Journal of Korean
Data & Information Science Society
2006, Vol. 17, No. 1, pp. 31 ~ 39

A Comparison on the Empirical Power of Some Normality Tests

Daehak Kim[1)] · Jun Hyeok Eom[2)] · Heong Chul Jeong[3)]

Abstract

In many cases, we frequently get a desired information based on the appropriate statistical analysis of collected data sets. Lots of statistical theory rely on the assumption of the normality of the data. In this paper, we compare the empirical power of some normality tests including sample entropy quantity. Monte carlo simulation is conducted for the calculation of empirical power of considered normality tests by varying sample sizes for various distributions.

Keywords : 경험적 검정력, 소표본 모의실험, 정규성 검정, 표본 엔트로피, 쿨백 라이블러 정보량

1. 서 론

오늘날의 지식, 정보 사회에서는 엄청난 자료들이 생성되고 있다. 그리고 우리는 정치, 경제, 사회, 자연과학 등의 여러 분야에서 이들 자료들 중 필요한 자료들만을 수집해서 적절한 통계적 분석을 통하여 유용한 정보를 얻고 있다. 이러한 과정에서 자료들을 수집하고 분석을 실시하기 전에 수집된 자료들이 통계적 정규성을 따르는지의 여부를 알아야 할 경우가 자주 발생한다.

대부분의 많은 통계적 분석들은 자료들이 좌우대칭인 종 모양의 형태를 갖는 정규분포를 따른다는 가정 하에 분석이 시행된다. 따라서 정규성 검정은 분석의 타당성을 높이기 위해 필요한 절차 중의 하나로서 현재까지 많은 검정방법들이 제안되어 왔다.

본 논문에서는 정규성 검정방법들의 여러 가지 상황에서의 검정력을 살펴보았다. 대표적인 정규성 검정방법인 카이제곱 검정을 포함하여, Kolmogorov-Smirnov 검정, Anderson-Darling 검정, Shapiro-Wilk 검정 등을 포함하고 또한 엔트로피를 이용한

---

1) 제 1 저자 : 교수, 경북 경산시 하양읍 대구가톨릭대학교 정보통계학과
E-mail : dhkim@cu.ac.kr

2) 대구가톨릭대학교 정보통계학과 석사

3) 조교수, 경기도 화성시 봉담읍 수원대학교 통계정보학과,
E-mail: jhc@suwon.ac.kr

표본 Entropy 검정, Kullback-Leibler 검정 등도 고려함으로서 종합적인 비교를 시도하였고 표본의 크기, 모집단의 분포 등을 다양하게 변화시키면서 경험적 검정력을 모의실험을 이용하여 구하였다. 2장에서는 정규성 검정방법들에 대하여 소개하였고, 3장에서는 실제자료에 대한 검정결과를 예를 들어 설명하였고 4장에서는 다양한 분포와 표본의 크기에 대하여 모의실험을 실시하였고 정규성 검정방법들의 경험적 검정력 비교와 표본 크기에 따른 경험적 검정력의 변화도 고려되었다.

2. 정규성 검정방법들

$X_1, X_2, \cdots, X_n$을 미지의 누적분포함수 $F(x)$로부터의 랜덤표본이라 하고 또 $Y_1 \leq Y_2 \leq \cdots \leq Y_n$을 랜덤표본 $X_1, X_2, \cdots, X_n$의 순서통계량이라 두자. 그리고 $F_n(x)$를 랜덤표본 $X_1, X_2, \cdots, X_n$에 기초한 경험적 분포함수라 두자. 즉

$$F_n(x) = \frac{1}{n} I_{(-\infty, x]}(X_i)$$

로 표현된다. 이때 $I(\cdot)$는 지시함수이다. 또 $F_0(x)$를 표준정규분포의 누적분포함수라 두고 다음과 같은 가설검정을 고려하자.

$$\begin{aligned} &H_0(\text{귀무가설}) : F(x) = F_0(x) \\ &H_1(\text{귀무가설}) : F(x) \neq F_0(x) \end{aligned} \tag{2.1}$$

즉, 주어진 자료들이 정규분포를 따른다는 귀무가설과 그렇지 않다는 대립가설에 대하여 가설검정문제를 고려하자. 이때 편의상 자료들은 평균이 0, 분산이 1을 갖도록 표준화되었다고 가정하고 정규성만을 검정하기로 하자. 잘 알려진 대표적인 정규성 검정방법들을 차례로 살펴보자.

카이제곱 검정은 Pearson(1900)에 의해 제안되었다. 가설 (2.1)에 대해 n개의 관측값을 c개의 계급으로 분류한 후 각 계급에서의 기대도수 $E_1, \cdots, E_c$와 각 계급에서의 관측도수 $O_1, \cdots, O_c$로부터 검정통계량

$$T = \sum_{j=1}^{c} \frac{(O_j - E_j)^2}{E_j} \tag{2.2}$$

을 계산하여 $T \geq \chi^2(\alpha, c-1)$이면 귀무가설 H_0를 기각한다. 귀무가설 H_0하에서 검정통계량 T는 근사적으로 자유도 $(c-1)$인 χ^2분포를 따름이 알려져 있다.

Smirnov(1939)에 의해 제안된 Kolmogorov-Smirnov검정방법은 경험적 분포와 표준정규분포 사이의 적합도를 검정하는 방법이다. 경험적 분포함수 $F_n(x)$가 이론적 분포함수 $F(x)$로 수렴하는 성질을 이용하여 $F_0(x)$와 $F_n(x)$로 각각 표기되는 표준정규분포와 경험적 분포의 근접도에 기초해서 적합성 검정을 구성하는 것은 타당하며

모든 $|F_n(x)-F_0(x)|$들의 최소상한인 검정통계량 D_n

$$D^{\cdot}n=\sup_x[\,|F^{\cdot}_n(x)-F_0(x)|\,] \tag{2.3}$$

을 이용하여 D_n의 값이 임계값보다 작을 경우 귀무가설 H_0를 기각하지 않는다. 이때 사용되는 임계값은 유의수준과 표본크기에 따라 달라진다.

Anderson-Darling 검정은 Nelson(1998)과 Stephens(1974), 그리고 Stephens(1982)에 의해 연구된 바 있다. 알려진 분포의 모집단으로부터 얻어진 자료인 경우에 유용한 검정방법으로 알려져 있으며, Kolmogorov-Smirnov 검정의 변형이라 할 수 있다. Anderson-Darling 검정통계량은

$$A^2=-n-S \tag{2.4}$$

이다. 이때 S는 알려진 분포의 누적분포함수 F와 순서통계량 Y_i에 대하여

$$S=\sum_{i=1}^{n}\frac{(2i-1)}{N}\left[\ln F(Y_i)+\ln\left(1-F(Y_{n+1-i})\right)\right]$$

이다. 실제로 정규성 검정에 이용되는 검정통계량은 A^2통계량을 수정한

$$A^*=A^2\left(1+\frac{0.75}{N}+\frac{2.25}{N^2}\right) \tag{2.5}$$

를 사용하며 통계량 A^*의 값이 임계값보다 작으면 가설 (2.1)의 귀무가설 H_0를 기각하지 않게 된다. 이 검정의 임계값은 유의수준에 관계없이 하나의 임계값만을 가지는 특징이 있다.

Shapiro와 Wilk(1965)에 의해 제안된 Shapiro-Wilk 검정은 적합도 검정과의 비교연구에서 가장 많이 사용되고 있는 검정방법이다. 표본의 기저분포에 관계없이 분포무관한 검정으로서 통계량

$$W=\frac{b^2}{S^2}=\left(\sum_{i=1}^{n}a_iY_i\right)^2\Big/\sum_{i=1}^{n}(Y_i-\overline{Y})^2 \tag{2.6}$$

을 이용한다. 여기서 a_i는 정규분포에서 크기 n인 표본의 순서통계량의 평균과 분산, 공분산으로부터 만들어진 상수들이고

$$b=\sum_{i=1}^{n}a_{n-i+1}(Y_{n-i+1}-Y_i),\ S^2=\sum_{i=1}^{n}(Y_i-\overline{Y})^2$$

이다. W의 값이 , 임계값보다 작으면 귀무가설 H_0를 기각하게 된다.

표본 Entropy 검정은 Vasicek(1976)에 의해 제안되고 Arizono와 Ohta(1989)에 의해 연구된 바 있다. 엔트로피는 확률밀도함수 $f(x)$에 대하여

$$H(f) = -\int_{-\infty}^{\infty} f(x)\ln f(x)\,dx$$

로 정의되고 누적분포함수 $F(\cdot)$를 이용하면

$$H(f) = \int_0^1 \ln\left\{\frac{d}{dp}F^{-1}(p)\right\}dp$$

로 나타난다. 엔트로피의 추정은 경험적 분포함수의 역함수를 이용하는데 $F^{-1}(p)$는

$$(Y_{i+m} - Y_{i-m})n/2m,\ (i-1)/n < p \le i/n,\ i = m+1, m+2, \cdots, n-m$$

로 추정된다. 여기서 m은 $n/2$보다 작은 양의 정수이다. 이를 이용하여 엔트로피 $H(f)$의 추정치는

$$\hat{H}(f) = n^{-1}\sum_{i=1}^{n}\ln\left\{\frac{n}{2m}(Y_{i+m} - Y_{i-m})\right\} \tag{2.7}$$

로 된다. 이때 $Y_i = Y_1 (i < 1)$이고 $Y_i = Y_n (i > n)$이다. 가설검정은 σ^2의 불편추정량 S^2을 이용한 통계량

$$\hat{E}(f) = \exp\{\hat{H}(f)\}/S \equiv \frac{n}{2mS}\left\{\prod_{i=1}^{n}(Y_{i+m} - Y_{i-m})\right\}^{1/n} \tag{2.8}$$

을 이용하여 수행된다. 이때

$$\overline{Y} = \sum_{i=1}^{n} Y_i/n,\ S^2 = \sum_{i=1}^{n}(Y_i - \overline{Y})^2/n$$

이고, 통계량 $\hat{E}(f)$은 귀무가설하에서

$$\hat{E}(f) \xrightarrow{p} \sqrt{2\pi e},\ n\to\infty,\ m\to\infty,\ m/n\to 0$$

로 된다. 귀무가설 H_0하에서 통계량 $\hat{E}(f)$은 분석적인 결과를 제공하지 못하므로 Vasicek(1976)은 모의실험을 통해서 $\hat{E}(f)$의 백분율을 이용하여 유의수준 α에서의 임계값을 구했다. $\hat{E}(f)$의 값이 임계값보다 작으면 귀무가설 H_0를 기각하게 된다.

Arizono와 Ohta(1989)에 의해 연구된 바 있는 Kullback-Leibler 검정은 평균과 분산의 두 모수를 포함하는 정규성 가설검정에 사용될 수 있는 특징을 가지고 있다.

Kullback-Leibler 정보량은

$$I(g:f) = \int_{-\infty}^{\infty} g(x)\ln\frac{g(x)}{f(x)}dx$$

로 정의되고 확률밀도함수 $g(x)$인 분포함수 $G(\cdot)$ G와 확률밀도함수 $f(x)$인 정규분포의 분포함수 $F(\cdot)$사이의 거리를 나타내며 $I(g:f) \geq 0$이고, 등호는 $g(x) = f9x)$의 경우에 한하여 유효하다. Kullback-Leibler 검정은 자료들이 평균이 μ이고 분산 σ^2인 정규분포 $N(\mu, \sigma^2)$로부터의 표본이다라는 귀무가설을 고려한다. 간단한 계산에 의해 $I(g:f)$는

$$I(g:f) = -H(g) + \ln\sqrt{2\pi\sigma^2} + \frac{1}{2}\int_{-\infty}^{\infty}\left(\frac{x-\mu}{\sigma}\right)^2 g(x)dx$$

로 되고 주어진 자료의 경험적 분포함수를 이용하여 $I(g:f)$를

$$\hat{I}(f,g) = \ln\left[\sqrt{2\pi e^2}\exp\left\{\frac{1}{2n}\sum_{i=1}^{n}\left(\frac{Y_i-\mu}{\sigma}\right)^2\right\}\Big/\left[\frac{n}{2m}\left\{\prod_{i=1}^{n}(Y_{i+m}-Y_{i-m})\right\}^{1/n}\right.\right. \tag{2.9}$$

로 추정하게 된다. 귀무가설 H_0하에서 $I(g:f) = 0$ 이고 또

$$\hat{I}(f:g)\xrightarrow{p}\sqrt{2\pi e},\ n\to\infty,\ m\to\infty,\ m/n\to 0$$

임을 이용하여 다음의 통계량 $\hat{K}$

$$\begin{aligned}\hat{K} &= \sqrt{2\pi}/exp\{\hat{I}(f,g)\} \\ &= n\left\{\prod_{i=1}^{n}\left(\frac{Y_{i+m}-Y_{i-m}}{\sigma}\right)\right\}^{1/n}\Big/\left[2m exp\left\{\frac{1}{2n}\sum_{i=1}^{n}\left(\frac{Y_i-\mu}{\sigma}\right)^2\right\}\right]\end{aligned} \tag{2.10}$$

을 유도할 수 있다. 귀무가설 H_0하에서

$$\hat{K}\xrightarrow{p}\sqrt{2\pi},\ n\to\infty,\ m\to\infty,\ m/n\to 0,\ 0 < \hat{K} \leq \sqrt{2\pi}$$

로 되고 모의실험을 이용해 백분율을 이용하여 Vasicek(1976)은 유의수준 α에서의 임계값을 구했다. 통계량 $\hat{K}$의 값이 임계값보다 작으면 귀무가설을 기각하게 된다.

3. 실제 사례를 통한 비교

2장에서 살펴본 정규성 검정방법들을 실제 자료를 이용하여 정규성 검정을 실시하여 보자. <표 3.1>은 Shapiro와 Wilk(1965)에 의해 사용된 남자의 몸무게(단위: 파운드b)를 측정한 자료이다. 모집단의 분포가 정규성을 따르는지 여러 가지 정규성 검정 방법으로 검정하여 보자. 이자료의 경우 표본의 크기 $n=11$이고 유의수준 $\alpha=.05$로 하였다.

<표 3.1> Shapiro와 Wilk(1965)의 몸무게 자료

표본수	몸무게(단위:파운드)										
11	148	154	158	160	161	162	166	170	182	195	236

먼저 카이제곱 검정을 실시하여 보자. 계급의 수 c를 8로 하고, (2.2)식의 통계량 T를 계산하면 11.385를 얻고 T의 값이 임계값 $\chi^2(.05,7)=114.07$ 보다 작으므로 H_0를 기각하지 못하게 된다.

Kolmogorov-Smirnov 검정의 경우는 (2.3)식을 이용하여 모든 $|F_n(x)-F_0(x)|$들의 최소 상한값인 D_n을 계산하면 D_n=0.2592을 얻고 통계량 D_n의 값이 임계값 0.39 보다 작으므로 H_0를 기각하지 못하게 된다.

Anderson-Darling 검정은 (2.4)식을 이용하여 $S=$-11.947과 통계량 A^2=0.947을 얻고 (2.5)식을 이용하여 A^2의 수정된 통계량 A^*=1.028을 얻는다. A^*의 값이 임계값 0.752 보다 크므로 귀무가설 H_0를 기각하게 된다.

Shapiro-Wilk 검정에서는 먼저 a_i의 값은 표본의 크기에 따라 Shapiro와 Wilk(1965)의 <표 5>를 참고하여 a_i=[0.5607, 0.3315, 0.2260, 0.1429, 0.0695, 0.00]를 얻고 (2.6)식을 이용하여 S^2=6,226과 b=70.081을 얻고 (2.6)식을 이용하여 통계량 W =0.7888을 얻는다. W의 값이 임계값 0.850보다 작으므로 가설 (2.1)의 귀무가설 H_0를 기각하게 된다.

또한 표본 Entropy 검정의 경우는 (2.7)식을 이용하여 $\hat{H}(f)$을 계산한 후에 (2.8)식을 이용하여 통계량 $\hat{E}(f)$을 계산하면 S=23.791, $\hat{H}(f)$=3.983, 그리고 $\hat{E}(f)$=2.2561을 얻게 된다. 통계량 $\hat{E}(f)$의 값이 m이 3일 때의 임계값 2.285보다 작으므로 가설 (2.1)의 귀무가설 H_0를 기각하게 된다.

마지막으로 Kullback-Leibler 검정을 살펴보면 (2.9)식을 이용하여$\hat{I}(f:g)$=0.6075를 얻고 (2.10)식을 이용하여 통계량 $\hat{K}$=1.3654를 얻는다. 통계량 $\hat{K}$의 값이 m이 3일 때의 임계값 1.15 보다 크므로 귀무가설 H_0를 기각하지 못하게 된다. 검정의 결과를 요약하여 <표 3.2>에 나타내었다.

<표 3.2> 검정결과 요약표

방법	카이제곱	K-S	A-D	S-W	S-E	K-L
통계량의 값	11.39	0.26	1.03	0.79	2.26	1.37
임계값	114.07	0.39	0.75	0.85	2.29	1.15
검정결과	채택	채택	기각	기각	기각	채택

여기서 K-S는 Kolmogorov-Smirnov검정을, A-D는 Anderson-Darling 검정을, S-W는 Shapiro-Wilk 검정을, S-E는 표본 Entropy 검정을 그리고 마지막으로 K-L은 Kullback-Leibler 검정을 의미한다.

위의 사례를 통하여 알 수 있듯이 같은 자료에 대하여 정규성 검정을 하였으나 검정방법에 따라 그 결과가 다르다는 것을 알 수 있다. 카이제곱 검정, Kolmogorov-Smirnov 검정, Kullback-Leibler 검정에서는 자료가 정규성을 따른다는 결과가 나왔으나, Anderson-Darling 검정, Shapiro-Wilk 검정, 표본 Entropy 검정에서는 정규성을 따르지 않는다는 상이한 결과가 나왔다. 이제 모의실험을 시행하여 검정방법에 따라 어떤 특징이 있는지 알아보도록 하겠다.

4. 모의실험과 그 결과

정규성 검정방법들의 경험적 검정력의 비교를 위하여 소표본 모의실험을 실행하였다. 모집단의 분포로서 표준정규분포, 로그정규분포(0,1), 자유도 2인 T분포, 자유도 2인 카이제곱분포, 일양분포(0, 1), 와이블분포(0.5, 1) 그리고 지수분포(1)등을 고려하였다. 표본의 크기는 각각의 분포에 대하여 10, 20, 30 그리고 40 으로 하였고 정확한 결과를 얻기 위하여 비교적 충분하다고 볼 수 있는 10,000번씩 반복 검정을 하였다. 유의수준은 0.05로 정하였다.

<표 4.1> $n=10,20$일 때 경험적 검정력

분포	카이제곱		K-S		A-D		S-W		S-E		K-L	
n	10	20	10	20	10	20	10	20	10	20	10	20
N	0.0017	0.0122	0.0010	0.0003	0.6099	0.0501	0.0527	0.0496	0.0494	0.0420	0.0030	0.0090
LN	0.0697	0.6761	0.0331	0.2307	0.6099	0.9149	0.6219	0.9400	0.6239	0.9244	0.3115	0.8270
T	0.0357	0.4001	0.0117	0.0861	0.3092	0.5307	0.2954	0.5328	0.1785	0.3093	0.0658	0.2176
χ^2_2	0.0288	0.4252	0.0062	0.0558	0.4515	0.7956	0.4737	0.8567	0.4973	0.8577	0.1600	0.6851
U	0.0009	0.0023	0.0010	0.0010	0.0782	0.1761	0.0781	0.1990	0.1551	0.4130	0.0126	0.1909
W	0.0235	0.4056	0.0035	0.0484	0.4191	0.7702	0.4433	0.8418	0.4664	0.8393	0.1361	0.6617
exp	0.0222	0.4039	0.0034	0.0441	0.4084	0.7834	0.4272	0.8369	0.4553	0.8399	0.1335	0.6602

<표 4.2> $n=30,40$일 때 경험적 검정력

분포	카이제곱		K-S		A-D		S-W		S-E		K-L	
n	30	40	30	40	30	40	30	40	30	40	30	40
N	0.0356	0.0580	0.0003	0.0002	0.9848	0.0472	0.0454	0.0439	0.4470	0.0463	0.0121	0.0132
LN	0.8840	0.9530	0.4512	0.6422	0.9848	0.9978	0.9912	0.9993	0.9868	0.9974	0.9658	0.9934
T	0.6360	0.7886	0.1670	0.2478	0.6844	0.7947	0.6617	0.7543	0.4731	0.6045	0.3712	0.5100
χ_2^2	0.6742	0.7898	0.1354	0.2406	0.9416	0.9859	0.9695	0.9955	0.9695	0.9945	0.9161	0.9796
U	0.0057	0.0222	0.0010	0.0008	0.2982	0.4308	0.4122	0.6830	0.6413	0.7933	0.4186	0.6045
W	0.6548	0.7872	0.1204	0.2314	0.9391	0.9871	0.9674	0.9970	0.9695	0.9951	0.9071	0.9797
exp	0.6492	0.7810	0.1240	0.2251	0.9330	0.9854	0.9683	0.9966	0.9687	0.9953	0.9097	0.9809

<표 4.1>과 <표 4.2>는 각 분포의 표본크기에 따른 검정방법들의 경험적 검정력을 나타낸 것이다. 여기서 3절의 경우와 마찬가지로 K-S는 Kolmogorov-Smirnov검정을, A-D는 Anderson-Darling 검정을, S-W는 Shapiro-Wilk 검정을, S-E는 표본 Entropy 검정을 그리고 마지막으로 K-L은 Kullback-Leibler 검정을 의미한다.

모의실험의 결과를 살펴보면 대체로 Anderson-Darling 검정, Shapiro-Wilk 검정, 표본 Entropy 검정에서 유의수준 0.05를 만족하고 있으며 또한 경험적 검정력이 다른 검정방법에 비해 상대적으로 높은 것을 알 수 있다. 모든 분포들에 대해서 표본이 커짐에 따라 경험적 검정력도 증가하는 것을 볼 수 있다. 카이제곱 검정에서는 일양분포를 제외한 나머지 분포들의 자료에 대해서는 다소 높은 경험적 검정력을 보였다.

5. 결 론

우리는 모의실험을 통하여 정규성 검정방법들의 경험적 검정력을 살펴보았다. 모의실험의 결과 대체로 Shapiro-Wilk 검정과 표본 Entropy 검정의 경험적 검정력이 가장 높았고, 그 다음으로 Anderson-Darling 검정, Kullback-Leibler 검정 순으로 경험적 검정력이 높은 것으로 나타났다. 또한 본 논문에서 고려한 분포들에 대해서 표본의 수가 커짐에 따라 당연한 결과이지만 경험적 검정력도 증가하는 것을 볼 수 있었다. 카이제곱 검정은 일양분포를 제외한 나머지 분포들의 자료에 대해서는 다소 높은 경험적 검정력을 보임을 알 수 있었다.

결론적으로 고려된 여러 가지 정규성 검정방법들은 저마다의 특징을 가지고 있지만 소표본 모의실험의 경험적 검정력 결과를 참고하면 정규성 검정의 한 방법으로 Shapiro-Wilk 검정과 표본 Entropy 검정, 그리고 Anderson-Darling 검정이 비교적 우위에 있음을 감안하여 적절히 활용하면 좋다고 사료된다.

참고문헌

1. Arizono, I. and Ohta, H. A. (1989). Test for Normality Based on Kullback-Leibler Information. *The American Statisticain,* 43, 20-22
2. Nelson, L. S. (1998). The Anderson-Darling Test for Normality. *Journal of Quality Technology*, 30, 298-299
3. Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can reasonably be supposed to have arisen from random sampling. *Philosophical Magazine* (5), 50, 157-175
4. Shapiro, S. S. and Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Sample). *Biometrika,* 52, 591-611
5. Smirnov, N. V. (1939). *Estimate of deviation between empirical distribution functions in two independent samples.* (Russian) Bulletin Moscow University, 2(2), 3-16
6. Stephens, M. A. (1974). EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association*, 69, 730-737
7. Stephens, M. A. (1982). *Anderson-Darling Test of Goodness of Fit.* Encyclopedia of Statistical Science, vol. 1, edited by S. Kots, N. L. Johnson, and C. B. Read. John Wiley & Sons, New York, NY, 81-85
8. Vasicek, O. (1976). A Test for Normality Based on Sample Entropy. *Journal of the Royal Statistical Society*, B, 38, 54-59

[2006년 1월 접수, 2006년 2월 채택]