응용통계연구 제 **21**권 2호, *2008*년, *pp. 265–273*

중도절단자료에 대한 수정된 SHAPIRO-WILK 지수 검정*

김남현[1)]

요 약

본 논문에서는 Kim (2001a)에서 제안한 지수분포에서의 수정된 Shapiro와 Wilk (1972) W_E-통계량을 중도절단자료에 적용하였다. 검정통계량은 Samanta와 Schwarz (1988)에서 W_E-통계량을 중도절단자료에 대해 수정한 것과 같은 방법으로 정규화 등간격(normalized spacings)을 이용하여 수정하였다. 그 결과 제안된 통계량은 귀무가설에서 중도절단이 없는 경우와 같은 분포를 갖고 표본크기만 변하게 된다. 제안된 통계량의 검정력을 Samanta와 Schwarz (1988)의 통계량과 비교한 결과, 중도절단이 없는 경우와 마찬가지로 중도절단이 있는 경우에도 변동계수가 1보다 크거나 같은 대립가설에서 제안된 통계량은 더 좋은 검정력을 나타내었다.

주요용어: 적합도검정, 지수검정, Shapiro-Wilk 통계량, 정규화 등간격.

1. 서론

$X_1, \ldots, X_n$을 연속확률분포함수 $F(x)$에서의 확률표본이고, 이 표본의 순서통계량을 $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$이라고 하자. $X_1, \ldots, X_n$이 지수분포의 모형에 적합한지, 즉

$$H_0 : F(x) = 1 - e^{-\frac{x-\alpha}{\beta}}, \quad x \geq \alpha \tag{1.1}$$

의 검정에 대해서 생각해 보자. 여기서 α와 β는 모두 미지라고 가정하자. 이 문제에 대한 전반적인 설명은 Ascher (1990), D'Agostino와 Stephens (1986), Doksum와 Yandell (1984), Spurrier (1984) 등을 참고로 한다.

위의 검정을 위한 대표적인 통계량 중의 하나는 Shapiro와 Wilk (1972)의 통계량인

$$W_E(n) = \frac{n(\bar{X} - X_{(1)})^2}{(n-1)S^2} \tag{1.2}$$

이다. 여기서 $\bar{X} = n^{-1}\sum_{i=1}^n X_i$, $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2$이다. $W_E(n)$-통계량은 모수 β의 두 추정량의 비를 기반으로 하는 통계량이며 이는 양쪽 검정통계량이다. $W_E(n)$의 귀무가설에서의 정확한 분포는 알려져 있지 않으나 Shapiro와 Wilk (1972)는 표본크기 n이 3부터 100일 때 W_E의 상방, 하방 분위수를 모의실험을 통하여 구하였다.

---

* 이 논문은 2007년도 홍익대학교 학술연구진흥비에 의하여 지원되었음.

1) (121-791) 서울시 마포구 상수동 72-1, 홍익대학교 기초과학과, 교수.
E-mail: nhkim@hongik.ac.kr

Spinelli와 Stephens (1987)에서 지적한 대로 $W_E(n)$-통계량은 대부분의 대립가설에서 좋은 검정력을 갖는다. 그러나 $W_E(n)$-통계량의 가장 심각한 단점은 이 검정법이 일치성(consistency)을 갖지 않는다는데 있다. 이는 D'Agostino와 Stephens (1986, 5.12절)도 언급하였다. Kim (2001a)에서는 $W_E(n)$-통계량의 이러한 단점을 보완한 통계량을 제안하였다.

또한 Samanta와 Schwarz (1988)은 $W_E(n)$-통계량을 중도절단 자료(censored data)에 대해서 수정하고 Brain과 Shapiro (1983)의 통계량과 비교하였다. r_1개의 최소자료와 r_2개의 최대자료가 절단되고 $X_{(r_1+1)} < X_{(r_1+2)} < \cdots < X_{(n-r_2)}$가 주어져 있을 때 이들이 제안한 통계량은 귀무가설에서 $W_E(n-r_1-r_2)$와 동일한 분포를 따른다.

본 논문에서는 Kim (2001a)에서 제안한 통계량을 Samanta와 Schwarz (1988)에서와 동일한 방법을 이용하여 중도절단 자료에 대한 통계량으로 확장하려 한다. 그리고 모의실험을 이용한 검정력 비교를 통하여 중도절단자료에서 나타나는 양상을 조사해보려고 한다.

2. 중도절단자료에서의 지수검정

앞에서 언급한 바와 같이 식 (1.2)의 $W_E(n)$ 통계량의 단점은 이 통계량을 이용한 검정법이 일치성을 갖지 않는다는 것이다. 즉, 이 검정법의 검정력이 표본의 수가 증가해도 1로 가까이 가지 않는 분포가 존재한다. 지수분포의 경우 모집단의 표준편차와 평균의 비인 변동계수(coefficient of variation) C_V가 $C_V = \sigma/\mu = 1$이므로

$$nW_E \xrightarrow{p} \frac{1}{C_V^2} = 1, \quad n \to \infty$$

이다. 그러므로 $C_V = 1$인 다른 분포에서도 $nW_E(n)$은 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 $(1/2, 3/2)$, $(1/4, 5/12)$, $(1/8, 9/56)$ 등이다.

Kim (2001a)에서는 $W_E(n)$의 이러한 단점을 보완하기 위해서 $W_E(n)$의 분모인 S^2을 β의 점근유효추정량(asymptotically efficient estimator, Kim (2001a)의 정리 3, 정리 4)인

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

또는

$$\widetilde{L}_n = \frac{1}{n-1}\sum_{j=2}^{n}\frac{(X_{(j)}-X_{(1)})^2}{\widetilde{v}_{jn}}, \quad \widetilde{v}_{jn} = \sum_{k=1}^{j}\frac{1}{n-k+1}$$

으로 대치한

$$N_E(n) = \frac{n(\bar{X}-X_{(1)})^2}{(n-1)^2 L_n} = \frac{n(\bar{X}-X_{(1)})^2}{(n-1)\sum_{j=2}^{n}\frac{(X_{(j)}-X_{(1)})^2}{v_{jn}}} \tag{2.1}$$

와

$$\widetilde{N}_E(n) = \frac{n(\bar{X} - X_{(1)})^2}{(n-1)^2 \widetilde{L}_n} = \frac{n(\bar{X} - X_{(1)})^2}{(n-1)\sum_{j=2}^{n} \frac{(X_{(j)} - X_{(1)})^2}{\widetilde{v}_{jn}}} \tag{2.2}$$

를 제안하였다. $V_{(1)} < \cdots < V_{(n)}$을 $F_0(x) = 1 - e^{-x}$, $x \geq 0$에서의 순서통계량이라고 할 때 $\widetilde{v}_{jn} = E(V_{(j)})$이고 v_{jn}은 이의 근사값으로 $v_{jn} = F_0^{-1}(j/(n+1))$이다. Kim (2001a)에서 언급한 바와 같이 일반적인 분포에서 $(n-1)N_E(n) \leq 1 + o_p(1)$이므로 $N_E(n)$-통계량은 적절한 c에 대해서 $N_E(n) < c$일 때 H_0를 기각하는 것이 합리적이다. $\widetilde{N}_E(n)$에 대해서도 마찬가지이다.

Kim (2001b)에서는 제안된 통계량의 극한분포를 브라운다리(Brownian bridge)의 적분의 형태로 구하였고, Kim (2002)에서는 이들 통계량에 근거한 검정법이 일치성(consistency)을 가짐을 살펴보았다.

본 논문에서는 식 (2.1), (2.2)의 N_E와 $\widetilde{N}_E$를 Samanta와 Schwarz (1988)의 방법을 적용하여 중도절단자료에 대한 통계량으로 확장하려고 한다.

우선 N-변환(normalized spacings, normalized waiting times)

$$T_j = (n-j+1)(X_{(j)} - X_{(j-1)}), \quad j = 2, \ldots, n$$

을 고려하자. 그러면 H_0에서 $T_2, \ldots, T_n$은 $i.i.d.$이고 분포함수 $F(x) = 1 - e^{-x/\beta}$, $x \geq 0$인 지수분포를 다른다. 식 (2.2)의 $\widetilde{N}_E(n)$를 T_j를 이용하여 다시 표현하면

$$\widetilde{N}_E(n) = \frac{\left(\sum_{j=2}^{n} T_j\right)^2}{n(n-1)\sum_{j=2}^{n}\left(\sum_{k=2}^{j}\frac{1}{n-k+1}T_k\right)^2 \Big/ \widetilde{v}_{jn}} = \frac{\left(\sum_{j=2}^{n} T_j\right)^2}{n(n-1)\sum_{j=2}^{n}\sum_{k=2}^{n} a_{jk}^{(n)} T_j T_k} \tag{2.3}$$

이고

$$\begin{aligned} a_{jk}^{(n)} &= \frac{1}{n-j+1}\frac{1}{n-k+1}\sum_{l=k}^{n}\frac{1}{\widetilde{v}_{ln}}, \quad j \leq k, \\ a_{kj}^{(n)} &= a_{jk}^{(n)} \end{aligned} \tag{2.4}$$

이다. 식 (2.3)은 제곱식을 전개하여 이차형식으로 다시 표현한 것이다. 만일 표본 중 가장 작은 r_1개와 가장 큰 r_2개가 중도절단되고 중간의 $n - r_1 - r_2$인 $X_{(r_1+1)} < X_{(r_1+2)} < \cdots < X_{(n-r_2)}$가 주어진다면 식 (2.3)의 $\widetilde{N}_E(n)$은

$$\widetilde{N}_E^C = \frac{\left(\sum_{j=2}^{n-r_1-r_2} T_{r_1+j}\right)^2}{(n-r_1-r_2)(n-r_1-r_2-1)\sum_{j=2}^{n-r_1-r_2}\left(\sum_{k=2}^{n-r_1-r_2} a_{jk}^{(n-r_1-r_2)} T_{r_1+j} T_{r_1+k}\right)} \tag{2.5}$$

표 3.1: 통계량 $\widetilde{N}_E(n)$의 하방 α 분위수

n \ α	0.01	0.025	0.05	0.10	0.15	0.25
10	0.0785	0.0838	0.0885	0.0936	0.0961	0.0990
15	0.0547	0.0585	0.0609	0.0629	0.0641	0.0656
20	0.0421	0.0446	0.0462	0.0476	0.0484	0.0492
25	0.0349	0.0362	0.0373	0.0383	0.0388	0.0394
30	0.0296	0.0307	0.0314	0.0321	0.0324	0.0329
35	0.0257	0.0266	0.0272	0.0277	0.0279	0.0282
40	0.0228	0.0234	0.0238	0.0242	0.0244	0.0247
45	0.0205	0.0210	0.0213	0.0216	0.0218	0.0220
50	0.0185	0.0189	0.0192	0.0195	0.0196	0.0198
70	0.0135	0.0137	0.0139	0.0140	0.0141	0.0142
85	0.0112	0.0114	0.0115	0.0116	0.0116	0.0117
100	0.0096	0.0097	0.0098	0.0099	0.0099	0.0099

로 수정할 수 있다. 여기서 $a_{jk}^{(n-r_1-r_2)}$는 식 (2.4)에 n 대신 $n-r_1-r_2$를 대입하여 얻어진다. $N_E(n)$도 똑같은 방법으로 수정하여 통계량 N_E^C를 얻을 수 있다. $\widetilde{N}_E^C$, N_E^C의 H_0에서의 분포는 각각 $\widetilde{N}_E(n-r_1-r_2)$, $N_E(n-r_1-r_2)$와 동일한 분포를 따른다. 앞에서 언급한 바와 같이 Kim (2001b)에서는 $N_E(n)$의 점근분포를 브라운 다리의 적분의 형태로 구하였다. $\widetilde{N}_E(n)$도 같은 점근분포를 갖는다. 유한한 n에 대해서는 모의실험을 통하여 귀무가설에서의 분포를 구하였고 $\widetilde{N}_E(n)$에 대한 결과를 표 3.1에 제시하였다. 식 (2.5)의 분모를 계산이 좀 더 용이한 형태로 다시 쓰면

$$m(m-1)\sum_{j=2}^{m}\frac{T_{r_1+j}}{m-j+1}\left(\frac{1}{m-j+1}\sum_{k=j}^{m}\frac{1}{\widetilde{v}_{r_1+k,m}}T_{r_1+j}+2\sum_{k=j+1}^{m}\frac{1}{m-k+1}\sum_{l=k}^{m}\frac{1}{\widetilde{v}_{r_1+l,m}}T_{r_1+k}\right)$$

이 된다. 여기서 $m=n-r_1-r_2$이다.

3. 모의실험 및 적용예제

식 (2.5)의 $\widetilde{N}_E^C$-통계량을 이용한 지수검정의 검정력을 알아보기 위하여 모의실험을 행하였다. 표 3.1은 통계량 $\widetilde{N}_E(n)$의 귀무가설에서의 분포의 하방 분위수를 각 표본크기에 대해서 모의실험을 통하여 구한 것이다. 이 결과는 지수분포에서 표본수 $N=10,000$을 이용하였다.

검정력 조사를 위하여 Kim (2001a)에서 고려한 것과 유사한 대립가설을 선택하였다. 따라서 표 3.3와 3.4의 대립가설의 확률밀도함수의 구체적인 형태는 Kim (2001a)을 참고한다. 표본크기는 Samanta와 Schwarz (1988)의 결과와 비교가능하도록 $n=30$으로 하였고 베타분포의 경우는 표본크기의 변화에 따른 검정력 변화를 보기 위하여 $n=50,100$도 고려하였다. 검정력 조사를 위한 표본수는 $N=2,500$을 사용하였다.

표 3.2: 자동차 고장시간자료

NA	16.3	17.3	18.6	22.9	NA	NA
NA	16.7	17.8	19.1	23.0	NA	NA
NA	16.9	17.9	19.5	24.6	NA	NA
NA	17.0	18.3	20.6	25.9	NA	NA
16.0	17.1	18.4	21.4	28.6	NA	NA

표 3.3: 통계량 $\widetilde{N}_E^C$, N_E^C, W_1의 검정력 (중도절단 없을 때, 오른쪽 중도절단의 경우)

중도절단 표본크기	$n = 30$	$r_1 = 0,\ r_2 = 0$			$r_1 = 0,\ r_2 = 5$			$r_1 = 0,\ r_2 = 10$		
	$n = 50$	$r_1 = 0,\ r_2 = 0$			$r_1 = 0,\ r_2 = 10$			$r_1 = 0,\ r_2 = 20$		
	$n = 100$	$r_1 = 0,\ r_2 = 0$			$r_1 = 0,\ r_2 = 15$			$r_1 = 0,\ r_2 = 30$		
대립가설	표본크기(n)	$\widetilde{N}_E^C$	N_E^C	W_1	$\widetilde{N}_E^C$	N_E^C	W_1	$\widetilde{N}_E^C$	N_E^C	W_1
	30	33	22	5	45	42	21	48	51	29
$B(1/2, 3/2)$	50	56	37	4	65	63	27	68	69	44
	100	92	80	2	88	87	53	92	91	61
	30	98	92	9	91	87	44	94	94	72
$B(1/4, 5/12)$	50	100	100	8	99	98	63	100	100	93
	100	100	100	4	100	100	78	100	100	98
	30	100	100	12	100	100	45	100	99	84
$B(1/8, 9/56)$	50	100	100	11	100	100	60	100	100	99
	100	100	100	6	100	100	68	100	100	98
$\chi^2(4)$	30	22	12	47	17	8	40	11	6	32
$U(0,1)$	30	88	67	96	38	18	67	12	5	35
$Weib(2.0)$	30	76	56	93	53	32	80	32	16	61
$(1/2)N$	30	18	8	47	10	4	27	6	3	18
$\chi^2(1)$	30	77	74	56	69	73	48	63	64	42
$weib(0.5)$	30	99	99	93	94	93	81	85	85	68
$lognorm(1)$	30	39	42	35	15	16	15	10	9	13
$(1/2)C$	30	89	88	87	45	49	37	19	20	17

표 3.3와 3.4은 $\widetilde{N}_E^C$, N_E^C와 Shapiro-Wilk 통계량을 중도절단자료로 일반화한 Samanta와 Schwarz (1988)의 통계량 W_1의 검정력을 제시한 것이다. 검정력은 $N = 2500$개의 표본 중 유의한 표본의 백분위를 소수 첫째자리에서 반올림한 것이다. 유의수준은 $\alpha = 0.1$을 이용하였다.

중도절단이 되지 않은 경우 ($r_1 = 0, r_2 = 0$)에는 고려한 모든 베타분포에서 Shapiro와 Wilk (1972)의 W_E(표의 W_1)-통계량은 표본크기가 커질수록 검정력이 감소하고 있다. 이러한 사실은 Kim (2001a)에서 이미 살펴본 바 있다. 오른쪽 중도절단의 경우에는 이러한 검정력 감소는 나타나지 않으나 $\widetilde{N}_E^C$나 N_E에 비해서 대체적으로 검정력이 현저히 떨어지는 것을 볼 수 있다. 왼쪽 중도절단의 경우에도 $B(1/2, 3/2)$ 외에는 유사한 현상을 볼 수 있다. 양쪽 중도절단의 경우에는 $r_1 = r_2 = 0$인 경우와 유사하게 표본크기가 증가해도 W_1은 검정력의 증가가 나타나지 않고 거의 일정하며 $\widetilde{N}_E^C$, N_E^C에 비해 검정력도 심하게 떨어진다.

대립가설이 $C_V > 1$인 감소고장률(decreasing failure rate: DFR) 분포인 $\chi^2(1)$, $weib(0.5)$,

표 3.4: 통계량 $\widetilde{N}_E^C$, N_E^C, W_1의 검정력 (왼쪽 중도절단, 양쪽 중도절단의 경우)

중도절단 표본크기	$n=30$ $n=50$ $n=100$	$r_1=5,\ r_2=0$ $r_1=10,\ r_2=0$ $r_1=15,\ r_2=0$			$r_1=10,\ r_2=0$ $r_1=20,\ r_2=0$ $r_1=30,\ r_2=0$			$r_1=2,\ r_2=3$ $r_1=5,\ r_2=5$ $r_1=7,\ r_2=8$			$r_1=5,\ r_2=5$ $r_1=10,\ r_2=10$ $r_1=15,\ r_2=15$		
대립가설	표본크기(n)	$\widetilde{N}_E^C$	N_E^C	W_1	$\widetilde{N}_E^C$	N_E^C	W_1	$\widetilde{N}_E^C$	N_E^C	W_1	$\widetilde{N}_E^C$	N_E^C	W_1
$B(1/2,3/2)$	30	11	5	10	8	3	20	25	21	10	16	16	10
	50	17	5	20	17	5	37	25	22	7	18	17	9
	100	50	23	23	49	17	55	49	42	13	28	24	10
$B(1/4,5/12)$	30	81	58	28	69	42	61	79	68	17	47	41	17
	50	97	87	51	95	80	90	90	80	11	54	46	19
	100	100	100	47	100	100	95	100	100	16	96	90	15
$B(1/8,9/56)$	30	100	99	33	99	95	74	100	99	22	94	87	25
	50	100	100	56	100	100	97	100	100	16	98	95	26
	100	100	100	51	100	100	97	100	100	18	100	100	16
$\chi^2(4)$	30	7	4	20	6	4	15	10	5	25	7	4	16
$U(0,1)$	30	80	52	93	61	33	85	51	26	77	23	9	50
$Weib(2.0)$	30	28	12	57	9	3	31	33	16	63	12	4	32
$(1/2)N$	30	11	4	33	7	2	21	10	4	27	7	3	19
$\chi^2(1)$	30	48	48	30	29	33	20	56	60	38	38	40	25
$weib(0.5)$	30	90	92	79	74	77	62	91	91	79	73	75	56
$longnorm(1)$	30	44	46	36	40	42	33	24	24	18	18	21	15
$(1/2)C$	30	87	87	85	83	85	80	59	65	54	40	46	34

$lognorm(1)$, $(1/2)C$에서는 $\widetilde{N}_E^C$, N_E^C-통계량이 W_1보다 좋은 검정력을 보여주나 대립가설이 $C_V < 1$인 증가고장률(increasing failure rate: IFR) 분포인 $\chi^2(4)$, $U(0,1)$, $weib(2.0)$, $(1/2)N$에서는 반대의 양상이 나타난다. 이는 중도절단의 형태와 관계없이 오른쪽, 왼쪽, 양쪽 중도절단 어느 경우에나 동일하다.

$\widetilde{N}_E^C$, N_E^C를 비교하면 중도절단의 형태나 대립가설의 형태에 관계없이 전반적으로 $\widetilde{N}_E^C$가 우수하거나 비슷한 검정력을 보여준다.

각각의 대립가설에서 중도절단 형태에 따라 나타나는 검정력의 양상은 세 통계량 모두 유사하다. 즉, 베타분포, $\chi^2(1)$, $\chi^2(4)$, $Weib(2.0)$ 등에서는 세 통계량 모두 오른쪽 중도절단의 경우 상대적으로 좋은 검정력을 보이며 $U(0,1)$, $lognorm(1)$, $(1/2)C$ 등에서는 왼쪽 중도절단의 경우 좋은 검정력을 보여준다.

결론적으로 대립가설이 $C_V \geq 1$인 분포에서는 W_1보다 제안된 $\widetilde{N}_E^C$-통계량이 효율적이며 $C_V < 1$인 대립가설에서는 W_1이 $\widetilde{N}_E^C$보다 효율적이라는 것을 중도절단자료에서도 확인할 수 있다.

다음으로 Brain과 Shapiro (1983)의 자동차 고장시간자료에 $\widetilde{N}_E^C$-통계량을 적용하여 보자. 환경수명검정을 위하여 35대의 자동차를 이용하였다. 4대는 고장시간 측정기에 연결하기도 전에 탈락하였다. 30시간이 지난 후 검사를 종료하였고 이때 10대의 차가 아직도 작동중이었다. 이 자료에 식 (1.1)의 이모수 지수분포모형이 적절한지 검정해 보자. 자료는 표 3.2에 제시하였다.

이 예제에서 $n = 35$, $r_1 = 4$, $r_2 = 10$, $n - r_1 - r_2 = 21$이고 $\widetilde{N}_E^C$의 값을 계산하면 $\widetilde{N}_E^C = 0.0434$이다. 따라서 p-값은 $p < 0.025$로 지수분포 모형에 대한 강한 반증을 보여준다. 이는 Brain과 Shapiro (1983), Samanta와 Schwarz (1988)의 결론과도 일치한다.

4. 결론

Kim (2001a)에서는 Shapiro와 Wilk (1972)의 지수분포의 검정을 위한 W_E-통계량의 단점을 보완한 수정된 통계량을 제안하였다. 본 논문에서는 이 통계량을 Samanta와 Schwarz (1988)의 방법을 이용하여 중도절단자료로 확장하였다.

Samanta와 Schwarz (1988)는 정규화 등간격을 이용하여 W_E-통계량을 중도절단자료에 대해 적용할 수 있는 W_1-통계량으로 수정하였다. 이들의 방법은 중도절단이 있는 경우에도 통계량의 귀무가설에서의 분포가 표본크기만 줄어들고 중도절단이 없는 경우와 동일하게 되므로, 중도절단자료에 대해서 수정된 통계량을 이용할 때 분포를 다시 구하지 않아도 된다는 장점이 있다.

제안된 통계량을 W_1-통계량과 비교한 결과 중도절단이 있는 경우에도 중도절단이 없는 경우와 마찬가지로 변동계수가 1보다 크거나 같은 대립가설에서 W_1보다 훨씬 좋은 검정력을 가짐을 볼 수 있었다.

참고문헌

Ascher, S. (1990). A survey of tests for exponentiality. *Communications in Statistics - Theory and Methods*, **19**, 1811–1825.

Brain, C. W. and Shapiro, S. S. (1983). A regression test for exponentiality: Censored and complete samples, *Technometrics*, **25**, 69–76.

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

Doksum, K. A. and Yandell, B. S. (1984). Tests for exponentiality, In *Handbook of Statistics 4: Nonparametric Methods*, (eds P. R. Krishnaiah and P. K. Sen), **8**, 579–612, North-Holland, Amsterdam.

Kim, N. (2001a). A modification of the W test for exponentiality, *The Korean Communications in Statistics*, **8**, 159–171.

Kim, N. (2001b). The limit distribution of a modified W-test statistic for exponentiality, *The Korean Communications in Statistics*, **8**, 473–481.

Kim, N. (2002). Consistency of a modified W test for exponentiality, *The Korean Communications in Statistics*, **9**, 629–637.

Samanta M. and Schwarz, C. J. (1988). The Shapiro-Wilk test for exponentiality based on censored data, *Journal of the American Statistical Association*, **83**, 528–531.

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

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

Spurrier, J. D. (1984). An overview of tests for exponentiality, *Communications in Statistics - Theory and Methods*, **13**, 1635–1654.

[2008년 1월 접수, 2008년 2월 채택]

A Modification of the Shapiro-Wilk Test for Exponentiality Based on Censored Data*

Namhyun Kim[1)]

ABSTRACT

Kim (2001a) presented a modification of the Shapiro and Wilk (1972) test for exponentiality based on the ratio of two asymptotically efficient estimates of scale. In this paper we modify this test statistic when the sample is censored. We use the normalized spacings based on the sample data, which was used in Samanta and Schwarz (1988) to modify the Shapiro and Wilk (1972) statistic to the censored data. As a result the modified statistics have the same null distribution as the uncensored case with a corresponding reduction in sample size. Through a simulation study it is found that the proposed statistic has higher power than Samanta and Schwarz (1988) statistic especially for the alternatives with the coefficient of variation greater than or equal to 1.

Keywords: Goodness of fit, exponentiality, Shapiro-Wilk statistic, normalized spacings.

* This work was supported by 2007 Hongik University Research Fund.

1) Professor, Dept. of Science, Hongik University, 72-1 Sangsu-dong, Mapo-gu, Seoul 121-791, Korea.
E-mail: nhkim@hongik.ac.kr