

Saddlepoint Approximation to the Linear Combination Based on Multivariate Skew-normal Distribution

Jonghwa Na^{a,1}

^aDepartment of Information & Statistics, Chungbuk National University

(Received August 19, 2014; Revised September 19, 2014; Accepted October 10, 2014)

Abstract

Multivariate skew-normal distribution(distribution that includes multivariate normal distribution) has been recently applied to many application areas. We consider saddlepoint approximation for a statistic of linear combination based on a multivariate skew-normal distribution. This approach can be regarded as an extension of Na and Yu (2013) that dealt saddlepoint approximation for the distribution of a skew-normal sample mean for a linear statistic and multivariate version. Simulations results and examples with real data verify the accuracy and applicability of suggested approximations.

Keywords: Multivariate skew-normal distribution, saddlepoint approximation, cumulant generating function.

1. 서론

Azzalini와 Dalla Valle (1996)는 p -차원의 다변량 왜정규분포의 밀도함수를 다음과 같이 정의하였다.

$$f_p(x) = 2\phi_p(x; \Omega)\Phi(\alpha'x), \quad x \in R^p, \quad (1.1)$$

여기서 $\phi_p(x; \Omega)$ 는 평균이 0이고 상관행렬이 Ω 인 p -차원의 다변량 정규분포의 밀도함수이며, $\Phi(\cdot)$ 는 일변량 표준정규분포 $N(0, 1)$ 의 분포함수이다. 또한 $\alpha(\in R^p)$ 는 형태모수로 불리며, $\alpha = 0$ 일 때 이 분포는 다변량 정규분포 $N_p(0, \Omega)$ 가 된다. Azzalini와 Capitanio (1999)는 위 분포를 $X \sim SN_p(\Omega, \alpha)$ 으로 나타내었다.

다변량 왜정규분포는 정규성을 벗어나는 다변량 자료의 적합에 유용하다. 예를 들어, 2차원의 경우 확률 등고선의 형태가 대략적으로 종모양을 취하지만 타원형에서 벗어나는 형태를 가지거나, 등고선의 간격이 정규분포의 형태와는 차이를 보이는 경우를 들 수 있다. 이러한 비정규성을 가지는 다변량 자료에 대해 정규분포 외의 뚜렷한 대안적인 분포 모형이 거의 없는 실정이다. 최근 다변량 왜정규분포에 대한

This work was supported by the research grant of the Chungbuk National University in 2013. This research was partially supported by the MSIP(The Ministry of Science, ICT and Future Planning), Korea, under the “SW master’s course of a hiring contract” support program(NIPA-2014-HB301-14-1011) supervised by the NIPA(National IT Industry Promotion Agency).

¹Department of Information and Statistics/ Business Data Convergence, Chungbuk National University, Naesudong-Ro 52, Seowon-Gu, Cheongju, Chungbuk, 361-763, Korea. E-mail: cherin@cbnu.ac.kr

응용이 매우 활발하다. 그 이유로 이 분포는 Azzalini (1985, 1986)가 처음 제안한 왜정규분포가 다변량으로 확장된 것으로, 일변량 분포와의 연계가 자연스럽고, 수학적으로 다루기 쉬우며, 다변량 정규분포를 특별한 경우로 포함할 뿐 아니라, 정규분포가 가지는 여러 가지 주요 성질을 보존하는 특징을 가지기 때문이다. 예를 들어, 이 분포는 적절한 조건하에서 이차형식의 분포가 카이제곱분포를 따르며, 선형 변환 및 주변분포에 대해 닫힌 성질을 가진다. 다변량 왜정규분포 및 그의 응용에 대해서는 Azzalini와 Capitanio (1999)와 Azzalini (2005)를 참고하기 바란다.

본 논문에서는 $X = (X_1, \dots, X_p)'$ 가 p -차원 다변량 왜정규분포 $SN_p(\Omega, \alpha)$ 를 따를 때, 다음의 선형 결합 통계량

$$h'X = \sum_{i=1}^p h_i X_i \quad (1.2)$$

의 분포함수에 대한 근사를 다루고자 한다. 여기서 $h' = (h_1, \dots, h_p)$ 는 상수벡터이다. 근사의 방법으로는 소표본에서도 정확도가 뛰어난 근사법으로 알려진 안장점근사를 사용하기로 한다. 2절에서는 관련된 선행연구로 Gupta와 Brown (2001)의 결과를 소개한다. 3절에서는 분포함수에 대한 안장점근사의 소개와 함께 이를 통한 근사 방법을 제시하였다. 4절에서는 모의실험 결과와 실제자료분석을 다루었으며, 5절은 결론으로 구성하였다.

2. 관련 선행연구

Azzalini와 Dalla Valle (1996)는 단변량 왜정규분포의 확장으로 다변량 왜정규분포를 다음과 같이 정의하였다. 먼저 $Y_0 \sim N(0, 1)$ 이고, 이와는 독립인 p -차원 확률벡터를 $Y = (Y_1, \dots, Y_p)'$ 라고 할 때, $(Y_0, Y)'$ 를 다음과 같이 정의하자.

$$\begin{pmatrix} Y_0 \\ Y \end{pmatrix} \sim N_{p+1} \left\{ 0, \begin{pmatrix} 1 & 0 \\ 0 & \Psi \end{pmatrix} \right\}.$$

위 식에서 Ψ 는 $p \times p$ 상관행렬이다. 만약 $\delta_i \in (-1, 1) (i = 1, \dots, p)$ 에 대해

$$X_j = \delta_1 |Y_0| + \sqrt{1 - \delta_1^2} Y_j, \quad j = 1, \dots, p \quad (2.1)$$

으로 정의하면, X_j 는 각각 단변량 왜정규분포 $SN(\lambda_i) (i = 1, 2, \dots, p)$ 를 따르며 ($\lambda_i = \delta_i / \sqrt{1 - \delta_i^2}$), 확률벡터 $X = (X_1, \dots, X_p)'$ 는 식 (1.1)의 다변량 왜정규분포의 밀도함수를 가진다. 이때 α 와 Ω 는 다음과 같이 정의된다.

$$\alpha' = \frac{\lambda' \Psi^{-1} \Delta^{-1}}{\sqrt{1 + \lambda' \Psi^{-1} \lambda}}, \quad (2.2)$$

위 식에서 Δ, λ, Ω 는 다음과 같이 정의된다.

$$\begin{aligned} \Delta &= \text{diag} \left((1 - \delta_1^2)^{\frac{1}{2}}, \dots, (1 - \delta_p^2)^{\frac{1}{2}} \right), \\ \lambda &= (\lambda_1, \dots, \lambda_p)', \quad \lambda_i = \frac{\delta_i}{(1 - \delta_i^2)^{\frac{1}{2}}}, \quad i = 1, \dots, p, \\ \Omega &= \Delta (\Psi + \lambda \lambda') \Delta. \end{aligned}$$

위 식에서 $p = 2$ 인 경우에 대해 Gupta와 Brown (2001)은 Y_0, Y_1, Y_2 가 서로 독립이며 $N(0, 1)$ 을 따르는 경우를 고려하였다. 즉, X_1 과 X_2 를 식 (2.1)과 같이 정의할 때, X_1 과 X_2 의 선형결합에 대한 분포함수가 다음의 식으로 주어짐을 보인 바 있다.

$$P(h_1X_1 + h_2X_2 < u) = 2 \int_0^\infty \Phi\left(\frac{u - (h_1\delta_1 + h_2\delta_2)t}{\sigma}\right) \phi(t) dt.$$

위 식에서 $\delta_i = \lambda_i / \sqrt{1 + \lambda_i^2}$ ($i = 1, 2$)이고, $\sigma^2 = h_1^2(1 - \delta_1^2) + h_2^2(1 - \delta_2^2)$ 이다. 특히, Gupta와 Brown은 $h_1 = 1, h_2 = -1, u = 0$ 인 경우 식 (2.1)의 결과가 다음과 같이 단순화 됨을 보인 바 있다.

$$\begin{aligned} P(X_1 < X_2) &= 2 \int_0^\infty \Phi\left(\frac{(\delta_2 - \delta_1)t}{\sqrt{2 - \delta_1^2 - \delta_2^2}}\right) \phi(t) dt \\ &= \frac{1}{\pi} \arctan\left(\frac{\delta_2 - \delta_1}{\sqrt{2 - \delta_1^2 - \delta_2^2}}\right) + \frac{1}{2}. \end{aligned} \quad (2.3)$$

위의 결과는 왜정규분포를 주변분포로 가지는 서로 연관된 두 확률변수 X_1 과 X_2 의 선형결합통계량의 분포함수를 다룬 것이다. 다음 절에서는 보다 일반적인 다변량 왜정규분포의 가정하에서의 선형결합에 대한 분포함수 문제를 다루고, 4절에서는 실제자료분석을 통해 Gupta와 Brown (2001)의 결과와 비교하기로 한다.

3. 다변량 왜정규 선형결합에 대한 안장점근사

3.1. 분포함수에 대한 안장점근사

안장점근사는 Daniels (1954)가 처음으로 통계학 분야에 소개한 근사 방법으로, 주어진 한 점에서 통계량의 밀도함수 또는 분포함수에 대한 근사를 제공한다. 통계량의 누울생성함수(cumulant generating function)에 기반한 안장점근사는 표본평균에 대한 근사를 시발점으로 그 동안 여러 유형의 통계량에 대해 연구되어 왔다. 이 가운데 Lugannani와 Rice (1980)는 표본평균의 분포함수에 대한 안장점근사를 제안하였으며, 본 연구에서는 이 근사의 개선된 형태인 Daniels (1987)의 결과를 다음과 같이 소개한다. 통계량 V_n 이 누울생성함수 $K_n(\lambda)$ 를 가진다고 하자. 이때 V_n 의 분포함수에 대한 안장점근사는 다음식으로 주어진다.

$$\Pr(V_n \leq v) \approx \Phi(\omega) + \phi(\omega) \left\{ \frac{1}{\omega} - \frac{1}{\xi} \right\}, \quad v \neq E(V_n). \quad (3.1)$$

여기서 $\Phi(\cdot)$ 와 $\phi(\cdot)$ 는 표준정규분포의 분포함수와 밀도함수를 나타내고, ω 와 ξ 는

$$\omega = \operatorname{sgn}(\hat{\lambda}) \left[2 \left\{ \hat{\lambda}x - K_n(\hat{\lambda}) \right\} \right]^{\frac{1}{2}}, \quad \xi = \hat{\lambda} \left\{ K_n''(\hat{\lambda}) \right\}^{\frac{1}{2}}$$

이고, $\hat{\lambda}$ 는 안장점 방정식 $K'_n(\lambda) = v$ 의 해이다.

3.2. 단변량 왜정규 선형결합의 경우

확률변수 X_i ($i = 1, \dots, p$)가 서로 독립이며 각각 $SN(\alpha_i)$ ($i = 1, \dots, p$)을 따를 때, $Y = h'X$ 의 적률생성함수는 다음과 같다.

$$M_Y(t) = E\left(e^{t \sum_{i=1}^p h_i X_i}\right) = \prod_{i=1}^p 2 \exp\left\{\frac{1}{2}(h_i t)^2\right\} \Phi(\delta_i h_i t).$$

여기서 $\delta_i = \alpha_i/\sqrt{1+\alpha_i}$ 이다. 위 결과로부터 안장점근사에 요구되는 누울생성함수의 1차와 2차 미분식의 결과는 다음과 같이 주어진다.

$$\begin{aligned} K_Y(t) &= \log M_Y(t) = n \cdot \log 2 + \sum_{i=1}^n \frac{1}{2} h_i^2 t^2 + \sum_{i=1}^n \log \Phi(\delta_i h_i t), \\ K'_Y(t) &= \sum_{i=1}^n h_i^2 t + \sum_{i=1}^n \frac{\delta_i h_i \phi(\delta_i h_i t)}{\Phi(\delta_i h_i t)}, \\ K''_Y(t) &= \sum_{i=1}^n h_i^2 + \sum_{i=1}^n \frac{\delta_i^2 h_i^2 \phi(\delta_i h_i t) \{ \delta_i h_i t \Phi(\delta_i h_i t) + \phi(\delta_i h_i t) \}}{\Phi^2(\delta_i h_i t)}. \end{aligned}$$

위 결과를 식 (3.1)에 대입하여 $Y = h'X$ 의 분포함수에 대한 안장점근사를 실시할 수 있다. 이 절차는 표본평균에 대한 Na와 Yu (2013)의 경우와 매우 유사하므로 본 논문의 4절에서의 모의실험에서 별도로 다루지는 않으나, 실제자료분석에서는 비교를 위해 그 결과를 함께 제시하기로 한다.

3.3. 다변량 왜정규 선형결합의 경우

X 가 $SN_p(\Omega, \alpha)$ 를 따를 때, X 의 적률생성함수는

$$M_X(t) = 2 \exp \left\{ \frac{1}{2} t' \Omega t \right\} \Phi(\delta' t), \quad \delta = \frac{\Omega \alpha}{\sqrt{1 + \alpha' \Omega \alpha}}$$

이다 (Azzalini와 Capitanio, 1999). 이로부터 선형결합 통계량 $Y = h'X$ 의 적률생성함수는

$$M_{h'X}(t) = 2 \exp \left\{ \frac{1}{2} t^2 h' \Omega h \right\} \Phi \left(\frac{\alpha' \Omega h t}{\sqrt{1 + \alpha' \Omega \alpha}} \right)$$

이다. 위 식으로부터 $Y = h'X$ 의 누울생성함수와 그의 1차, 2차 미분식은 다음과 같이 주어진다.

$$\begin{aligned} K_{h'Z}(t) &= \log M_{h'Z}(t) = \log 2 + \frac{1}{2} t^2 h' \Omega h + \log \Phi(\delta_h t), \\ K'_{h'Z}(t) &= t h' \Omega h + \frac{\phi(\delta_h t)}{\Phi(\delta_h t)}, \\ K''_{h'Z}(t) &= h' \Omega h - \frac{\delta_h^2 \phi(\delta_h t) [\delta_h t \Phi(\delta_h t) + \phi(\delta_h t)]}{\Phi^2(\delta_h t)}, \end{aligned}$$

여기서 $\delta_h = \alpha' \Omega h / \sqrt{1 + \alpha' \Omega \alpha}$ 이다. 위 식의 유도과정에는 다음의 사실이 사용되었다.

$$\frac{d}{dt} \Phi(\delta t) = \delta \cdot \phi(\delta t), \quad \frac{d}{dt} \phi(t) = -\delta^2 t \cdot \phi(\delta t).$$

위 결과를 식 (3.1)에 대입하여 $Y = h'X$ 의 분포함수에 대한 안장점근사를 실시할 수 있다.

4. 모의실험과 실제자료분석

4.1. 모의실험

이 절에서는 다변량 왜정규분포에서의 선형결합통계량 $Y = h'X$ 의 분포함수에 대한 안장점근사에 대한 모의실험을 수행한다. 모의실험 설계는 다음과 같다. 먼저 $X \sim SN_p(\Omega; \alpha)$ 이고, $p = 2, 3$ 인 경우에 대해 각각 다음의 상관행렬

Table 4.1. Design of simulation

$p = 2$			$p = 3$		
$\alpha' = (\alpha_1, \alpha_2)$	Ω	$h' = (h_1, h_2)$	$\alpha' = (\alpha_1, \alpha_2, \alpha_3)$	Ω	$h' = (h_1, h_2, h_3)$
(1, 3)	Ω_1	(2, 5)	(1, 3, 5)	Ω_2	(1, 2, 5)
(-1, 3)		(2, 5)	(-1, 3, -5)		(1, 2, 5)
(1, 3)		(1, -1)	(1, 3, 5)		(1, -2, 5)
(-1, 3)		(1, -1)	(-1, 3, -5)		(1, -2, 5)

Table 4.2. Saddlepoint approximation to the distribution function of linear combination

$P(h'X < x)$							
$h' = (2, 5), \Omega = \Omega_1, \alpha' = (1, 3)$				$h' = (2, 5), \Omega = \Omega_1, \alpha' = (-1, 3)$			
x	Normal	Saddle	S.Exact	x	Normal	Saddle	S.Exact
-3.0	0.0239	0.0035	0.0034	-7.0	0.0138	0.0104	0.0103
-1.5	0.0566	0.0239	0.0236	-5.5	0.0296	0.0248	0.0249
0.0	0.1170	0.0926	0.0919	-4.0	0.0579	0.0526	0.0525
1.5	0.2130	0.2223	0.2248	-2.5	0.1041	0.1002	0.1003
3.0	0.3439	0.3847	0.3921	-1.0	0.1726	0.1722	0.1721
4.5	0.4969	0.5410	0.5537	0.5	0.2646	0.2692	0.2687
6.0	0.6503	0.6837	0.6903	2.0	0.3764	0.3858	0.3857
7.5	0.7824	0.7922	0.7956	3.5	0.4998	0.5187	0.5120
9.0	0.8799	0.8709	0.8718	5.0	0.6232	0.6336	0.6340
10.5	0.9416	0.9241	0.9242	6.5	0.7351	0.7415	0.7421
12.0	0.9752	0.9577	0.9576	8.0	0.8271	0.8286	0.8295
13.5	0.9908	0.9777	0.9777	9.5	0.8957	0.8931	0.8935
15.0	0.9970	0.9889	0.9889	11.0	0.9420	0.9373	0.9374
16.5	0.9992	0.9948	0.9947	12.5	0.9704	0.9654	0.9653
18.0	0.9998	0.9977	0.9976	14.0	0.9861	0.9820	0.9820
				15.5	0.9941	0.9912	0.9911
				17.0	0.9977	0.9959	0.9959

$$\Omega_1 = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}, \quad \Omega_2 = \begin{pmatrix} 1 & 0.3 & 0.5 \\ 0.3 & 1 & 0.7 \\ 0.5 & 0.7 & 1 \end{pmatrix}$$

을 가정하고, 아래의 Table 4.1의 조건에서 모의실험을 수행하였다.

모의실험 결과는 아래의 Table 4.2 ~ Table 4.5에 제시하였다. 각 표에서 Normal, Saddle, S.Exact는 다음을 의미한다.

Normal: 정규근사

Saddle: Lugannani-Rice 방법에 의한 분포함수 안장점근사

S.Exact: 반복횟수 10만번의 모의실험을 통한 정확한 값

4.2. 실제자료분석과 응용

4.2.1. 인장-강도 모형에의 응용 X_1 과 X_2 가 각각 $SN(\alpha_1)$ 과 $SN(\alpha_2)$ 를 따를 때, 인장-강도 모형(stress-strength model)에서의 관심확률 $P(X_1 < X_2)$ 에 대한 근사를 생각하자. X_1 과 X_2 가 종속인

Table 4.3. Saddlepoint approximation to the distribution function of linear combination

$P(h'X < x)$							
$h' = (1, -1), \Omega = \Omega_1, \alpha' = (1, 3)$				$h' = (1, -1), \Omega = \Omega_1, \alpha' = (-1, 3)$			
x	Normal	Saddle	S.Exact	x	Normal	Saddle	S.Exact
-3.0	0.0093	0.0094	0.0093	-4.0	0.0002	0.0007	0.0007
-2.5	0.0278	0.0279	0.0278	-3.5	0.0014	0.0031	0.0032
-2.0	0.0698	0.0698	0.0697	-3.0	0.0073	0.0112	0.0112
-1.5	0.1493	0.1490	0.1483	-2.5	0.0285	0.0347	0.0347
-1.0	0.2737	0.2730	0.2725	-2.0	0.0863	0.0909	0.0904
-0.5	0.4348	0.4341	0.4335	-1.5	0.2049	0.2011	0.2005
0.0	0.6077	0.6073	0.6067	-1.0	0.3881	0.3750	0.3748
0.5	0.7615	0.7615	0.7607	-0.5	0.6008	0.5884	0.5887
1.0	0.8746	0.8750	0.8746	0.0	0.7868	0.7848	0.7856
1.5	0.9437	0.9441	0.9440	0.5	0.9091	0.9153	0.9157
2.0	0.9785	0.9788	0.9787	1.0	0.9696	0.9759	0.9759
2.5	0.9931	0.9933	0.9932	1.5	0.9921	0.9952	0.9952
3.0	0.9981	0.9982	0.9981	2.0	0.9984	0.9994	0.9993
3.5	0.9996	0.9996	0.9996				
4.0	0.9999	0.9999	0.9999				

Table 4.4. Saddlepoint approximation to the distribution function of linear combination

$P(h'X < x)$							
$h' = (1, 2, 5), \Omega = \Omega_2, \alpha' = (1, 3, 5)$				$h' = (1, 2, 5), \Omega = \Omega_2, \alpha' = (-1, 3, -5)$			
x	Normal	Saddle	S.Exact	x	Normal	Saddle	S.Exact
-2.0	0.0401	0.0011	0.0011	-15.0	0.0297	0.0340	0.0342
-0.5	0.0802	0.0244	0.0232	-13.5	0.0516	0.0560	0.0560
1.0	0.1450	0.1228	0.1221	-12.0	0.0848	0.0884	0.0885
2.5	0.2381	0.2676	0.2759	-10.5	0.1319	0.1336	0.1334
4.0	0.3569	0.4160	0.4268	-9.0	0.1945	0.1935	0.1929
5.5	0.4915	0.5533	0.5612	-7.5	0.2724	0.2684	0.2681
7.0	0.6271	0.6707	0.6763	-6.0	0.3633	0.3568	0.3566
8.5	0.7485	0.7668	0.7693	-4.5	0.4626	0.4549	0.4550
10.0	0.8450	0.8409	0.8416	-3.0	0.5644	0.5571	0.5573
11.5	0.9132	0.8955	0.8952	-1.5	0.6621	0.6567	0.6571
13.0	0.9560	0.9339	0.9333	0.0	0.7499	0.7474	0.7472
14.5	0.9799	0.9598	0.9594	1.5	0.8238	0.8243	0.8240
16.0	0.9917	0.9764	0.9763	3.0	0.8822	0.8849	0.8847
17.5	0.9970	0.9867	0.9866	4.5	0.9253	0.9292	0.9293
19.0	0.9990	0.9928	0.9928	6.0	0.9552	0.9593	0.9595
				7.5	0.9746	0.9781	0.9783
				9.0	0.9864	0.9891	0.9891

경우와 서로 독립인 경우에 대해 안장점근사의 결과를 Table 4.6에 제시하였다. Table 4.6에서 GB는 Gupta와 Brown (2001)에 의한 식 (2.3)의 결과이며, Saddle은 안장점근사, S.Exact는 10만번의 모의 실험을 통한 정확한 값이다. 이 가운데 안장점근사(Saddle)의 경우 종속인 경우는 Gupta와 Brown의 결과와의 비교를 위해 식 (2.2)에서 $\Psi = I_2$ 를 적용하였으며, 이로부터 계산된 다변량 왜정규 분포의 모수값을 식 (3.3)에 대입하여 안장점근사를 적용하였다. 독립인 경우는 3.2절의 방법을 적용한 결과가

Table 4.5. Saddlepoint approximation to the distribution function of linear combination

$P(h'X < x)$							
$h' = (1, -2, 5), \Omega = \Omega_2, \alpha' = (1, 3, 5)$				$h' = (1, -2, 5), \Omega = \Omega_2, \alpha' = (-1, 3, -5)$			
x	Normal	Saddle	S.Exact	x	Normal	Saddle	S.Exact
-4.0	0.0213	0.0147	0.0146	-11.0	0.0043	0.0133	0.0136
-3.0	0.0416	0.0332	0.0330	-10.0	0.0112	0.0244	0.0250
-2.0	0.0753	0.0670	0.0668	-9.0	0.0263	0.0429	0.0434
-1.0	0.1265	0.1215	0.1211	-8.0	0.0554	0.0722	0.0723
0.0	0.1980	0.1995	0.1995	-7.0	0.1055	0.1164	0.1155
1.0	0.2897	0.2990	0.2992	-6.0	0.1822	0.1798	0.1774
2.0	0.3975	0.4132	0.4133	-5.0	0.2866	0.2656	0.2618
3.0	0.5139	0.5319	0.5322	-4.0	0.4132	0.3752	0.3693
4.0	0.6290	0.6446	0.6452	-3.0	0.5495	0.5057	0.4992
5.0	0.7336	0.7432	0.7440	-2.0	0.6802	0.6482	0.6446
6.0	0.8208	0.8233	0.8242	-1.0	0.7916	0.7859	0.7858
7.0	0.8874	0.8841	0.8843	0.0	0.8761	0.8961	0.8970
8.0	0.9341	0.9275	0.9276	1.0	0.9331	0.9626	0.9633
9.0	0.9642	0.9567	0.9567	2.0	0.9674	0.9905	0.9907
10.0	0.9820	0.9753	0.9753	3.0	0.9856	0.9984	0.9983
11.0	0.9916	0.9866	0.9864				
12.0	0.9964	0.9930	0.9929				

Table 4.6. Saddlepoint approximation to the probability $P(X_1 < X_2)$ when $X_1 \sim SN(\alpha_1)$ and $X_2 \sim SN(\alpha_2)$

α_1	α_2	$P(X_1 < X_2)$				
		$X_1, X_2 : \text{Dependent case}$			$X_1, X_2 : \text{Independent case}$	
		GB	Saddle	S.Exact	Saddle	S.Exact
0	2	0.7180	0.7179	0.7178	0.7179	0.7179
0.5	2	0.6339	0.6339	0.6338	0.6156	0.6155
1	2	0.5701	0.5701	0.5699	0.5513	0.5512
2	2	0.5000	0.5000	0.5000	0.5000	0.5000
5	2	0.4444	0.4444	0.4443	0.4739	0.4740
10	2	0.4312	0.4312	0.4313	0.4693	0.4694
20	2	0.4275	0.4275	0.4272	0.4681	0.4683

다. 그 결과 대부분의 경우 안장점근사의 정도가 모의실험을 통한 정확한 값은 물론 기존의 Gupta와 Brown의 결과와도 매우 유사함을 확인하였다. 따라서 본 논문에서 제시한 안장점근사는 매우 제한된($p = 2$ 인 경우) 통계량에 적용되는 Gulpa와 Brown (2001)에 대한 대안적 방법으로 사용될 수 있으며, 인장-강도 모형에서 보다 일반적인 독립성 가정(X_1 과 X_2 가 독립)하에서도 동일한 절차를 제공한다.

4.2.2. Roberts의 IQ 자료 다음의 Table 4.7은 Roberts (1966)에서 다루어진 1971년 대기업 보험회사에 고용된 87명의 백인 남성과 52명의 비백인 남성의 IQ 자료이다. 이 자료에 대해 정규분포와 비표준화(non-standard) 왜정규분포를 적합한 결과는 다음의 Table 4.7 및 Figure 4.1과 같다. 모수추정 방법으로는 최대가능도추정법을 사용하였으며, 왜정규분포에 대한 최대가능도추정에 대해서는 Gupta와 Brown (2001)을 참고하기 바란다. 그 결과 왜정규분포 가정하에서의 추정 결과를 정규분포

Table 4.7. Maximum likelihood estimates for Roberts IQ data

Parameter	Maximum likelihood estimates			
	White(X_1)		Non-white(X_2)	
	Normal	Skew Normal	Normal	Skew Normal
μ	112.86	105.72	106.65	97.45
σ	9.59	11.91	8.31	12.34
α		1.14		2.55

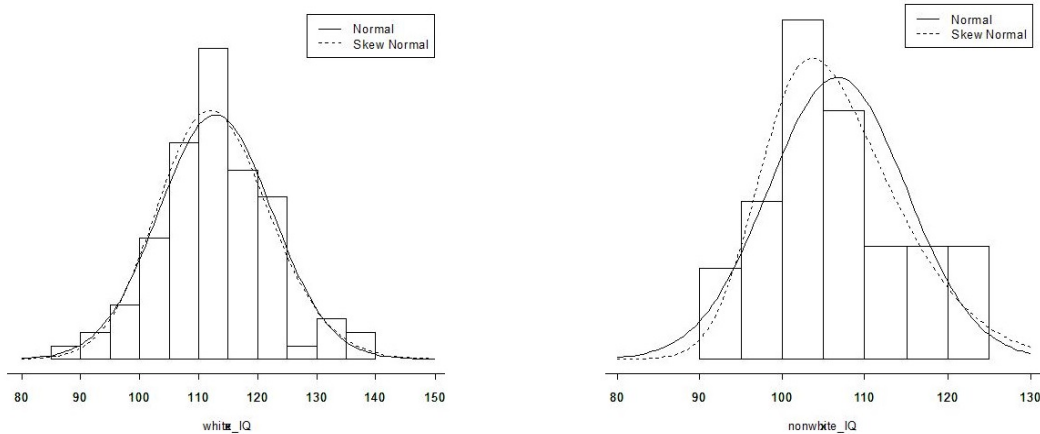


Figure 4.1. Model fitting for Roberts IQ data

Table 4.8. Saddlepoint approximation to the probability $P(Z_1 < Z_2)$ for IQ data

White (Z_1)	Non-white (Z_2)	$P(Z_1 < Z_2)$				
		(a) Z_1, Z_2 : Dependent case			(b) Z_1, Z_2 : Independent case	
$\hat{\alpha}_1$	$\hat{\alpha}_2$	GB	Saddle	S.Exact	Saddle	S.Exact
1.1432	2.5462	0.5739	0.5739	0.5738	0.5501	0.5499

를 가정한 경우와 비교해 보면, 상대적으로 평균(μ)은 크게, 표준편차(σ)는 작게 추정되었음을 알 수 있다.

Z_1 과 Z_2 를 X_1 과 X_2 의 표준화된 확률변수라고 할 때, Z_1 과 Z_2 는 동일한 왜도모수를 가지는 왜정규분포를 따르게 된다. 다음의 Table 4.8은 $P(Z_1 < Z_2)$ 에 대한 안장점근사의 결과이다. 안장점근사는 식 (3.1)이 사용되었으며, 비교를 위해 식 (2.2)의 Gupta와 Brown의 결과(GB)와 10만번 모의실험을 통한 정확한 값(S.Exact)을 함께 제시하였다. 그 결과 인장-강도 경우에서와 같이 안장점근사의 정도가 Gupta와 Brown의 결과와 일치하며, 10만번 모의실험을 통한 결과와도 매우 근사함을 알 수 있다.

5. 결론

본 논문에서는 서로 연관된 왜정규 확률변수들의 선형결합 통계량의 분포함수 추정 문제를 다루었다. 안장점근사를 이용한 분포함수의 추정 방법을 제시하고, 모의실험을 수행한 결과 매우 정확한 근사 결과를 제시함을 확인하였다. 실제 인장-강도 모형과 Roberts (1966)의 자료 분석을 통해 Gupta와 Brown (2001)에 대한 대안으로 사용될 수 있음을 확인하였다. 본 논문에서 제시된 다변량 왜정규분포를 기반

으로 한 선형결합 통계량에 대한 안장점근사는 금융자료에서의 선형결합 포트폴리오의 위험관리를 위한 측도인 VaR(Value at Risk)과 ES(Expected Shortfall)의 계산 등에 활용될 수 있다. 또한 본 연구에서 다루어진 왜정규분포 기반 선형결합에 대한 안장점근사는 Na (2008)에서 다루어진 이차형식 통계량의 경우로 확장될 수 있으며, 정규분포와 왜정규분포의 보다 확장된 형태인 타원(elliptical) 및 왜-타원(skew-elliptical) 분포군에도 유사하게 적용될 수 있다.

References

- 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.
- Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distributions, *Journal of the Royal Statistical Society, Series B*, **61**, 579–602.
- Azzalini, A. (2005). The skew-normal distribution and related multivariate families (with discussion), *Scandinavian Journal of Statistics*, **32**, 159–188.
- Daniels, H. E. (1954). Saddlepoint approximations in statistics, *The Annals of Mathematical Statistics*, **25**, 631–650.
- Daniels, H. E. (1987). Tail probability approximations, *International Statistical Review*, **55**, 37–48.
- Gupta, R. C. and Brown, N. (2001). Reliability studies of the skew-normal distribution and its application to a strength-stress model, *Communications in Statistics-Theory and Methods*, **30**, 2427–2445.
- 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 the Korean Data & Information Science Society*, **19**, 497–504.
- Na, J. H. and Yu, H. K. (2013). Saddlepoint approximation for distribution function of sample mean of skew-normal distribution, *Journal of the Korean Data & information Science Society*, **24**, 1211–1219.
- Roberts, C. (1966). A correlation model useful in the study of twins, *Journal of the American Statistical Association*, **61**, 1184–1190.

다변량 왜정규분포 기반 선형결합통계량에 대한 안장점근사

나중화^{a,1}

^a충북대학교 정보통계학과

(2014년 8월 19일 접수, 2014년 9월 19일 수정, 2014년 10월 10일 채택)

요약

다변량 왜정규분포는 다변량 정규분포를 포함하는 분포로 최근 많은 응용분야에서 활용되고 있다. 본 논문에서는 다변량 왜정규분포를 기반으로 하는 선형결합통계량의 분포함수에 대한 안장점근사를 다루었다. 이는 단변량 왜정규분포 기반 표본평균에 대한 Na와 Yu (2013)의 결과를 선형결합 및 다변량의 경우로 확장한 것이다. 모의실험과 실제자료분석을 통해 제안된 근사법의 유용성과 정확도를 확인하였다.

주요용어: 다변량 왜정규분포, 안장점근사, 누울생성함수.

이 논문은 2013학년도 충북대학교 학술연구지원사업에 의하여 연구되었음. 본 논문은 미래창조과학부 및 정보통신산업진흥원의 고용계약형 SW석사과정 지원사업(NIPA-2014-HB301-14-1011)의 부분지원 연구결과로 수행되었음.

¹(361-763) 충북 청주시 서원구 내수동로 52, 충북대학교 정보통계학과/비즈니스데이터융합학과.

E-mail: cherin@cbnu.ac.kr