

쿨백-레이블러 정보함수에 기초한 와이블분포와 극단값 분포에 대한 적합도 검정 *

김종태¹⁾ 이우동²⁾

요약

공학의 응용분야인 신뢰수명론에서 와이블분포는 매우 중요한 역할을 해왔다. 그러나 와이블분포는 분포자체가 가지고 있는 형상모수의 영향으로 인하여 적합도 검정에 있어서 어려움의 대상이 되어 왔다. 이 논문은 쿨백-레이블러 정보 (Kullback-Leibler Information)을 이용한, 와이블 분포의 모수들에 영향을 받지 않은 검정 통계량을 제시함으로 위의 문제점을 해결하고, 제시된 검정 통계량에 대한 점근적 성질들과 검정력을 분석하였다. 제시된 검정 통계량은 기존의 검정 통계량들보다 검정력 비교에 있어서 더 우수한 검정력들을 보였고, 또한 실제 자료에 의한 적합도 검정의 예제를 보였다.

1. 서론

수많은 신뢰수명 분야에 관련된 문헌의 데이터들은 일반적으로 와이블분포를 따른다고 가정을 한다. 그리고 그러한 가정들이 참인지 아닌지를 검정하는 적합도 검정(goodness of fit test) 문제는 분석에 있어서 매우 중요하며 많은 학자들의 관심의 대상이다. 이러한 신뢰수명분야에 있어서 중요한 역할을 하는 와이블분포에 대한 적합도 검정을 위해서 다음과 같은 절차를 고려해 보자.

임의의 확률변수 T 가 확률밀도함수 $f(t; \cdot)$ 와 분포함수 $F(t; \cdot)$ 를 가진다고 가정 하자. 그리고 T_1, T_2, \dots, T_n 을 이 분포로부터의 확률표본이라 하자. 그러면, 주어진 확률표본이 모수 α, β 를 갖는 와이블분포(Weibull distribution)를 따르는지를 검정하기 위한 귀무가설은 다음과 같다.

$$H_0 : f(t; \cdot) = f_0(t; \alpha, \beta), \quad (1.1)$$

여기에서 모수 α, β 를 갖는 와이블분포의 확률밀도함수 $f_0(t; \alpha, \beta)$ 는 아래와 같다.

$$f_0(t; \alpha, \beta) = (\beta/\alpha)(t/\alpha)^{\beta-1} \exp\{-(t/\alpha)^\beta\}, \quad t > 0. \quad (1.2)$$

가설 (1.1)에 대한 대립가설은 아래와 같다.

$$H_a : f(t; \cdot) \neq f_0(t; \alpha, \beta). \quad (1.3)$$

* 이 논문은 1998년도 대구대학교 학술연구비 지원에 의한 연구임

1) (712-714) 경북 경산시 진량면 산15, 대구대학교 자연과학대학 통계학과, 조교수
2) (712-240) 경북 경산시 점촌동 산75, 경산대학교 자연과학대학 통계학과, 조교수

와이블분포의 모수들에 의한 영향력을 제거하기 위해 식 (1.2)에서 $X = \log T$ 로 변수변환을 시키면, X 에 대한 확률밀도함수는 아래의 극단값분포(extreme value distribution)가 된다.

$$f_X(x; a, b) = (1/b) \exp((x - a)/b) \exp(-\exp((x - a)/b)), -\infty < x < \infty. \quad (1.4)$$

여기서 $a = \log \alpha$ 이고 $b = 1/\beta$ 이다. 여기서 a 와 b 는 확률밀도함수 식 (1.4)에서 각각 위치모수(location parameter)와 척도모수(scale parameter)가 되므로 와이블분포가 가지는 모수로 인한 분석의 어려움은 제거되어 진다. 또한, $Y = (X - a)/b$ 로 변수변환을 한다면 Y 에 대한 확률밀도함수는 아래의 표준극단값분포(standard extreme value distribution)를 따른다.

$$f_Y(y) = \exp(y) \exp(-\exp(y)), -\infty < y < \infty. \quad (1.5)$$

그러므로 T 의 분포가 모수 α, β 를 갖는 와이블분포인지를 검정하는 적합도 검정은 확률변수

$$Y = (\log T - \log \alpha)/(1/\beta)$$

가 확률밀도함수 (1.5)를 갖는 표준극단값분포를 하는가를 검정하는 적합도 검정문제와 일치한다.

쿨백-레이블러 정보함수를 이용한 적합도 검정에 대한 연구는 Vasicek(1976), Arizono와 Ohta (1989), Dudewicz와 Van Der Meulen(1981) Ebrahimi, Habibullah 와 Soofi (1992), Soofi, Ebrahimi와 Habibullah(1995)등 많은 학자들에 의해 연구되었다. 2절에서는 와이블분포에 대한 쿨백-레이블러 정보함수의 추정량을 구하고 이 추정량에 대한 점근적 성질들을 조사한다. 3절에서는 2절에서 구한 추정량을 이용하여 검정 통계량을 제시하고 이 검정 통계량에 대한 기각값들을 구하였다. 4절에서는 열전달 장치의 고장시간에 대한 실제 데이터를 가지고 적합도 검정을 한다. 5절에서는 모의실험을 통해 제시한 검정 통계량과 기존의 검정 통계량들과 검정력측면에서 비교할 것이다. 마지막으로 6절에서는 결론과 문제점들을 제시한다.

2. 와이블 분포에 대한 쿨백-레이블러 정보함수 추정

쿨백-레이블러 정보함수 $I(f, f_0)$ 는 f 와 f_0 에 비의 척도로서의 정보를 제공하여 준다. 그러므로 와이블분포의 적합도 검정의 귀무가설 H_0 에 대한 검정에 있어서 쿨백-레이블러 정보함수가 큰 값을 가질때는 귀무가설 H_0 이 기각이 되어 진다. 이러한 이유로 와이블분포의 쿨백-레이블러 정보함수 추정량 그 자체가 와이블분포에 대한 적합도 검정의 검정 통계량으로서 역할을 한다. 2절에서는 와이블분포의 쿨백-레이블러 정보함수의 추정량을 구하고 그에 대한 점근적 성질들을 제시한다.

클백-레이블러 정보함수는 아래와같이 정의 되어 진다.

$$\begin{aligned} I(f, f_0) &= \int_{-\infty}^{\infty} \log\{f(t)/f_0(t)\}f(t)dt \\ &= \int_{-\infty}^{\infty} \log\{f(t)\}f(t)dt - \int_{-\infty}^{\infty} \log\{f_0(t)\}f(t)dt \\ &= -H(f) - \int_{-\infty}^{\infty} \log\{f_0(t)\}f(t)dt, \end{aligned}$$

여기서 $H(f) = -\int_{-\infty}^{\infty} \log\{f(t)\}f(t)dt = \int_0^1 \log\{\frac{d}{dp}F^{-1}(p)\}dp$ 로서 엔트로피(entropy)라고 한다. X_1, X_2, \dots, X_n 이 확률밀도함수 f 로 부터 추출된 확률표본이라 할 때, Vasicek(1976)은 엔트로피 $H(f)$ 에 대한 추정량 H_{mn} 을 아래와 같이 제안했다.

$$H_{mn} = (1/n) \sum_{i=1}^n \log\{\frac{n}{2m} \{X_{(i+m)} - X_{(i-m)}\}\}, \tag{2.1}$$

여기서 $X_{(i)}$ 는 i 번째 순서 통계량이고 m 은 $\frac{n}{2}$ 보다 작은 정수이며, 식 (2.1) 에서 만약 $j < 1$ 이면 $X_{(j)} = X_{(1)}$ 이고, $j > n$ 이면, $X_{(j)} = X_{(n)}$ 이다. 그리고 Vasicek(1976)은 H_{mn} 의 점근적 성질을 아래와 같이 밝혔다.

정리 2.1 확률변수 X 가 분포함수 F 와 확률밀도함수 f 를 갖고 $Var(X) < \infty$ 라 하자. X_1, X_2, \dots, X_n 이 F 로 부터 추출된 확률표본일 때, $n \rightarrow \infty, m \rightarrow \infty$ 에 대하여 $m/n \rightarrow 0$ 이면,

$$H_{mn} \xrightarrow{p} H(f).$$

이제 와이블분포의 확률표본 T_i 에 대해 $Y_i = (\log T_i - \log \alpha)/(1/\beta)$ 로 변수변환 시켜 만든 표준극단값분포에 대한 클백-레이블러 정보함수, $I(f, f_Y)$ 는 다음과 같이 표현된다.

$$\begin{aligned} I(f, f_Y) &= -H(f) - \int_{-\infty}^{\infty} \log\{f_Y(y)\}f(y)dy \\ &= -H(f) - E(Y) + E(e^Y). \end{aligned}$$

$H(f)$ 에 대한 추정량인 식 (2.1)의 추정량과 적률에 의한 추정량을 이용하여 $I(f, f_Y)$ 에 대한 추정량 I_{mn} 을 다음과 같이 제안 한다.

$$I_{mn} = -H_{mn}(Y) - \bar{Y} + \frac{1}{n} \sum_{i=1}^n e^{Y_i}. \tag{2.2}$$

다음의 정리들은 $I(f, f_Y)$ 에 대한 추정량 I_{mn} 에 대한 점근적 성질을 밝힌 것이다.

정리 2.2 와이블분포의 모수 α, β 를 알고 있을 경우, 변수변환에 의한 Y_1, Y_2, \dots, Y_n 이 표준극단값 확률밀도함수 식 (1.5)에서 추출된 확률표본이고, $n \rightarrow \infty, m \rightarrow \infty$ 에 대하여 $m/n \rightarrow 0$ 이면,

$$I_{mn} \xrightarrow{p} I(f, f_Y)$$

를 따른다.

증명: 정리 2.1의 결과와 $\bar{Y} \xrightarrow{p} E(Y)$, $\frac{1}{n} \sum_{i=1}^n \exp(Y_i) \xrightarrow{p} E(e^Y)$, 그리고 Slutsky 정리에 의하여 밝혀진다. \square

그러나 와이블분포의 모수들, α 와 β 를 모르는 경우에는 극단값분포 식 (1.4)의 모수 a 와 b 에 대한 추정량 \hat{a} 와 \hat{b} 를 구하여

$$Y^* = \frac{X - \hat{a}}{\hat{b}}$$

로 변수변환시킨 Y^* 를 이용하여 쿨백-레이블러 정보함수에 대한 추정량을 얻을 수 있다. 특히 와이블분포를 변환시킨 극단값분포의 모수 a , b 에 대한 최우추정량 (maximum likelihood estimates) \hat{a} , \hat{b} 에 대하여 최우추정량의 불변성 (invariance property)와 점근적 일치성 (consistency)을 이용하여 와이블분포의 모수를 모르는 경우에도 쿨백-레이블러 정보함수에 대한 추정량을 다음과 같이 제안 한다. $Y_i^* = \frac{X_i - \hat{a}}{\hat{b}}$, $i = 1, 2, \dots, n$ 이라 두고, Y_i^* 를 식 (2.2)의 Y_i 에 대입하자. 그리고

$$I_{mn}^* = -H_{mn}^* - \bar{Y}^* + \frac{1}{n} \sum_{i=1}^n e^{Y_i^*}, \quad (2.3)$$

여기서 $H_{mn}^* = (1/n) \sum_{i=1}^n \log\{\frac{n}{2m} \{Y_{(i+m)}^* - Y_{(i-m)}^*\}\}$ 이고 $\bar{Y}^* = (1/n) \sum_{i=1}^n Y_i^*$ 이다.

정리 2.3 와이블분포의 모수들인 α 와 β 를 모르는 경우, Y_1, Y_2, \dots, Y_n 이 확률밀도함수 식 (1.5)에서 추출된 확률표본 이고, \hat{a} , \hat{b} 가 각각 a , b 의 일치추정량이라고 하자. 만약 $n \rightarrow \infty$, $m \rightarrow \infty$ 일 때 $m/n \rightarrow 0$ 이면,

$$I_{mn}^* \xrightarrow{p} I(f, f_Y) \quad (2.4)$$

이 된다.

증명: \hat{a} , \hat{b} 가 a , b 의 일치추정량이므로,

$$H_{mn}^* \xrightarrow{p} H(f) \quad (2.5)$$

이며, 정리 2.2의 결과에 의해 식 (2.4)가 얻어진다. \square

3. 쿨백-레이블러 정보함수를 이용한 적합도검정

2절에서 제안된 I_{mn}^* 의 값이 클수록 적합도 검정의 귀무가설을 기각할 확률이 높아지므로 쿨백-레이블러 정보함수의 추정량은 적합도 검정의 검정 통계량으로 이용되어 진다. 이 절에서 먼저 쿨백-레이블러 정보함수의 추정량 I_{mn}^* 를 이용하여 와이블분포에 대한 적합도 검정 통계량 $KL_{mn} = \exp(-I_{mn}^*)$ 를 다음과 같이 제안한다. 그러므로 이때 제안된 검정 통계량 KL_{mn} 의 값이 적을 수록 귀무가설을 기각할 확률이 높아진다.

$$KL_{mn} = \exp(H_{mn}^* + \bar{Y}^* - \frac{1}{n} \sum_{i=1}^n e^{Y_i^*})$$

표 3.1: 기각값을 최대로하는 m

sample size window size	
n	m
5	2
6 - 19	3
20 - 34	4
35 - 64	5
65 - 70	6
71 - 94	7
95 - 120	8
121 - 130	9
131 - 190	10
191 - 200	11
201 - 230	12
230 -	15

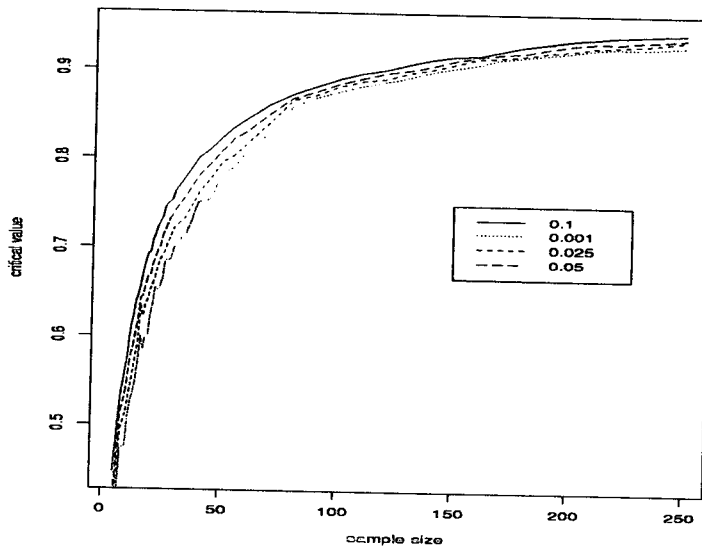


그림 3.1: KL_{mn} 의 기각치

표 3.2: $\alpha = 0.05$ 에서 KL_{mn} 의 기각치

sample size	window size	critical values					
		n	m	0.01	0.025	0.05	0.1
5	2			0.321430	0.368656	0.408796	0.446972
6	3			0.389221	0.422846	0.440466	0.472085
7	3			0.430790	0.455159	0.480774	0.506808
8	3			0.472440	0.495325	0.515181	0.532678
9	3			0.474279	0.502031	0.523212	0.548788
10	3			0.475986	0.512946	0.538906	0.565767
11	3			0.507675	0.527729	0.557704	0.584580
12	3			0.524246	0.544066	0.578191	0.609244
13	3			0.535997	0.560322	0.588742	0.621734
14	3			0.549154	0.578822	0.604336	0.639302
15	3			0.569340	0.592560	0.618288	0.645394
16	3			0.599751	0.634170	0.643280	0.658910
17	3			0.585534	0.622318	0.644456	0.671857
18	3			0.598272	0.634818	0.657599	0.682898
19	3			0.601549	0.641689	0.667019	0.693216
20	4			0.622000	0.652383	0.667294	0.694149
21	4			0.635827	0.659325	0.680486	0.707961
22	4			0.654373	0.667556	0.689078	0.715315
23	4			0.652145	0.677674	0.696193	0.724153
24	4			0.656383	0.685241	0.700522	0.729242
25	4			0.664202	0.692536	0.713228	0.734603
26	4			0.682748	0.695242	0.719416	0.744376
27	4			0.686403	0.704491	0.726222	0.748863
28	4			0.685139	0.709767	0.733068	0.751633
29	4			0.691369	0.717220	0.735790	0.754124
30	4			0.700754	0.723018	0.740087	0.762242
35	5			0.716758	0.735559	0.759459	0.780797
40	5			0.751377	0.761475	0.780242	0.800279
45	5			0.754547	0.776715	0.793594	0.810437
50	5			0.783181	0.795774	0.808484	0.823422

표 3.2: (계속)

sample size	window size	critical values			
n	m	0.01	0.025	0.05	0.1
55	5	0.789618	0.803162	0.822242	0.834911
60	5	0.805987	0.815223	0.829409	0.843670
70	6	0.829492	0.840031	0.849903	0.860592
80	7	0.859053	0.866274	0.868304	0.872809
90	7	0.869330	0.872318	0.876090	0.882064
100	8	0.876471	0.882340	0.885387	0.890861
110	8	0.883205	0.888408	0.891593	0.897152
120	8	0.887835	0.891125	0.897538	0.901544
130	9	0.893052	0.897856	0.901209	0.907929
140	10	0.899562	0.902717	0.906380	0.913415
150	10	0.904738	0.908936	0.912578	0.917194
160	10	0.908717	0.915187	0.917007	0.918304
170	10	0.914710	0.916683	0.920502	0.924214
180	10	0.917530	0.919161	0.922400	0.929901
190	10	0.920434	0.923248	0.927610	0.933175
200	11	0.923258	0.925570	0.931857	0.936832
210	12	0.927084	0.929463	0.934264	0.939231
220	12	0.929340	0.930839	0.934168	0.941556
230	12	0.928433	0.932473	0.936345	0.942854
240	15	0.930122	0.934946	0.937808	0.944119
250	15	0.931915	0.938612	0.940289	0.945249

표 3.1는 제안된 검정통계량 KL_{mm} 에 대한 기각값을 구하기 위해 각각의 와이블분포의 난수에 대하여 20,000번의 반복횟수로 기각값중 검정력을 최대로 하는, 즉 가장 큰 기각값을 갖는 m 의 값을 모의실험(Monte-Carlo simulation)을 통하여 얻은 결과이다. 이때 유의수준 α 은 0.01, 0.05와 0.1을 두었으나 똑 같은 결과의 m 의 값을 얻었다.

표 3.2은 표 3.1에 의해 구한 m 의 값을 이용하여 각각의 표본 크기에 대하여 신뢰계수 $\alpha = 0.05$ 에 대한 반복횟수 20,000번의 모의실험을 결과를 통하여 얻은 와이블분포의 적합도 검정의 기각값들이다. 표 3.2에 나타난 기각값들은 여러가지 와이블분포의 모수들을 변형시켜 모의실험을 해도 거의 똑같은 안정된 기각값을 가진다. 표본의 크기에 따른 기각치의 변화를 그림 3.1으로 나타내었다.

다음 절에서는 이절에서 제시한 검정통계량과 기각값을 가지고 실제 예제에 대한 적용을 할 것이다.

4. 예제와 검정력 비교를 위한 모의실험

다음의 예제는 Martz 와 Wallner(1982)의 395쪽에 있는 와이블분포를 따른다고 알려진 가솔린 정류에서의 열전달 장치기구에 대한 9개의 고장시간이다($n = 9$).

0.41 0.58 0.75 0.83 1.00
1.08 1.17 1.25 1.35

이 자료를 기초로 하여 모수를 모르는 경우, 와이블분포를 따르는지를 적합도 검정을 해보자. 먼저 모수 α 와 β 에 대한 최우추정량을 구해보면 $\hat{\alpha} = 1.0408$, $\hat{\beta} = 3.6825$ 이다. 이 때, $n = 9$ 에 해당하는 m 의 값은 표 3.1에서 3이므로 $KL_{mn} = 0.65142$ 이며, 표 3.2에서 유의수준 0.05에서의 기각치가 0.523212 이므로 “위의 자료는 와이블분포를 따른다”는 귀무가설을 기각 하지 못한다. 그러므로 쿨백-레이블러 정보함수의 추정을 이용한 적합도 검정의 결과는 “열전달장치기구에 대한 고장시간은 와이블 분포를 따른다”는 결과를 얻는다.

다음의 자료는 Mann,Schafer과 Singpurwalla(1974)의 215쪽에 수록된 6개의 알루미늄 조각에 대한 부식시간을 조사한 자료이다.($n = 6$)

3.45 5.00 5.90 7.30 8.60 10.20

위의 자료를 이용한 모수의 최우추정값은 각각 $\hat{\alpha} = 7.532$, $\hat{\beta} = 3.364$ 로 추정되었으며, $n = 6$ 에 해당하는 m 의 값은 3이고, KL_{mn} 통계량의 값은 0.572이다. 유의수준 0.05에서 기각치가 0.440466이므로 “위의 자료는 와이블분포를 따른다”는 귀무가설을 기각 하지 못한다.

제안된 검정통계량 KL_{mn} 과 기존의 검정통계량들, Kolmogorov-Smirnov (D), Kuiper (V), Cramer von Mises (W^2), Watson Statistics (U^2), Anderson-Darling (A^2), Finkelstein and Schafer (S^*) 검정통계량과의 검정력을 비교하기 위하여 모의실험을 하였다.

또한 대립가설의 분포로는 Uniform 분포 ($U(a, b)$), Inverse Gaussian 분포 ($IG(\mu, \lambda)$), Log-normal 분포 ($LN(\mu, \sigma)$), Gamma 분포 ($\Gamma(\gamma, \delta)$)를 두었다.

각각의 대립가설의 분포에 있어서 표본의 계수 $n = 10, 20, 30$ 을 주었고 모의실험의 반복횟수는 10,000번을 하여 신뢰계수 $\alpha = 0.05$ 로 검정력을 계산하였다. 모의실험의 결과는 다음과 같다. Uniform 분포는 표 4.1에, Inverse Gaussian 분포는 표 4.2에, Log-normal 분포는 표 4.3에, Gamma 분포는 표 4.4에 그 결과를 작성 하였다.

표 4.1의 Uniform 분포와, 표 4.2의 Inverse Gaussian 분포, 표 4.4의 Gamma 분포의 검정력 비교에 있어서 제안된 검정통계량 KL_{mn} 은 기존의 다른 통계량의 검정력보다 뛰어난 검정력을 가짐을 알 수 있다. 그러나 표 4.2의 Inverse Gaussian 분포에 있어서는 추출된 표본의 계수가 적은 경우 ($n = 10$)에는 제안된 검정 통계량의 검정력이 우수하지만 표본의 계수가 점점 크질수록 Cramer von Mises(W^2), Watson (U^2), Anderson-Darling (A^2), Finkelstein and Schafer (S^*)의 검정 통계량이 더 우수함을 알 수 있다.

표 4.1: Uniform 분포에 대한 검정력 비교 $H_a : X \sim U(a, b)$

n		KL_{mn}	W^2	U^2	A^2	D	V	S^*
10	$U(0, 1)$	0.14270	0.15600	0.13940	0.19210	0.13930	0.11870	0.15920
	$U(0.3, 1.7)$	0.11770	0.08600	0.09050	0.08970	0.07060	0.08890	0.08810
	$U(0.5, 1.5)$	0.13130	0.08810	0.09120	0.08670	0.06960	0.09030	0.09010
20	$U(0, 1)$	0.40070	0.30580	0.27060	0.38970	0.25340	0.22440	0.32000
	$U(0.3, 1.7)$	0.34820	0.15290	0.16930	0.19010	0.11530	0.16260	0.17690
	$U(0.5, 1.5)$	0.38830	0.15400	0.16550	0.17790	0.11390	0.16590	0.17270
30	$U(0, 1)$	0.60210	0.44460	0.39860	0.55740	0.34580	0.32760	0.46550
	$U(0.3, 1.7)$	0.56500	0.24940	0.26840	0.30750	0.16260	0.24590	0.27710
	$U(0.5, 1.5)$	0.62330	0.24710	0.26250	0.29840	0.16160	0.24460	0.27740

표 4.2: Inverse Gaussian 분포에 대한 검정력 비교 $H_a : X \sim IG(\mu, \lambda)$

n		KL_{mn}	W^2	U^2	A^2	D	V	S^*
10	$IG(1, 0.5)$	0.18860	0.13820	0.13830	0.12470	0.11090	0.12530	0.14740
	$IG(1, 0.3)$	0.19990	0.14160	0.14190	0.12930	0.11020	0.12610	0.15060
	$IG(1, 0.1)$	0.23230	0.16750	0.16530	0.15450	0.13000	0.14730	0.17710
20	$IG(1, 0.5)$	0.28560	0.26570	0.25430	0.27690	0.19780	0.21920	0.28230
	$IG(1, 0.3)$	0.30910	0.28970	0.27760	0.29940	0.21290	0.24000	0.30610
	$IG(1, 0.1)$	0.38880	0.35150	0.33360	0.36870	0.25670	0.29890	0.37190
30	$IG(1, 0.5)$	0.44860	0.41910	0.39800	0.46340	0.31120	0.36440	0.43620
	$IG(1, 0.3)$	0.49790	0.45890	0.43390	0.50290	0.33560	0.39400	0.47520
	$IG(1, 0.1)$	0.59710	0.53130	0.50810	0.58710	0.39600	0.47310	0.55490

표 4.3: Log-normal 분포에 대한 검정력 비교 $H_a : X \sim LN(\mu, \sigma)$

n		KL_{mn}	W^2	U^2	A^2	D	V	S^*
10	$LN(-0.5, 1)$	0.12860	0.11010	0.11110	0.09830	0.09470	0.10080	0.11210
	$LN(-0.25, 0.5)$	0.12750	0.10950	0.11100	0.09740	0.09210	0.09980	0.11160
	$LN(-0.1, 0.2)$	0.12650	0.11060	0.11120	0.09730	0.09500	0.10130	0.11260
20	$LN(-0.5, 1)$	0.18750	0.20780	0.20360	0.21810	0.15930	0.18070	0.21910
	$LN(-0.25, 0.5)$	0.18860	0.20530	0.20120	0.21600	0.15640	0.17880	0.21630
	$LN(-0.1, 0.2)$	0.18700	0.20380	0.19960	0.21380	0.15260	0.17550	0.21560
30	$LN(-0.5, 1)$	0.26670	0.30850	0.29480	0.33640	0.22520	0.26160	0.32350
	$LN(-0.25, 0.5)$	0.26550	0.31010	0.29690	0.33940	0.22580	0.26170	0.32550
	$LN(-0.1, 0.2)$	0.26710	0.31140	0.29700	0.33980	0.22630	0.26310	0.32600

표 4.4: Gamma 분포에 대한 검정력 비교 $H_a : X \sim \Gamma(\gamma, \delta)$

n		KL_{mn}	W^2	U^2	A^2	D	V	S^*
10	$\Gamma(10, 10)$	0.09090	0.07190	0.07190	0.06380	0.06760	0.06610	0.07230
	$\Gamma(15, 15)$	0.10860	0.08670	0.08760	0.08000	0.07530	0.08020	0.09120
	$\Gamma(20, 20)$	0.12110	0.08230	0.08620	0.07840	0.07010	0.08010	0.09020
20	$\Gamma(10, 10)$	0.09730	0.11910	0.11710	0.11690	0.10520	0.10910	0.12770
	$\Gamma(15, 15)$	0.12380	0.15380	0.14900	0.16180	0.12050	0.12210	0.16850
	$\Gamma(20, 20)$	0.17600	0.15420	0.14840	0.16020	0.10940	0.12510	0.16990
30	$\Gamma(10, 10)$	0.12200	0.17000	0.16460	0.17580	0.13590	0.15000	0.17400
	$\Gamma(15, 15)$	0.17000	0.20250	0.19620	0.21500	0.15320	0.17010	0.21220
	$\Gamma(20, 20)$	0.20590	0.19480	0.18300	0.22290	0.12420	0.15270	0.20720

5. 결론과 문제점

와이블분포의 클백-레이블러 정보함수의 추정을 이용하여 제시한 검정 통계량은 실제 자료의 정보가 와이블분포를 따른다는 가정하에서 모수에 대한 최우추정량을 이용함으로써 적합도 검정의 귀무가설이 채택되어 진다면 그 자료에 대한 와이블 분포의 모수로서 최우추정량의 값을 대체함으로 자료의 정확한 분포의 모양을 찾을 수 있다.

모의실험에 있어서 Log-normal 분포를 제외한 다른 분포들에 있어서 제안된 검정통계량의 값이 우수하였다. 그러나 여전히 의문점이 남는 것은 왜 Log-normal 분포에서 Cramer von Mises 형태의 검정통계량이 더 우수인가에 대한 결론은 여전히 해결하지 못하였다.

참고문헌

- [1] Arizono, I. and Ohta, H. (1989). A test for normality based on Kullback-Leibuller information, *The American Statistician* . Vol. 43. 20-22.
- [2] Dudewicz, E. J. and Van Der Meulen, E. C. (1981). Entropy-based tests of uniformity, *Journal of the American Statistical Association* . Vol. 76. 967-974.
- [3] Ebrahimi, N., Habibullah, M., and Soofi, E.S. (1992). Testing exponentiality based on Kullback-Leibler Information, *Journal of Royal Statistical Society, Ser B* . Vol. 54. 739-748.
- [4] Mann, N. R., Schafer, R. E. and Singpurwalla, N. D. (1974). *Methods for Statistical Analysis of Reliability and Life Data*, John Wiley & Sons, Inc.
- [5] Martz, F.M. and Wallner R.A. (1982). *Bayesian Reliability Analysis*, John Wiley & Sons, Inc.
- [6] Soofi, E. S., Ebrahimi, N., and Habibullah, M. (1995). Information distinguishability with application to analysis of failure data, *Journal of the American Statistical Association* . Vol. 90. 657-668.
- [7] Vasicek, O. (1976). A test for normality based on sample entropy, *Journal of Royal Statistical Society, Ser B* . Vol. 38. 54-59.

[1997년 12월 접수, 1998년 4월 최종수정]

**A Test for Weibull Distribution and Extreme Value
Distribution
Based on Kullback-Leibler Information ***

Jong Tae Kim ¹⁾ Woo Dong Lee ²⁾

ABSTRACT

In this paper, a test of fit for Weibull distribution on the estimated Kullback-Leibler information is proposed. The test uses the Vasicek entropy estimates, so to compute it a window size m must first be fixed, and then is obtained critical values computed by Monte Carlo simulations. The power of the proposed test under various alternatives is compared with that of other famous tests. The use of the test is shown in an illustrative example.

* This paper was supported by research fund, Taegu University, 1998

1) Dept of Statistics, Taegu University, Kyunsan, Kyungpook 712-714, Korea.

2) Dept of Statistics, Kyungsan University, Kyungsan, Kyungpook 712-240, Korea.