

## 왜정규 표본평균의 분포함수에 대한 안장점근사<sup>†</sup>

나중화<sup>1</sup> · 유혜경<sup>2</sup>

<sup>1</sup>충북대학교 정보통계학과 · <sup>2</sup>국립보건연구원

접수 2013년 8월 5일, 수정 2013년 8월 24일, 게재확정 2013년 8월 30일

### 요약

최근 많은 통계 이론과 응용 문제에 정규분포의 대안으로 왜정규분포에 대한 활용이 높아지고 있다. 본 논문에서는 왜정규분포에 기반한 표본평균의 분포함수에 대한 안장점근사를 다루었다. 안장점근사는 기존의 정규근사에 비해 매우 뛰어난 정확성을 보일 뿐 아니라, 소표본에서도 정확한 근사 결과를 제공한다. 본 논문에서 제시한 왜정규분포에 관련된 안장점근사는 복잡한 계산이 요구되는 기존의 Gupta와 Chen (2001)과 Chen 등 (2004)에 대한 근사적 방법으로 사용될 수 있다. 모의실험을 통해 표본평균의 분포함수에 대한 제안된 안장점근사의 정확도를 확인하고, 실제 자료에 대한 응용으로 Roberts (1966)의 쌍둥이 자료의 분석에 적용하였다.

주요용어: 누울생성함수, 안장점근사, 왜정규분포, 표본평균.

### 1. 서론

왜정규분포 (skew-normal distribution)는 Azzalini (1985, 1986)에 의해 처음 소개된 분포로 밀도함수가 다음과 같이 주어진다.

$$f_{\alpha}(x) = 2\phi(x)\Phi(\alpha x), \quad -\infty < x < \infty. \quad (1.1)$$

위 식에서  $\phi(\cdot)$ 와  $\Phi(\cdot)$ 는 각각 표준정규분포의 밀도함수와 분포함수를 나타낸다. 여기서  $\alpha$ 는 왜도모수로  $-\infty < \alpha < \infty$ 이며, 위 분포를  $SN(\alpha)$ 로 나타낸다.

위 분포는 왜도모수의 값에 따라  $\alpha = 0$ 일 때 표준정규분포가 되며,  $\alpha \rightarrow \infty$ 일 때 반정규분포를 극한 분포로 가진다. 또한 이 분포는 정규분포와 유사한 성질로  $X \sim SN(\alpha)$ 일 때  $X^2 \sim \chi^2(1)$ 이 성립한다. 다양한 왜도와 첨도를 포괄하는 위 분포는 정규분포를 벗어나는 실제의 자료에 대한 적합 모형으로 유용성이 뛰어나며, 수학적으로 다루기에 편리하고, 정규분포와 같이 다변량 분포로의 일반화도 쉬운 장점을 가진다. Arnold 등 (1993)은 실제의 예제를 통해 왜정규분포의 실용성을 확인한 바 있으며, Henze (1986)와 Pouramahdi (2007)는 독립인 두 정규 확률변수와의 관계성을 통해, Azzalini와 Dalla Valle (1996)는 이변량 정규분포의 조건부 분포를 통해 위 분포를 해석하고 있다. 이처럼 다양한 실용적, 해석적 의미를 가지는 왜정규분포에 대한 응용이 활발하게 이루어지고 있다. 관련 연구로는 Kim (2012), Kang 등(2011)이 있다.

<sup>†</sup> 이 논문은 2012년도 충북대학교 학술연구지원사업의 연구비 지원에 의하여 연구되었음.

<sup>1</sup> 교신저자: (361-763) 충북 청주시 흥덕구 내수동로 52 충북대학교 정보통계학과, 교수.  
E-mail: cherin@cdu.ac.kr

<sup>2</sup> (363-951) 충북 청원군 오송읍 오송생명2로 187, 국립보건연구원, 선임연구원.

본 논문에서는 두 가지 주제를 다루기로 한다. 첫째, 적합도 검정 등에 유용하게 사용되는 왜정규분포의 분포함수 (또는 꼬리확률)에 대한 근사를 다룬다. 둘째, 왜정규 분포로부터의 확률표본에 기초한 표본평균의 분포함수에 대한 근사를 다루고자 한다. 본 논문에서 사용된 근사법은 안장점근사 (saddle-point approximation) 로, 이 근사법은 대표본은 물론 소표본에서도 매우 뛰어난 근사 결과를 제공하는 것으로 알려져 있다. 첫 번째 주제는 두 번째 주제의 특별한 경우 ( $n = 1$ 인 경우)로 취급될 수 있으며, 따라서 본 논문에서 다루게 될 안장점근사는 표본평균에 대한 내용을 중심으로 소개한다. 2절에서는 왜정규 분포함수에 대한 기존의 계산법으로 Owen (1956)과 Gupta와 Chen (2001)의 방법을 소개하고, 아울러 표본평균과 관련된 Chen 등 (2004)의 연구 결과를 간략히 소개한다. 3절에서는 표본평균의 분포함수에 대한 안장점근사법을 소개하고, 앞서 언급한 두 가지 주제에 대한 안장점근사의 결과를 제시하였다. 4절에서는 Roberts (1966)의 쌍둥이 자료를 이용한 실제자료 분석을 다루었고, 5절은 결론으로 구성하였다.

## 2. 왜정규 기반 분포함수

### 2.1. 왜정규 분포함수

식 (1.1)의 왜정규분포의 적률생성함수는,  $\delta = \alpha/\sqrt{1 + \alpha^2}$ 이라 할 때, 다음과 같이 주어진다.

$$M_X(t) = 2 \exp\left(\frac{t^2}{2}\right) \Phi(\delta t). \quad (2.1)$$

이로부터, 왜정규분포의 평균과 분산은 다음과 같이 주어짐을 쉽게 확인할 수 있다.

$$E(X) = \sqrt{\frac{2}{\pi}}\delta, \quad \text{Var}(X) = 1 - \frac{2}{\pi}\delta^2.$$

한편, 본 논문에서의 첫 번째 관심사인 왜정규분포의 분포함수는 다음과 같이 주어진다.

$$\begin{aligned} F_\alpha(x) &= 2 \int_{-\infty}^x \int_{-\infty}^{\alpha t} \phi(u)\phi(t)du dt \\ &= \Phi(x) - 2T(x; \alpha). \end{aligned} \quad (2.2)$$

위 식에서  $T(x; \alpha)$ 는 Owen (1956)의 함수로 다음과 같이 정의된다.

$$T(x; \alpha) = \int_x^\infty \int_0^{\alpha t} \phi(u)\phi(t)du dt.$$

Gupta와 Chen (2001)은 다양한  $\alpha$ 와  $x$ 값에 대해, Simpson의 방법을 이용한 수치적분을 통해 위 분포함수를 계산하고, 그 결과를 방대한 표로 제시한 바 있다.

### 2.2. 왜정규 표본평균의 분포함수

본 논문의 두 번째 관심사인 표본평균의 분포함수에 대한 기존 연구 결과를 소개하면 다음과 같다. Chen 등 (2004)은  $SN(\alpha)$  분포로부터의 확률표본에 기초한 표본평균 ( $\bar{X}$ )의 밀도함수를 직접 유도한 바 있다. 분포이론에 기초한 그들의 유도과정은 Helmert 변환 등의 복잡한 과정을 사용하고 있으며, 그 결과가 다변량 정규분포의 누적분포 항을 포함하고 있을 뿐 아니라, 분포함수의 값을 얻기 위해서는 이를 다시 수치적분 해야 하는 어려움이 따른다. 이 계산과정은 매우 복잡하여 분포함수 (또는 꼬리확률)의 계산 등이 요구되는 실제 응용문제에 사용되기에는 적합하지 않다.

### 3. 표본평균에 대한 안장점근사

#### 3.1. 분포함수에 대한 안장점근사

$X_1, \dots, X_n$ 을 누울생성함수 (CGF)  $K(\xi)$ 를 가지는 분포로부터의 확률표본이라 하자. 이때 표본평균  $T = \bar{X}$ 의 분포함수에 대한 안장점근사식은 다음과 같이 주어진다 (Lugannani와 Rice, 1980).

$$F_T^{LR}(t) \approx \Phi(w) + \phi(w) \left( \frac{1}{w} - \frac{1}{v} \right). \tag{3.1}$$

여기서  $w$ 와  $v$ 는 각각

$$w = \text{sgn}(\hat{\xi}) \left[ 2n \left\{ \hat{\xi}t - K(\hat{\xi}) \right\} \right]^{1/2}, \quad v = \hat{\xi} \left\{ nK''(\hat{\xi}) \right\}^{1/2}$$

으로 주어진다. 또한  $\hat{\xi}$ 은 안장점 (saddlepoint)으로 불리는 값으로, 다음의 안장점 방정식 (saddlepoint equation)의 해이다.

$$K'(\xi) = t.$$

여기서  $K'(\cdot)$ 은 누울생성함수의  $\xi$ 에 대한 1차 미분을 나타낸다. 즉, 위 식에서  $w$ 와  $v$ , 그리고  $\hat{\xi}$ 은 모두  $t$ 의 함수이며, 위 근사식의 계산을 위해서는 모집단 분포의 누울생성함수와 근사를 원하는  $t$  값에서의 안장점의 계산이 요구된다.

위 근사식 (3.1)의 변형으로 Barndorff-Nielsen (1986, 1990)은 다음의 근사식

$$F_T(t) \approx \Phi \left\{ w + \frac{1}{w} \log \left( \frac{v}{w} \right) \right\} \tag{3.2}$$

을 제안하였으며, Jensen (1992)은 이 근사식의 상대오차가  $O(n^{-3/2})$ 으로 주어지며, 위의 근사식 (3.1)과 근사의 정도가 동일함을 보인 바 있다. 식 (3.2)는 식 (3.1)의 구조상 그 결과가 음의 값이 나올 수도 있는 문제점을 보완한 것이라 말할 수 있으나, 이 경우를 제외한 계산 결과에는 큰 차이가 없으므로, 본 논문에서는 식 (3.1)의 근사식을 사용하기로 한다.

식 (3.1)의 안장점근사는 매우 뛰어난 근사 결과를 제공한다. 다만  $t = E(T)$ 인 점에서는 안장점의 계산이 불안정하므로 다음의 식을 사용할 수 있다 (Daniels, 1987).

$$F_T(E(T)) \approx \frac{1}{2} - \frac{\lambda_3(0)}{6\sqrt{2\pi n}}. \tag{3.3}$$

여기서  $\lambda_3(0) = K'''(0) / \left\{ K''(0) \right\}^{3/2}$ 이며,  $K''(\cdot)$ 과  $K'''(\cdot)$ 은 각각 누울생성함수의  $\xi$ 에 대한 2차와 3차 미분을 나타낸다. 안장점근사에 대한 소개와 응용에 대한 연구로는 Na (2008) 가 있다.

#### 3.2. 왜정규 표본평균의 경우

$X_1, \dots, X_n$ 을 왜정규분포  $SN(\alpha)$ 로부터의 확률표본이라 하자. 이때, 표본평균의 분포함수에 대한 안장점근사는 식 (3.1)을 통해 구해진다. 식 (2.1)로부터,  $K'(\xi)$ 와  $K''(\xi)$ 은 다음의 식으로 주어진다.

$$K'(\xi) = t + \frac{\delta\phi(\delta t)}{\Phi(\delta t)},$$

$$K''(\xi) = 1 - \frac{\delta^2\phi(\delta t) \{ \delta\Phi(\delta t) + \phi(\delta t) \}}{\Phi^2(\delta t)}.$$

위 결과로부터 식 (3.1)을 적용하기 위해서는 안장점방정식의 해를 수치적으로 구하는 과정이 요구된다.

### 4. 모의실험

#### 4.1. 왜정규 분포함수에 대한 안장점근사

아래의 Table 4.1 ~ Table 4.3은 왜정규확률변수의 분포함수에 대하여 안장점근사를 실시한 결과이다. 각 칸에서 위의 값은 식 (2.2)를 이용한 정확한 값이고, 아래의 값은 안장점근사의 결과로 식 (3.1)에서  $n = 1$ 인 경우의 결과이다. 표에서 제시된 다양한  $\alpha (= 1, 5, 20)$ 와  $x$ 값에 대해 두 값이 거의 일치함을 확인할 수 있다. 아래 Table에 대한 설명으로, Table 4.1에서  $x$ 값의 범위는 -2.0 에서 2.9 가 사용되었으며, 0.1 단위로 변화시켜 근사값을 구하였다. 한편 왜정규분포가 음의 왜도값을 가지는 경우( $\alpha < 0$ )의 분포함수는, 다음의 관계로부터 쉽게 구해질 수 있으므로, 모의실험에서 제외되었다.

$$\begin{aligned}
 F_{-\alpha}(x) &= \int_{-\infty}^x 2\phi(x)\Phi(-\alpha x)dx \\
 &= \int_{-x}^{\infty} 2\phi(x)\Phi(\alpha x)dx \\
 &= 1 - F_{\alpha}(-x).
 \end{aligned}$$

**Table 4.1** Saddlepoint approximations to the distribution function of skew-normal ( $\alpha = 1$ )

		$P(SN(1) < x)$									
$x$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
-2.0	0.0005	0.0003	0.0002	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	
	0.0005	0.0003	0.0002	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	
-1.0	0.0252	0.0184	0.0132	0.0094	0.0065	0.0045	0.0030	0.0020	0.0013	0.0008	
	0.0252	0.0184	0.0133	0.0094	0.0065	0.0045	0.0030	0.0020	0.0013	0.0008	
-0.0	0.2500	0.2118	0.1770	0.1460	0.1187	0.0952	0.0752	0.0585	0.0449	0.0339	
	0.2500	0.2118	0.1771	0.1461	0.1188	0.0953	0.0753	0.0586	0.0449	0.0339	
0.0	0.2500	0.2914	0.3355	0.3818	0.4296	0.4781	0.5267	0.5746	0.6212	0.6658	
	0.2500	0.2914	0.3355	0.3817	0.4294	0.4779	0.5264	0.5743	0.6208	0.6653	
1.0	0.7079	0.7471	0.7831	0.8158	0.8450	0.8708	0.8934	0.9129	0.9294	0.9434	
	0.7074	0.7466	0.7826	0.8153	0.8446	0.8705	0.8931	0.9126	0.9292	0.9432	
2.0	0.9550	0.9646	0.9724	0.9787	0.9837	0.9876	0.9907	0.9931	0.9949	0.9963	
	0.9549	0.9645	0.9723	0.9786	0.9836	0.9876	0.9907	0.9931	0.9949	0.9963	

**Table 4.2** Saddlepoint approximations to the distribution function of skew-normal ( $\alpha = 5$ )

		$P(SN(5) < x)$									
$x$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
-0.0	0.0628	0.0308	0.0127	0.0043	0.0012	0.0003	0.0000	0.0000	0.0000	0.0000	
	0.0645	0.0318	0.0131	0.0045	0.0012	0.0003	0.0000	0.0000	0.0000	0.0000	
0.0	0.0628	0.1104	0.1712	0.2402	0.3121	0.3832	0.4515	0.5161	0.5763	0.6319	
	0.0645	0.1117	0.1703	0.2358	0.3043	0.3732	0.4408	0.5058	0.5671	0.6242	
1.0	0.6827	0.7287	0.7699	0.8064	0.8385	0.8664	0.8904	0.9109	0.9281	0.9426	
	0.6766	0.7241	0.7666	0.8043	0.8372	0.8658	0.8903	0.9110	0.9285	0.9430	
2.0	0.9545	0.9643	0.9722	0.9786	0.9836	0.9876	0.9907	0.9931	0.9949	0.9963	
	0.9549	0.9647	0.9725	0.9788	0.9838	0.9878	0.9908	0.9932	0.9950	0.9963	
3.0	0.9973	0.9981	0.9986	0.9990	0.9993	0.9995	0.9997	0.9998	0.9999	0.9999	
	0.9973	0.9981	0.9986	0.9990	0.9993	0.9995	0.9997	0.9998	0.9999	0.9999	

**Table 4.3** Saddlepoint approximations to the distribution function of skew-normal ( $\alpha = 20$ )

		$P(SN(20) < x)$									
$x$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
0.0	0.0159	0.0800	0.1585	0.2358	0.3108	0.3829	0.4515	0.5161	0.5763	0.6319	
	0.0172	0.0801	0.1538	0.2273	0.2999	0.3711	0.4399	0.5056	0.5676	0.6247	
1.0	0.6827	0.7287	0.7699	0.8064	0.8385	0.8664	0.8904	0.9109	0.9281	0.9426	
	0.6772	0.7248	0.7673	0.8049	0.8378	0.8662	0.8906	0.9113	0.9287	0.9431	
2.0	0.9545	0.9643	0.9722	0.9786	0.9836	0.9876	0.9907	0.9931	0.9949	0.9963	
	0.9550	0.9648	0.9726	0.9789	0.9839	0.9878	0.9908	0.9932	0.9950	0.9963	
3.0	0.9973	0.9981	0.9986	0.9990	0.9993	0.9995	0.9997	0.9998	0.9999	0.9999	
	0.9973	0.9981	0.9986	0.9990	0.9993	0.9995	0.9997	0.9998	0.9999	0.9999	

4.2. 왜정규 표본평균의 분포함수에 대한 안장점근사

아래의 Figure 4.1 ~ Figure 4.3은  $SN(\alpha)$ 분포로부터의 확률표본에 기초한 표본평균의 분포함수에 대한 안장점근사를 실시한 결과이다. 모의실험은  $\alpha=1, 5, 20$ 에 대해  $n=5, 30$ 인 경우에 대해 수행하였으며, 그 결과를 정규근사 및 정확한 값과 비교하였다. 각 그림에서 Normal, Saddlepoint, Sim.Exact는 각각 다음을 의미한다.

Normal: 정규근사

Saddlepoint: 식 (2.1)에 의한 안장점근사

Sim.Exact: 10만번의 모의실험의 결과 (정확한 값)

모의실험 결과로부터  $\alpha$ 값이 커짐에 따라 정규근사의 결과는 표본이 큰 경우에는 정확도를 유지하나, 소표본의 경우에는 정확도가 점점 떨어짐을 확인할 수 있다. 이에 비해 안장점근사의 결과는 소표본의 경우에 대해서도 매우 정확한 근사 결과를 제공함을 알 수 있다.

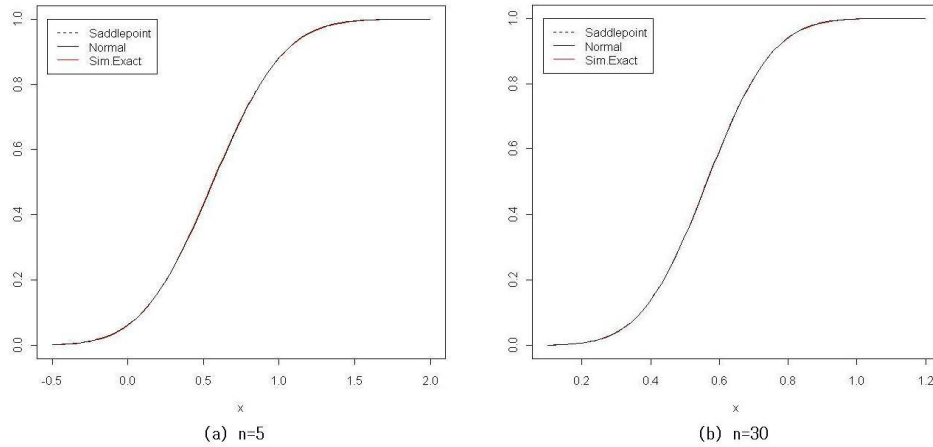


Figure 4.1 Saddlepoint approximation to the distribution function of sample mean of skew-normal ( $\alpha = 1$ )

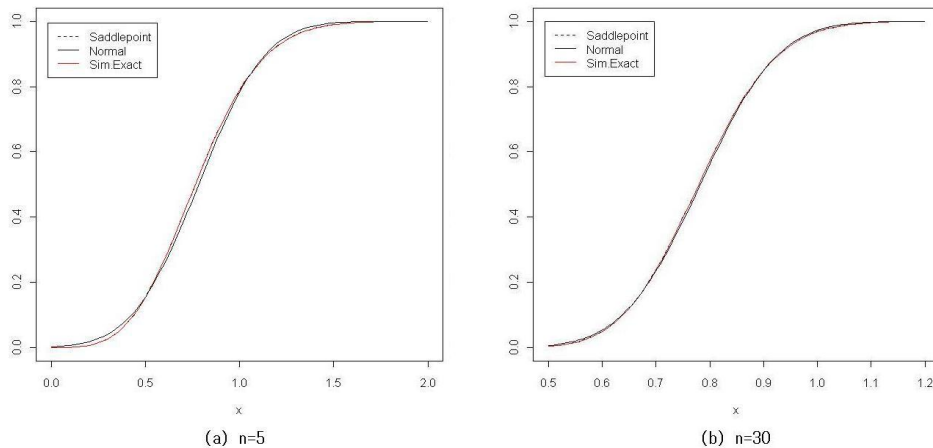


Figure 4.2 Saddlepoint approximation to the distribution function of sample mean of skew-normal ( $\alpha = 5$ )

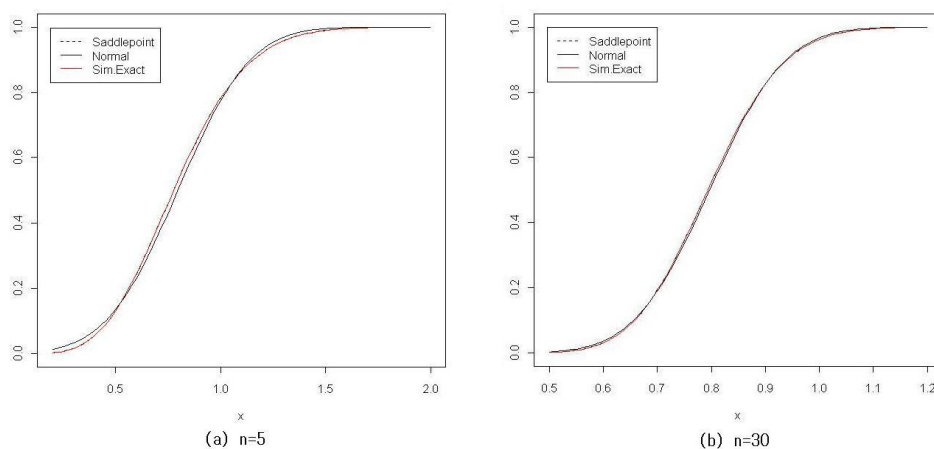


Figure 4.3 Saddlepoint approximation to the distribution function of sample mean of skew-normal ( $\alpha = 20$ )

## 5. 실제응용사례

Roberts (1966)는 쌍둥이 중 태어나서 처음으로 먼저 감기에 걸리는 나이에 대한 연구를 위해 다음의 상관모형을 고려하였다. 먼저  $X$ 와  $Y$ 를 쌍둥이 각각이 처음으로 감기에 걸린 나이라 하면,  $Z = \min(X, Y)$ 는 먼저 감기에 걸린 나이로 정의된다. Roberts의 관심사는  $n$ 개의 확률표본  $Z_1, \dots, Z_n$ 을 통해,  $X$ 와  $Y$  (쌍둥이 각각이 감기에 걸리는 나이) 간의 상관계수  $\rho$ 를 추정하고자 하였다. 이 경우  $(X, Y)$ 의 모형으로 평균과 분산이 동일한 값을 가지는 이변량정규분포  $N_2(\mu, \mu, \sigma^2, \sigma^2, \rho)$ 를 가정할 수 있다. 위 모형으로부터 Roberts는  $W = (Z - \mu)/\sigma$ 의 분포가

$$W = \frac{(Z - \mu)}{\sigma} \sim SN\left(-\sqrt{\frac{1-\rho}{1+\rho}}\right) \quad (5.1)$$

임을 보였다 (Roberts가  $W$ 의 밀도함수를 유도할 당시에는 왜정규분포가 소개되기 이전이었다). 또한 Roberts는 상관계수  $\rho$ 의 추정량으로  $\bar{\rho}$ 를 다음과 같이

$$\bar{\rho} = 1 + \frac{\pi}{n-1} - \frac{\pi \left(\sum_{i=1}^n \frac{Z_i - \mu}{\sigma}\right)^2}{n(n-1)} \quad (5.2)$$

으로 제안하고, 이 추정량이 불편성과 완비성을 만족함을 보인 바 있다. 또한  $\bar{\rho}$ 의 점근분포가

$$\bar{\rho} \sim N\left(\rho, \frac{4(1-\rho)(\pi-1+\rho)}{n}\right) \quad (5.3)$$

임을 보였다.

본 논문에서는 추정량  $\bar{\rho}$ 의 분포함수에 대한 안장점근사를 실시하고, 이를 정규근사 및 모의실험을 통한 정확한 값과의 비교를 수행한다. 먼저 안장점근사를 위해  $\bar{\rho}$ 를,  $W_i = (Z_i - \mu)/\sigma$ 를 통해, 다음과 같이 표현하자.

$$\bar{\rho} = 1 + \frac{\pi}{n-1} - \frac{\pi n \bar{W}^2}{(n-1)}. \quad (5.4)$$

식 (5.4)로부터  $\bar{\rho}$ 의 분포함수는 다음과 같이 계산된다.

$$P(\bar{\rho} \leq a) = P\left(\bar{W}^2 \geq \frac{n-1}{\pi n} \left(\frac{\pi}{n-1} + 1 - a\right)\right)$$

$$\begin{aligned}
 &= 1 - P(-b < \overline{W} < b) \\
 &= 1 - P(\overline{W} < b) + P(\overline{W} < -b).
 \end{aligned}
 \tag{5.5}$$

여기서  $b = b(a, n) = \sqrt{1/n + (n-1)(1-a)/\pi n}$ 이다. 위 식으로부터 주어진  $n$ 과  $a$  값에 대해 근사식 (3.1)을 통해 안장점근사를 실시할 수 있다.

아래의 Table 5.1은 고정된  $\mu = 0, \sigma^2 = 1, \rho (= -0.9, 0, 0.9)$ 에 대해  $n$ 과  $a$ 값의 변화에 따른  $\bar{\rho}$ 의 분포함수에 대한 안장점근사의 결과를 제시하였다. 안장점근사의 정확도를 확인하기 위해 모의실험을 통한 정확한 값과 함께 정규근사의 결과도 함께 제시하였다. 모의실험은 10만번의 이변량정규분포로부터의 난수에 기초하여 구하였으며, 정규근사는 식 (5.3)을 이용하여 구한 값이다. 정규근사의 경우 표본이 작은 경우에는 정확도가 매우 떨어지나, 안장점근사의 경우에는 소표본의 경우를 비롯한 모의실험이 이루어진 전 범위에서 그 결과가 매우 정확함을 알 수 있다. 쌍둥이 자료의 특성상  $\rho = -0.9$  (음의 상관)인 경우보다는  $\rho = 0.9$  (양의 상관)의 경우가 더 큰 실제적인 의미를 가진다고 말할 수 있다.

**Table 5.1** Saddlepoint approximation to the distribution function of Roberts' estimator

n	a	P( $\bar{\rho} < a$ )								
		$\rho = -0.9$			$\rho = 0$			$\rho = 0.9$		
		Normal	Saddle	Exact	Normal	Saddle	Exact	Normal	Saddle	Exact
10	-0.99	0.4631	0.4029	0.4051	0.1424	0.1648	0.1663	0.0000	0.0205	0.0207
	-0.9	0.5000	0.4335	0.4342	0.1654	0.1801	0.1813	0.0000	0.0232	0.0229
	-0.8	0.5410	0.4691	0.4667	0.1937	0.1986	0.1969	0.0000	0.0267	0.0268
	0	0.8229	0.7785	0.7791	0.5000	0.4102	0.4094	0.0049	0.0822	0.0818
	0.8	0.9599	0.9805	0.9808	0.8063	0.7383	0.7398	0.3872	0.2738	0.2739
	0.9	0.9681	0.9894	0.9896	0.8346	0.7838	0.7849	0.5000	0.3233	0.3260
50	0.99	0.9742	0.9946	0.9945	0.8576	0.8246	0.8239	0.6018	0.3780	0.3784
	-0.99	0.4179	0.3920	0.3927	0.0084	0.0220	0.0222	0.0000	0.0000	0.0000
	-0.9	0.5000	0.4697	0.4703	0.0148	0.0312	0.0301	0.0000	0.0000	0.0000
	-0.8	0.5910	0.5601	0.5596	0.0266	0.0452	0.0436	0.0000	0.0000	0.0000
	0	0.9809	0.9908	0.9908	0.5000	0.4579	0.4573	0.0000	0.0021	0.0021
	0.8	1.0000	1.0000	1.0000	0.9734	0.9915	0.9912	0.2607	0.2179	0.2191
100	0.9	1.0000	1.0000	1.0000	0.9852	0.9982	0.9981	0.5000	0.3673	0.3651
	0.99	1.0000	1.0000	1.0000	0.9916	0.9998	0.9999	0.7180	0.5839	0.5842
	-0.99	0.3848	0.3676	0.3701	0.0004	0.0024	0.0027	0.0000	0.0000	0.0000
	-0.9	0.5000	0.4785	0.4767	0.0011	0.0046	0.0048	0.0000	0.0000	0.0000
	-0.8	0.6276	0.6074	0.6059	0.0031	0.0092	0.0100	0.0000	0.0000	0.0000
	0	0.9983	0.9996	0.9997	0.5000	0.4701	0.4701	0.0000	0.0000	0.0000
100	0.8	1.0000	1.0000	1.0000	0.9969	0.9998	0.9999	0.1823	0.1754	0.1748
	0.9	1.0000	1.0000	1.0000	0.9989	1.0000	1.0000	0.5000	0.3984	0.3979
	0.99	1.0000	1.0000	1.0000	0.9996	1.0000	1.0000	0.7927	0.7429	0.7422

## 6. 결론

최근 많은 응용분야에서 왜정규분포에 대한 활용이 높아지고 있다. 본 논문에서 다룬 왜정규 표본평균에 대한 분포함수에 대한 안장점근사는 기존의 Owen (1956) 함수 또는 Gupta와 Chen (2001)의 계산을 단순화할 뿐 아니라, 표본평균에 대한 Chen 등 (2004)의 결과를 이용한 분포함수 계산을 보다 편리하게 수행할 수 있게 한다. 제안된 근사법은 정확한 값의 계산에 요구되는 복잡한 계산 과정을 대신할 수 있는 매우 정확한 근사적 방법을 제공하므로, 적합도 검정을 비롯한 많은 통계적 응용문제에 직접 활용될 수 있다.

## References

- Arnold, B. C., Beaver, R. J., Groeneveld, R. A. and Meeker, W. Q. (1993). The nontruncated marginal of a truncated bivariate normal distribution. *Psychometrika*, **58**, 471-478.
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171-178.
- Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica*, **46**, 199-208.
- Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, **83**, 715-726.
- Barndorff-Nielsen, O. E. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika*, **73**, 307-322.
- Barndorff-Nielsen, O. E. (1990). Approximate interval probabilities. *Journal of the Royal Statistical Society B*, **52**, 485-496.
- Chen, J. T., Gupta, A. K. and Nguyen, T. T. (2004). The density of the skew-normal sample mean and its application. *Journal of Statistical Computation and Simulation*, **74**, 487-494.
- Daniels, H. E. (1987). Tail probability approximations. *International Statistical Review*, **55**, 37-48.
- Gupta, A. K. and Chen, T. (2001). Goodness-of-fit tests for the skew-normal distribution. *Communications in Statistics Simulation & Computation*, **30**, 907-930.
- Henze, N. (1986). A probabilistic representation of the skew-normal distribution. *Scandinavian Journal of Statistics*, **13**, 271-275.
- Jensen, J. L. (1992). The modified signed rank likelihood statistic and saddlepoint approximations. *Biometrika*, **79**, 693-703.
- Kang, S. G., Kim, D. H. and Lee, W. D. (2011). Noninformative priors for the reliability function of two-parameter distribution. *Journal of Korean Data & Information Science Society*, **22**, 361-369.
- Kim, J. D. (2012). Moment of the ratio and approximate MLEs of parameters in a bivariate Pareto distribution. *Journal of Korean Data & Information Science Society*, **23**, 1213-1222.
- Lugannani, R. and Rice, S. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables. *Advanced Applied Probability*, **12**, 475-490.
- Na, J. H. (2008). Saddlepoint approximation to quadratic form and application to intraclass correlation coefficient. *Journal of Korean Data & Information Science Society*, **19**, 497-504.
- Owen, D. B. (1956). Tables for computing bivariate normal probabilities. *The Annals Mathematical Statistics*, **27**, 1075-1090.
- Pourahmadi, M. (2007). Construction of skew-normal random variables: Are they linear combination of normal and half normal? *Journal of Statistical Theory and Applications*, **3**, 314-328.
- Roberts, C. (1966). A correlation model useful in the study of twins. *Journal of the American Statistical Association*, **61**, 1184-1190.



## Saddlepoint approximation for distribution function of sample mean of skew-normal distribution<sup>†</sup>

Jong-Hwa Na<sup>1</sup> · Hye-Kyung Yu<sup>2</sup>

<sup>1</sup>Department of Information and Statistics, Chungbuk National University

<sup>2</sup>Korea National Institute of Health

Received 5 August 2013, revised 24 August 2013, accepted 30 August 2013

### Abstract

Recently, the usage of skew-normal distribution, instead of classical normal distribution, is rising up in many statistical theories and applications. In this paper, we deal with saddlepoint approximation for the distribution function of sample mean of skew-normal distribution. Comparing to normal approximation, saddlepoint approximation provides very accurate results in small sample sizes as well as for large or moderate sample sizes. Saddlepoint approximations related to the skew-normal distribution, suggested in this paper, can be used as an approximate approach to the classical method of Gupta and Chen (2001) and Chen et al. (2004) which need very complicated calculations. Through simulation study, we verified the accuracy of the suggested approximation and applied the approximation to Robert's (1966) twin data.

*Keywords:* Cumulant generating function, saddlepoint approximation, sample mean, skew-normal distribution.

---

<sup>†</sup> This work was supported by the research grant of Chungbuk National University in 2012.

<sup>1</sup> Corresponding author: Professor, Department of Information and Statistics, Chungbuk National University, Chungbuk 361-763, Korea. E-mail: cherin@cbu.ac.kr

<sup>2</sup> Researcher, Korea National Institute of Health, Chungbuk 363-951, Korea.