u,w)\,du\,dw \tag{2.17}$$

$$K(u,w)=K(w,u)=u(1-w)=\min(u,v)-uw,\quad 0\le u\le w\le 1 \tag{2.18}$$

$$B_1=\int_0^1 J(u)H'(u)[u(1-u)]^{1/2}\,du \tag{2.19}$$

$$B_2=\int_\delta^{1-\delta}J(u)du,\qquad 0<\delta<1/2 \tag{2.20}$$

라고 하자.

가정 C*. B_1과 B_2가 절대 수렴한다.

정리 2. (Chernoff 외(1967)의 정리 3.) **가정 B, C*, D**가 성립하고, 식(2.17)의 σ^2이 절대수렴하며,

$$\mu_n=\frac{1}{n}\sum_{j=1}^{n}J\left(\frac{j}{n+1}\right)H\left(\frac{j}{n+1}\right) \tag{2.21}$$

가 $\mu_n=\mu+o(n^{-1/2})$을 만족하면,

$$\sqrt{n}(T_n-\mu)\xrightarrow{d}N(0,\sigma^2)$$

이 성립한다. 여기서 μ와 σ^2은 각각 식(2.15)와 식(2.17)이다.

Fact 2. 적분의 정의를 이용하면,

$$\frac{j}{n+1}<1-e^{-\widetilde{v_{jn}}}<\frac{j}{n+1/2}$$

가 성립하고, $v_{jn}=-\log(1-j/(n+1))$이므로 평균값 정리(mean value theorem)를 이용하면,

$$0\le\widetilde{v_{jn}}-v_{jn}<\frac{j}{(2n+1)(n+1/2-j)}, \tag{2.22}$$

$$\left|\frac{1}{\widetilde{v_{jn}}}-\frac{1}{v_{jn}}\right| < \frac{1}{\left(-\log\left(1-\frac{j}{n+1/2}\right)\right)^2}\frac{j}{(2n+1)(n+1/2-j)} \tag{2.23}$$

가 성립한다.

보조정리 2.

$$T_n=\frac{1}{n}\sum_{j=1}^{n} V_{(j)}^2/\widetilde{v_{jn}},$$

$V_{(1)},\ldots,V_{(n)}$이 $\exp(0,1)$에서의 순서통계량일 때, **가정** **A, B, C, D**가 성립한다.

증명. 식(2.7)에서

$$c_{jn}=1/\widetilde{v_{jn}} \tag{2.24}$$

이고, $G(x)$가 $\exp(0,1)$의 분포함수, 즉 $G(x)=F_0(x)$이므로 식(2.8)에서

$$h(x)=x^2=\widetilde{H}(v)=v^2 \tag{2.25}$$

$$h(x)=H(u)=(F_0^{-1}(u))^2=(-\log(1-u))^2 \tag{2.26}$$

이다.

가정 A는 식(2.25)로부터 자명하다.

가정 B는 $H'(u)=2F_0^{-1}(u)/f_0(F_0^{-1}(u))=(-2\log(1-u))/(1-u)$에서

$$\varepsilon < u < 1-\varepsilon \text{일 때 } H'(u)\le(-2\log\varepsilon)/\varepsilon$$

이므로 평균값 정리에 의해서 만족된다.

가정 C가 만족됨을 보이자. 식(2.10), (2.11), (2.24), (2.25)에서, $a_{jn}=2$, $\sigma_n^2=4$이다. 그리고 식(2.23), (2.24), (2.26)에서 식(2.12)는

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\widetilde{v_{jn}}}\frac{2F_0^{-1}(j/(n+1))}{f_0(F_0^{-1}(j/(n+1)))}\left(\frac{j}{n+1}\left(1-\frac{j}{n+1}\right)\right)^{1/2}\\
&\le\frac{2}{n}\sum_{j=1}^{n}\left|\frac{1}{\widetilde{v_{jn}}}-\frac{1}{v_{jn}}\right|v_{jn}\left(\frac{j/(n+1)}{1-j/(n+1)}\right)^{1/2}+\frac{2}{n}\sum_{j=1}^{n}\left(\frac{j/(n+1)}{1-j/(n+1)}\right)^{1/2}\\
&\le\frac{2}{n}\sum_{j=1}^{n}\frac{-\log(1-j/(n+1))}{(-\log(1-j/(n+1/2)))^2}\frac{j}{(2n+1)(n+1/2-j)}\left(\frac{j/(n+1)}{1-j/(n+1)}\right)^{1/2}\\
&\qquad+\frac{2}{n}\sum_{j=1}^{n}\left(\frac{j/(n+1)}{1-j/(n+1)}\right)^{1/2}\\
&=\frac{2}{2n+1}O\left(\int_{1/(n+1)}^{n/(n+1)}\frac{\sqrt{x}}{(1-x)^{3/2}}dx\right)+O\left(\int_{1/(n+1)}^{n/(n+1)}\left(\frac{x}{1-x}\right)^{1/2}dx\right)\\
&=O\left(\frac{1}{\sqrt{n}}\right)+O(1)\\
&=O(1)=O(\sigma_n)
\end{aligned}$$

이고 식(2.13)도 위와 마찬가지 방법으로 보일 수 있다.

가정 D를 보이기 위해서 $0<\delta_0<1/2$인 임의의 δ_0를 택하자.

(a) $0<u_1, u_2<\delta_0$이고 $K^{-1}<u_1/u_2<K$일 때

$$\frac{H'(u_1)}{H'(u_2)}=\frac{1-u_2}{1-u_1}\frac{-\log(1-u_1)}{-\log(1-u_2)}$$

$$1-\delta_0<\frac{1-u_2}{1-u_1}<\frac{1}{1-\delta_0}$$

이다. 또한 $-\log(1-u)\geq u$에서

$$\frac{-\log(1-u_1)}{-\log(1-u_2)}\leq\frac{u_1}{u_2}\left(1+\frac{1}{2}u_1+\frac{1}{3}u_1^2+\cdots\right)\leq K\left(1+\frac{1}{2}\delta_0+\frac{1}{3}\delta_0^2+\cdots\right)$$

이고 $1+1/2\delta_0+1/3\delta_0^2+\cdots$은 수렴급수이므로 이를 C_1이라고 하면 $\frac{-\log(1-u_1)}{-\log(1-u_2)}\leq KC_1$이다.

마찬가지로 $\frac{-\log(1-u_1)}{-\log(1-u_2)}\geq\frac{1}{KC_1}$ 임을 보일 수 있으므로

$$\frac{1-\delta_0}{KC_1}<\frac{H'(u_1)}{H'(u_2)}<\frac{KC_1}{1-\delta_0}$$

이다.

(b) $1-\delta_0<u_1, u_2<1$이고 $K^{-1}<(1-u_1)/(1-u_2)<K$일 때

$$-\log(1-u_2)\geq-\log\delta_0,$$

$$-\log\left(\frac{1-u_2}{K}\right)>-\log(1-u_1)$$

이므로

$$\frac{-\log(1-u_1)}{-\log(1-u_2)}<\frac{\log K}{-\log\delta_0}+1$$

이고 마찬가지 방법으로 $\frac{-\log(1-u_1)}{-\log(1-u_2)}>1/\left(\frac{\log K}{-\log\delta_0}+1\right)$이다. 따라서 $C_2=\frac{\log K}{-\log\delta_0}+1$이라고 하면 $\frac{1}{KC_2}<\frac{H'(u_1)}{H'(u_2)}<KC_2$이 성립한다.

□

정리 3. 식(2.6)의 $\widehat{N_E}$의 분모, 분자인 $(\overline{X}-X_{(1)})^2$, $\widehat{L_n}$는 $X_1,\ldots,X_n$이 $\exp(\alpha,\beta)$에서의 표본일 때 β^2의 점근유효추정량이다. 즉

$$\sqrt{n}((\overline{X}-X_{(1)})^2-\beta^2)\xrightarrow{d}N(0,4\beta^2) \tag{2.27}$$

$$\sqrt{n}(\widehat{L_n}-\beta^2)\xrightarrow{d}N(0,4\beta^4) \tag{2.28}$$

이다.

증명. **Fact 1**에 의하면,

$$\overline{X}-X_{(1)}=\overline{Y}=\frac{1}{n-1}\sum_{j=1}^{n-1}Y_{(j)} \tag{2.29}$$

$$\widetilde{L_n}=\frac{1}{n-1}\sum_{j=1}^{n-1}Y_{(j)}^2/\widetilde{v_{jn}} \tag{2.30}$$

이고, $Y_{(1)},\ldots,Y_{(n-1)}$은 $\exp(0,\beta)$에서의 순서통계량이다. 식(2.27)은 중심극한정리(central limit theorem)와 Cramér 정리(Fergerson(1996, 7절) 참조)에 의해서 자명하다.

$V_{(i)}$가 $\exp(0,1)$에서의 순서통계량일 때 $Y_{(i)}=\beta V_{(i)}$이므로 식(2.30)에 의해서 $\beta=1$일 때 식(2.28)이 성립함을 보이면 충분하다. **정리 1**와 **보조정리 2**에 의해서

$$\sqrt{n}\,(\widetilde{L_n}-\mu_n)\xrightarrow{d}N(0,4) \tag{2.31}$$

이 성립한다. 식(2.9), (2.24), (2.25)에서 $\mu_n=n^{-1}\sum_{j=1}^{n}\widetilde{v_{jn}}$이므로 식(2.22)을 이용하면

$$\begin{aligned}\mu_n-1&=\frac{1}{n}\sum_{j=1}^{n}(\widetilde{v_{jn}}-v_{jn})+\frac{1}{n}\sum_{j=1}^{n}v_{jn}-1\\ &\le\frac{1}{n}\sum_{j=1}^{n}\frac{j}{(2n+1)(n+1/2-j)}+\frac{1}{n}\sum_{j=1}^{n}\Big(-\log\Big(1-\frac{j}{n+1}\Big)\Big)-\int_0^1-\log(1-x)\,dx\\ &=O\Big(\frac{\log n}{n}\Big)=o\Big(\frac{1}{\sqrt{n}}\Big)\end{aligned} \tag{2.32}$$

이므로, 식(2.31), (2.32)에서

$$\sqrt{n}\,(\widetilde{L_n}-1)\xrightarrow{d}N(0,4)$$

이고 정리는 성립한다.

□

정리 4. 식(2.3)의 L_n은 $X_1,\ldots,X_n$이 $\exp(\alpha,\beta)$에서의 표본일 때 β^2의 점근유효추정량이다. 즉,

$$\sqrt{n}\,(L_n-\beta^2)\xrightarrow{d}N(0,4\beta^4)$$

이다.

증명. **정리 3**의 증명에서와 마찬가지 이유로,

$$L_n=\frac{1}{n-1}\sum_{j=1}^{n-1}V_{(j)}^2/v_{jn},$$

$V_{(1)},\ldots,V_{(n-1)}$이 $\exp(0,1)$에서의 순서통계량일 때를 고려하면 충분하다.

식(2.14)에서 $X_{(1)},\ldots,X_{(n)}$의 분포함수 $G(x)$가 $G(x)=1-e^{-x}$이고,

$$J(u)=(-\log(1-u))^{-1},\ \ h(x)=H(u)=(-\log(1-u))^2$$

일 때를 고려하자.

이 경우 식(2.19), (2.20)의 B_1, B_2는 $B_1=2\int_0^1\Big(\frac{u}{1-u}\Big)^{1/2}du\le 2$이고, 고정된 δ, $0<\delta<1/2$에 대해서 $B_2=\int_\delta^{1-\delta}(-\log(1-u))^{-1}du<\infty$이므로 **가정 C***가 만족된다.

또한 식(2.15)의 μ는 $\mu=1$이고 식(2.21)의 μ_n은 $\mu_n=\frac{1}{n}\sum_{j=1}^{n}(-\log(1-j/(n+1)))$에서 $\mu_n-1=o(n^{-1/2})$을 만족한다. 식(2.17)의 σ^2은

$$\begin{aligned}\sigma^2&=\int_0^1\int_0^1\frac{4K(u,w)}{(1-u)(1-w)}\,dudw\\&=8\int_0^1\int_u^1\frac{u(1-w)}{(1-u)(1-w)}\,dwdu\\&=8\int_0^1 udu=4\end{aligned}$$

이므로 **정리** 2과 **보조정리** 2에 의해서

$$\sqrt{n}(L_n-1)\xrightarrow{d}N(0,4)$$

이고 정리는 성립한다.

□

정리 3과 **정리** 4의 사실, 즉 식(2.3)과 (2.5)의 L_n, $\widetilde{L_n}$이 점근유효추정량이라는 사실과 관련하여 제안된 식(2.4), (2,6)의 통계량, N_E, $\widetilde{N_E}$가 일치성을 만족할 것이라 예상되나 이에 대한 연구는 완성되지 않은 상태이다. 이를 대신하여, W_E, N_E, $\widetilde{N_E}$-통계량의 검정력을 모의실험을 통하여 비교하여 보았다.

3. 모의실험결과 및 결론

2절에서 소개한 수정된 W_E-통계량인 식(2.4), (2.6)의 N_E, $\widetilde{N_E}$의 효율성을 소표본일 때 조사하기 위해서, 모의실험(simulation)을 행하였다. 이를 위하여 여러 가지 대립가설에서 크기 $n=10,\ 20,\ 50,\ 100$인 표본을 S-plus를 이용하여 추출하였다. 이 때 표본의 수는 $N=5000$을 사용하였다. 대립가설의 분포로는 신뢰성이론 등에서 지수분포의 대립가설로 많이 이용되는 와이블, 감마분포 등을 포함하여 여러 가지 분포가 고려되었다. <표 2>에서 $\chi^2(m)$은 자유도가 m인 카이제곱분포이고, $U(0,1)$은 $(0,1)$에서의 균등분포이다. 또한 Weib(m)은 확률밀도함수가 $g(x)=mx^{m-1}e^{-x^m}$, $x>0$인 분포이고, lognorm(m)은 $g(x)=C\exp[-(\log x)^2/(2m^2)]$, $x>0$인 로그정규분포이다. 그리고 $(1/2)N$은 Y가 표준정규분포 $N(0,1)$을 따를 때 $X=|Y|$의 분포이며 $(1/2)C$는 Y가 중앙값이 0인 코쉬분포를 따를 때 $X=|Y|$의 분포이다. $g(x)$가 확률밀도함수, $G(x)$가 확률분포함수일 때 고장률 또는 위험률(failure rate, hazard rate)을 $h(x)=g(x)/(1-G(x))$, $x>0$이라고 하면, 이 중 $\chi^2(4)$, $U(0,1)$, Weib(1.5), $(1/2)N$은 증가고장률(increasing failure rate : IFR) 분포이고, $\chi^2(1)$, Weib(0.8), lognorm(1), $(1/2)C$는 감소고장률(decreasing failure rate : DFR) 분포이다. IFR 분포는 변동계수 C_V가 $C_V<1$이고 DFR 분포는 $C_V>1$이다.

<표 1>은 통계량 $(n-1)^2 N_E$과 $(n-1)^2 \widetilde{N_E}$의 하방 α 분위수를 나타낸다. 이 값들은 지수분포에서 표본의 수 $N=10,000$을 추출하여 구하였다.

<표 2>에서는 세 가지 통계량의 검정력을 유의수준 $100\alpha=10\%$에서 비교하고 있다. 이를 통하여 보면, 우선 특이할만한 사실은 베타분포의 경우는 $B(1/2,3/2)$, $B(1/4,5/12)$, $B(1/8,9/56)$ 모두에서 Shapiro와 Wilk(1972)의 W_E-통계량은 표본크기가 커질수록 검정력이 점점 감소하고, 이러한 사실은 2절의 앞부분에서 이미 지적한 바 있다. 그러나 제안된 N_E, $\widetilde{N_E}$-통계량은 이 경우 검정력이 매우 우수함을 볼 수 있다. 또한 베타분포 $B(a,b)$의 경우 a가 작아질수록 검정력이 증가하고 있음을 볼 수 있다.

또한 대립가설이 DFR 분포일 경우에는 N_E, $\widetilde{N_E}$-통계량이 W_E-통계량보다 대체적으로 (Weib(0.8) 제외) 우수한 검정력을 보여주고 있고 N_E와 $\widetilde{N_E}$는 거의 비슷한 양상을 나타낸다. 그러나 대립가설이 IFR 분포일 경우에는 W_E-통계량이 특히 표본크기가 작을 때 ($n=10$, 20) N_E, $\widetilde{N_E}$ 보다 훨씬 우수한 검정력을 보여주고 $\widetilde{N_E}$가 N_E보다 고려된 모든 경우에 좋은 검정력을 보여준다.

	하방 α 분위수										
n	0.005	0.01	0.025	0.05	0.1	0.5	0.9	0.95	0.975	0.99	0.995
	통계량 $(n-1)^2 N_E$										
10	4.87	5.20	5.76	6.13	6.55	7.39	7.70	7.75	7.79	7.81	7.83
20	13.10	13.59	14.45	15.07	15.68	16.85	17.26	17.32	17.37	17.41	17.44
50	40.95	41.95	43.15	44.01	44.78	46.18	46.68	46.76	46.81	46.86	46.90
100	89.90	90.95	92.32	93.24	94.10	95.68	96.26	96.35	96.41	96.48	96.52
	통계량 $(n-1)^2 \widetilde{N_E}$										
10	6.16	6.45	6.87	7.23	7.60	8.34	8.67	8.73	8.77	8.81	8.83
20	14.69	15.22	16.11	16.68	17.18	18.19	18.60	18.67	18.72	18.77	18.79
50	43.53	44.34	45.42	46.21	46.82	48.01	48.50	48.58	48.64	48.70	48.73
100	93.05	94.05	95.18	95.92	96.54	97.86	98.43	98.52	98.59	98.65	98.68

<표 1> 통계량 $(n-1)^2 N_E$과 $(n-1)^2 \widetilde{N_E}$의 하방 α 분위수

결론적으로 소표본이고 대립가설이 $C_V \geq 1$인 분포일 경우에는 W_E-통계량보다는 제안된 N_E, $\widetilde{N_E}$-통계량이 더 효율적이며 소표본이고 $C_V < 1$인 대립가설에서는 W_E가 $\widetilde{N_E}$, N_E보다 더 효율적이라고 할 수 있다.

N_E와 $\widetilde{N_E}$를 비교하면, 각각이 반 정도의 대립가설에서 다른 통계량보다 우수하지만, N_E가 우수할 경우에는 검정력의 차이가 매우 작고, 반면 $\widetilde{N_E}$가 우수할 경우에는 그 차이가 대체적으로 크다고 할 수 있다. 따라서 전체적으로 $\widetilde{N_E}$가 N_E보다 우수하다는 결론을 내릴 수 있다.

대립가설	표본크기(n)	W_E	N_E	$\widetilde{N_E}$
$B(1/2,3/2)$	10	8.72	13.10	17.22
	20	5.92	18.62	22.40
	50	4.36	40.82	57.40
	100	2.38	80.68	91.70
$B(1/4,5/12)$	10	14.92	33.46	47.94
	20	10.00	69.94	83.64
	50	6.82	99.82	100.00
	100	5.34	100.00	100.00
$B(1/8,9/56)$	10	21.16	70.48	86.16
	20	15.06	98.86	99.80
	50	10.70	100.00	100.00
	100	8.88	100.00	100.00
$\chi^2(4)$	10	18.92	2.24	5.02
	20	32.36	4.84	11.36
	50	68.52	36.58	53.00
	100	93.36	85.60	93.00
$U(0,1)$	10	48.66	4.54	17.24
	20	84.60	34.00	59.90
	50	99.94	97.90	99.92
	100	100.00	100.00	100.00
Weib(1.5)	10	22.34	1.76	4.64
	20	42.18	6.24	16.36
	50	85.12	47.54	68.36
	100	99.28	93.44	98.10
$(1/2)N$	10	16.50	1.62	4.68
	20	32.12	3.86	10.12
	50	68.36	20.60	41.60
	100	94.20	62.22	80.96
$\chi^2(1)$	10	25.64	38.26	39.98
	20	40.82	63.02	60.34
	50	75.00	92.10	92.50
	100	94.66	99.60	99.74
Weib(0.8)	10	25.42	25.54	25.26
	20	39.14	38.76	34.64
	50	74.28	61.22	57.96
	100	94.78	80.86	80.26
lognorm(1)	10	19.98	23.88	24.06
	20	29.14	34.34	32.98
	50	44.76	55.62	49.48
	100	60.80	72.92	69.58
$(1/2)C$	10	47.46	56.96	55.94
	20	73.26	80.88	78.48
	50	96.30	97.86	97.06
	100	99.84	99.94	99.92

<표 2> W_E, N_E, $\widetilde{N_E}$-통계량의 검정력 비교 (유의수준 $\alpha=0.1$)

References

[1] Chernoff, H., Gastwirth, J. L. and Johns, M. V. (1967). "Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation," *Annals of Mathematical Statistics*, 38, 52-73.

[2] D'Agostino, R. B. and Stephens, M. A. (1986). *Goodness-of-fit Techniques,* Marcel Dekker, New York.

[3] de Wet, T. and Venter, J. H. (1973). "A goodness of fit test for a scale parameter family of distributions," *South African Statistical Journal*, 7, 35-46.

[4] Ferguson, T. S. (1996). *A course in large sample theory,* Chapman & Hall.

[5] Gail, M. H. and Gastwirth, J. L. (1978a). "A scale-free goodness-of-fit test for the exponential distribution based on the Lorenz curve," *Journal of the American Statistical Association*, 73, 787-793.

[6] Gail, M. H. and Gastwirth, J. L. (1978b). "A scale-free goodness-of-fit test for the exponential distribution based on the Gini Statistic," *Journal of the Royal Statistical Society*, Series B, 40, 350-357.

[7] Jackson, O. A. Y. (1967). "An analysis of departures from the exponential distribution," *Journal of the Royal Statistical Society*, Series B, 29, 540-549.

[8] Shapiro, S. S. and Wilk, M. B. (1972). "An analysis of variance test for the exponential distribution (complete samples)," *Technometrics*, 14, 355-370.

[9] Spinelli, J. J. and Stephens, M. A. (1987). "Tests for exponentiality when origin and scale parameters are unknown," *Technometrics*, 29, 471-476.

[10] Stephens, M. A. (1978). "On the W test for exponentiality with origin known," *Technometrics*, 20, 33-35.