응용통계연구 제 14권 2호, 2001년, pp. 333-344

평균 벡터의 평활함수모형에 대한 안부점근사 *

-스튜던트화 분산을 중심으로

나종화 [1)] 김주성 [2)]

요 약

통계적 추론에 사용되는 많은 통계량들은 평균벡터의 평활함수의 형태로 표현이 가능하다. 본 연구에서는 이들 통계량들의 분포함수에 대한 안부점근사법을 제시하였다. 이 방법은 Na(1998)에서 제시된 일반적 통계량의 분포함수에 대한 안부점근사법이 평균벡터의 평활함수모형에 특히 유용하게 사용될 수 있음을 보인 것이다. 이 근사법은 정규근사에 비해 근사의 정도가 뛰어나며, 특히 통계량의 꼬리부분의 확률에 대해서도 정확도가 그대로 유지되는 장점이 있어 정밀한 추론이 요구되는 많은 문제에 효과적으로 사용될 수 있다. 모의 실험에 사용할 평균벡터의 평활함수 모형으로는 스튜던트화 분산을 고려하였다.

주요용어: 평활함수모형, 안부점근사, 누율과 적률, 스튜던트화 분산, 꼬리확률, 확률적 전개.

1. 서론

각종 통계량의 분포에 대한 대표적인 근사로는 중심극한정리 또는 Pollard(1985)의 기법에 기초한 정규근사를 들 수 있다. 그러나 점근정규성에 기초한 이들 근사법들은 큰 표본을 요구하거나, 통계적 추론에 중요한 부분인 꼬리확률에 대해 정확도가 심각히 떨어지는 경우가 발생한다. (Ritcey(1985)). 이에 대한 개선책으로 Edgeworth 근사를 사용할 수 있다. 서로 독립이며 동일한 분포를 따르는 확률변수들의 합을 비롯하여, U- 및 L-통계량을 비롯한 여러가지 일반적 통계량들에 대한 Edgeworth 근사의 타당성에 대해 활발한 연구가 진행되어 왔다. (Field와 Ronchetti(1990) 참고.) 그러나 통계량의 분포함수에 대한 Edgeworth 근사는 일종의 다항함수를 이용한 근사로써 오차항 역시 다항함수의 형태로 특정한 패턴을 보이기 때문에 근사의 결과가 안정적이지 못하며, 분포함수에 대한 근사의 결과가 (0,1)의 범위를 벗어나기도 하는 단점을 가지고 있어 실재의 통계적 추론과 관련된 응용문제에 직접 사용되기에는 많은 문제를 포함하고 있다.

최근 이러한 단점을 보완하기 위한 새로운 근사법으로 안부점근사(saddlepoint approximation)에 대한 연구가 활발히 진행되고 있다. 이 방법은 Daniels(1954)가 처음 통계학 분야

* 이 논문은 1999년도 학술진흥재단의 연구비에 의하여 지원되었음.(KRF-99-003-D00059)
1) (361-763) 충북 청주시 흥덕구 개신동 산 48, 충북대학교 통계학과, 부교수
E-mail: cherin@cbucc.chungbuk.ac.kr
2) (361-763) 충북 청주시 흥덕구 개신동 산 48, 충북대학교 통계학과, 교수
E-mail: kimjs@cbucc.chungbuk.ac.kr

에 소개한 방법으로 기존의 Edgeworth 근사가 가지는 단점을 보완할 뿐 아니라 중간크기의 표본은 물론 소표본의 경우에도 근사의 정도가 뛰어나며, 통계량의 꼬리부분의 확률에도 뛰어난 정확도를 유지하는 장점 때문에 정밀한 추론이 요구되는 많은 통계적 문제에 효과적으로 활용될 수 있다. 안부점근사의 방법은 주로 표본평균을 비롯한 비교적 간단한 형태의 통계량에 대한 연구가 진행되어 왔다. (Barndoff-Nielsen과 Cox(1979), Lugannani와 Rice(1980), Daniels(1987)등 참고.) 최근 Easton과 Ronchetti(1986)는 일반적 통계량의 밀도함수의 근사에 안부점 근사의 방법을 사용하였으며, Na(1998)는 이들의 기법을 통계량의 분포함수 또는 꼬리확률에 직접 적용하는 방법을 제시하였다. 이 방법들은 통계량의 처음 4차 까지의 누율(cumulants)에 기초한 근사법으로 표본분산 및 스튜던트화 평균을 비롯한 비교적 간단한 형태의 통계량들에 대해 다루어 졌다.

본 논문에서는 Na(1998)에서 제시된 방법이 효과적으로 사용될 수 있는 일반적 모형으로 평균벡터(mean vector)의 평활함수모형 (smooth function model)을 제시하고, 이에 대한 안부점 근사의 과정을 2절에서 소개하였다. 특히 이 모형은 통계적 추론에 사용되는 많은 형태의 통계량을 포함하는 모형으로 그 범위가 대단히 넓다고 말할 수 있다. 본 연구에서는 지금까지 잘 다루어지지 않은 복잡한 형태의 통계량으로 알려진 스튜던트화 분산(studentized variance)을 중심으로 다루었으며, 3절에서는 안부점근사의 과정에 필요한 통계량의 누율 계산과정을 다루었고, 4절에서는 안부점근사의 효율성을 모의실험을 통해 확인하였다.

2. 평활함수모형과 안부점근사

2.1. 평활함수모형

먼저 $X_1, X_2, \cdots, X_n$은 서로 독립이고 동일한 분포를 가지는 m-차원의 확률벡터로써 평균을 μ라 하고 표본평균 $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$ 이라 하자. 함수 A를 R^m에서 R로 가는 평활함수로써 다음과 같이 정의하자.

$$A(x) = \{g(x) - g(\mu)\}/h(\mu).$$

여기서 $g(\cdot)$는 적절한 평활함수(smooth function)이고 $h^2(\mu)$은 $\sqrt{n}g(\bar{X}n)$의 극한분산을 의미하며, 이때 $A(\mu) = 0$을 만족한다. 벡터 x의 i번째 원소를 $x^{(i)}$라 하자. $Z = \sqrt{n}(\bar{X}_n - \mu)$이 $O_p(1)$이고

$$a_{i_1 \cdots i_j} = (\partial^j / \partial x^{(i_1)} \cdots \partial x^{(i_j)}) A(x) | x = \mu$$

이라 정의할 때, 통계량 $V_n = \sqrt{n}A(\bar{X}_n)$의 테일러 전개는 다음과 같이 주어진다.

$$V_n = \sqrt{n}A(\bar{X}_n) = V_{nr} + O_p(n^{-r/2}), \quad r \geq 1, \tag{2.1}$$

여기서

$$\begin{aligned} V_{nr} = \sum_{i=1}^{m} a_i Z^{(i)} \quad &+ \quad n^{-1/2}\frac{1}{2}\sum_{i_1=1}^{m}\sum_{i_2=2}^{m} a_{i_1 i_2} Z^{(i_1)} Z^{(i_2)} + \cdots \\ &+ \quad n^{-(r-1)/2}\frac{1}{r!}\sum_{i_1=1}^{m} \cdots \sum_{i_r=1}^{m} a_{i_1 \cdots i_r} Z^{(i_1)} \cdots Z^{(i_r)}. \end{aligned}$$

Hall(1992)은 평균벡터의 평활함수로 표현되는 통계량 $V_n = \sqrt{n}A(\bar{X}_n)$에 대해서는 식(2.1)과 같은 확률적 전개(stochastic expansion)가 가능하고, 이러한 모형에 대한 전개식에서 V_{nr}에 대한 j차 누율(k_{jn})은 다음의 형태로 주어짐을 보인 바 있다. 즉,

$$k_{jn} = n^{-(j-2)/2}(c_{j,1} + n^{-1}c_{j,2} + n^{-2}c_{j,3} + \cdots), \quad j \geq 1 \tag{2.2}$$

이며 $c_{j,l}$은 상수 a들에만 의존하는 값이다. 또한, Hall(1992)은 식(2.2)의 전개를 만족하는 경우 적절한 수학적 가정 하에서 (Cramer의 조건등) 분포함수에 대한 Edgeworth 전개가 타당성을 가짐을 보인 바 있다. 그러나 Edgeworth 전개는 통계적 추론과 관련된 응용문제에 적용될 경우 앞서 언급한 바와 같은 많은 단점을 가지고 있기 때문에 이에 대한 보완이 요구된다고 말할 수 있다.

2.2. 안부점근사

확률표본 $X_1, X_2, \cdots, X_n$에 기초한 통계량 $V_n = V_n(X_1, \cdots, X_n)$의 분포함수에 대한 안부점근사는 V_n의 누율생성함수(Cumulant Generating Function, CGF)를 $K_n(t)$라 할 때 다음의 식으로 주어진다. (이에 대한 자세한 유도과정은 Na(1998)의 부록을 참고.)

$$P_r(V_n \leq v) = \begin{cases} \Phi(w) + \phi(w)\left\{\dfrac{1}{w} - \dfrac{1}{\zeta} + O(n^{-3/2})\right\}, v \neq E(V_n). \\ \dfrac{1}{2} - \dfrac{R_n^{(3)}(0)}{6\sqrt{2\pi(R_n''(0))^3}}\{1 + O(n^{-3/2})\}, v = E(V_n). \end{cases} \tag{2.3}$$

위 식에서 $\phi(\cdot)$와 $\Phi(\cdot)$는 각각 표준정규분포의 밀도함수와 분포함수를 나타내고 w와 ζ는 다음의 식으로 정의되는 값이다.

$$\begin{aligned} w &= [2n\{t_0 v - R_n(t_0)\}]^{1/2} sgn(t_0), \\ \zeta &= t_0\{nR_n''(t_0)\}^{1/2}. \end{aligned}$$

여기서 $R_n(t) = K_n(nt)/n$이고 $R_n'(\cdot), R_n''(\cdot), R_n^{(3)}(\cdot)$은 각각 $R_n(\cdot)$의 1, 2, 3차 미분을 의미하며 t_0는 다음의 안부점 방정식(saddlepoint equation) 을 만족하는 근이며, sgn(t_0)는 부호함수(sign function)를 의미한다. 이 경우 t_0는 3차 방정식의 근이 되므로 세 개의 값이 존재하나 본 논문에서는 0보다 큰 실수값 중 최소의 값을 택하여 사용한다. 그 이유는 Daniels(1954)의 밀도함수에 대한 안부점 근사의 유도과정을 참고하면 좋을 것이다.

$$R_n'(t_0) = v. \tag{2.4}$$

한편, 통계량 V_n이 $n^{1/2}A(\bar{X}_n)$의 형태로 주어지는 평활함수모형의 분포함수에 대한 안부점근사를 실시하기 위해서는 이 통계량에 대한 CGF가 필요하다. 그러나 이 모형에 포함되는 대부분의 통계량에 대해서 정확한 CGF가 알려져 있지 않다. 따라서, 본 논문에서는 통계량 $V_n = \sqrt{n}A(\bar{X}_n)$의 CGF를 다음의 $\tilde{K}_n(t)$로 근사하기로 한다.

$$\tilde{K}_n(t) = k_{1n}t + \frac{k_{2n}}{2!}t^2 + \frac{k_{3n}}{3!}t^3 + \frac{k_{4n}}{4!}t^4. \tag{2.5}$$

위 식에서 $k_{1n} = E(V_n)$, $k_{2n} = Var(V_n)$이고, k_{3n}, k_{4n}은 각각 통계량 V_n의 3차와 4차 누율을 의미한다.

또한, 식(2.5)의 근사에 요구되는 통계량에 대한 처음 4차까지의 누율 역시 정확히 계산되는 경우(예를들어, 표본분산의 경우)는 극히 제한적이므로 이를 다시 식(2.2)의 형태로 근사하기로 한다. 즉, $c_{1,1} = 0$이고 $c_{2,1} = 1$일 때,

$$\begin{aligned} k_{1n} &= n^{-1/2}c_{1,2} + n^{-3/2}c_{1,3} + O(n^{-5/2}) \\ k_{2n} &= 1 + n^{-1}c_{2,2} + n^{-2}c_{2,3} + O(n^{-3}) \\ k_{3n} &= n^{-1/2}(c_{3,1} + n^{-1}c_{3,2} + n^{-2}c_{3,3} + O(n^{-3})) \\ k_{4n} &= n^{-1}(c_{4,1} + n^{-1}c_{4,2} + n^{-2}c_{4,3} + O(n^{-3})) \end{aligned} \tag{2.6}$$

이며, 본 논문에서는 위 식의 처음 수정항(first correction term)까지의 근사식을 사용하기로 한다. 처음 4차까지의 적률(moment) $\mu'_{rn}(r \geq 1)$에 대한 근사식은 누율과 적률에 대한 다음의 관계식

$$\begin{aligned} k_{1n} &= E(V_n), \quad k_{2n} = E(V_n^2) - (EV_n)^2 \\ k_{3n} &= E(V_n^3) - 3E(V_n^2)E(V_n) + 2(EV_n)^3 \\ k_{4n} &= E(V_n^4) - 4E(V_n^3)E(V_n) - 3(EV_n^2)^2 \\ &\quad +12E(V_n^2)(EV_n)^2 - 6(EV_n)^4 \end{aligned} \tag{2.7}$$

으로부터(Kendall 과 Stuart(1977) 참고) 아래에 주어진 차수(order)까지의 전개가 필요함을 쉽게 확인할 수 있다.

$$\begin{aligned} \mu'_{1n}(V_n) &= E(V_n) = n^{-1/2}A + O(n^{-3/2}), \\ \mu'_{2n}(V_n) &= E(V_n^2) = 1 + n^{-1}B + O(n^{-2}), \\ \mu'_{3n}(V_n) &= E(V_n^3) = n^{-1/2}C + O(n^{-3/2}), \\ \mu'_{4n}(V_n) &= E(V_n^4) = 3 + n^{-1}D + O(n^{-2}). \end{aligned} \tag{2.8}$$

이상에서 소개한 안부점근사는 누율생성함수가 알려져 있지 않은 통계량에 대해서도 처음 4차까지의 누율에 대한 적절한 차수까지의 확률적 전개를 통해 쉽게 구할 수 있다. 따라서 본 논문에서는 2.1절에서 소개된 바와 같이 누율에 대한 확률적 전개가 가능한 Hall(1992)이 제안한 평균벡터의 평활함수 모형에까지 안부점근사가 가능함을 제안하고, 그 예로써 상당히 복잡한 형태의 통계량으로 알려진 스튜던트화 분산을 중심으로 구체적인 근사과정을 다루고자 한다.

3. 스튜던트화 분산과 누율계산

$X_1, X_2, \cdots, X_n$을 평균이 μ이고 분산$\sigma^2(> 0)$인 모집단으로부터의 확률표본이라 하자. 표준화된 확률변수 $W_i = (X_i - \mu)/\sigma(i = 1, \cdots, n)$이고 $Z_r(r = 1, 2, 3, 4)$을 각각 다음과 같

이 정의하자.

$$(Z_1, Z_2, Z_3, Z_4) = \sqrt{n}(\bar{W}, \overline{W^2} - 1, \overline{W^3} - \nu_3, \overline{W^4} - \nu_4).$$

여기서 $\overline{W^r} = \sum_{i=1}^n W_i^r/n$이고 $\nu_r(r = 1, 2, \cdots)$은 $E(W_i^r)$이다. 또한 $\hat{\sigma_n^2} = \sum_{i=1}^n (X_i - \bar{X})^2/n$이라 할 때, $\hat{\sigma_n^2}$의 기대치와 극한분산은 다음과 같이 주어진다.

$$\begin{aligned} E(\hat{\sigma_n^2}) &= (1 - \frac{1}{n})\sigma^2, \\ Var(\hat{\sigma_n^2}/\sigma^2) &= n^{-1}(\rho_4 + 2) + O(n^{-2}). \end{aligned}$$

여기서 $\rho_4 = \nu_4 - 3$이다. 따라서 $\sqrt{n}\hat{\sigma_n^2}$의 극한분산을 τ^2이라 할 때, 이의 추정량은 다음과 같다.

$$\hat{\tau^2} = \widehat{k_4 + 2\sigma^4} = \widehat{\mu_4 - \sigma^4} = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^4 - (\hat{\sigma_n^2})^2.$$

위 식에서 k_4와 μ_4는 각각 X_i의 제 4차 누율(cumulant)과 중심적률(central moment)을 의미한다. 이제 스튜던트화 분산을 $V_n = \sqrt{n}(\hat{\sigma_n^2} - \sigma^2)/\hat{\tau}$으로 정의할 때, 이에 대한 확률적 전개를 생각하자.

$$V_n = \sqrt{n}(\hat{\sigma_n^2}/\sigma^2 - 1)/(\hat{\tau^2}/\sigma^4)^{1/2} = (Z_2 - Z_1^2/\sqrt{n})/(\hat{\tau^2}/\sigma^4)^{1/2} \quad (3.1)$$

이고

$$\begin{aligned} \hat{\tau^2}/\sigma^4 &= \overline{W^4} - 4\overline{W^3}\bar{W} + 6\overline{W^2}(\bar{W})^2 - 4(\bar{W})(\bar{W})^3 + (\bar{W})^4 - \{\overline{W^2} - (\bar{W})^2\}^2 \\ &= \overline{W^4} - 4(\overline{W^3})(\bar{W}) + 8(\overline{W^2})(\bar{W})^2 - (\overline{W^2})^2 - 4(\bar{W})^4 \end{aligned}$$

이며, $\nu_1 = 0,\ \nu_2 = 1$이라 할 때 $\overline{W^r} = \nu_r + \frac{1}{\sqrt{n}}Z_r$이므로 $\hat{\tau^2}/\sigma^4$의 확률적 전개는 다음과 같이 주어진다.

$$\begin{aligned} \hat{\tau^2}/\sigma^4 = (\nu_4 - 1) \quad &+ \quad n^{-1/2}(Z_4 - 4\nu_3 Z_1 - 2Z_2) \\ &+ \quad n^{-1}(-4Z_3Z_1 + 8Z_1^2 - Z_2^2) + O_p(n^{-3/2}). \end{aligned} \quad (3.2)$$

식(3.2)를 식(3.1)에 대입하면 V_n에 대한 다음의 전개식을 얻을 수 있다.

$$\begin{aligned} V_n &= a^{-1/2}Z_2 + n^{-1/2}\left\{-a^{-1/2}Z_1^2 - \frac{1}{2}Z_2(Z_4 - 4\nu_3Z_1 - 2Z_2)a^{-3/2}\right\} \\ &+ n^{-1}\left\{\frac{1}{2}Z_2(4Z_3Z_1 - 8Z_1^2 + Z_2^2)a^{-3/2} + \frac{1}{2}Z_1^2(Z_4 - 4\nu_3Z_1 - 2Z_2)a^{-3/2}\right. \\ &+ \left.\frac{3}{8}Z_2(Z_4 - 4\nu_3Z_1 - 2Z_2)^2a^{-5/2}\right\} + O_p(n^{-3/2}). \end{aligned} \quad (3.3)$$

여기서, $a = \nu_4 - 1$이다. 위 식으로부터 V_n의 처음 4차까지의 적률(moments)을 구하면 다음과 같다. 이에 대한 자세한 유도과정은 부록에 수록하였다.

$$\begin{aligned}
\mu'_{1n}(V_n) &= n^{-1/2}a^{-3/2}(4\nu_3^2 + \nu_4 - \nu_6)/2 + O(n^{-3/2}), \\
\mu'_{2n}(V_n) &= 1 + n^{-1}\{a^{-2}(2\nu_6 + 8\nu_3\nu_5 - 17\nu_4 + 5\nu_4^2 - 32\nu_3^2 + 10) \\
&\quad +2a^{-3}(\nu_6 - 2\nu_4 - 4\nu_3^2)^2\} + O(n^{-2}), \\
\mu'_{3n}(V_n) &= n^{-1/2}a^{-3/2}(-7\nu_6/2 + 15\nu_4/2 + 12\nu_3^2 - 4) + O(n^{-3/2}), \\
\mu'_{4n}(V_n) &= 3 + n^{-1}\{a^{-3}(-184\nu_4\nu_6 - 208\nu_3^2\nu_6 + 720\nu_3^2\nu_4 + 384\nu_3^4 - 512\nu_3^2 \\
&\quad +348\nu_4^2 + 28\nu_6^2 + 448\nu_4 + 128\nu_6 + 144) \\
&\quad +a^{-2}(96\nu_3\nu_5 + 44\nu_6 - 2\nu_8 - 72\nu_4 - 456\nu_3^2 + 30) \\
&\quad +a^{-1}(24\nu_4 - 78) + 21\} + O(n^{-2}).
\end{aligned} \tag{3.4}$$

여기서 $\nu_r(r = 3, 4, \cdots)$은 표준화된 확률변수 W_i의 r차 적률을 의미한다. 또한 위의 결과를 식(2.7)에 대입하면 통계량의 처음 4차까지의 누율들을 쉽게 계산할 수 있다.

4. 모의실험

이 절에서는 평활함수모형의 분포함수에 대한 안부점근사에 대한 모의실험 결과를 소개하였다. 여기서는 스튜던트화 분산을 중심으로 그 결과를 제시하고 있으나, 이보다 훨씬 간단한 형태의 평활함수 모형의 통계량인 표본분산 및 스튜던트화 평균등에 대한 모의실험의 결과는 이미 Na(1998)에서 소개된 바 있으며 관심있는 독자는 이를 참고하면 좋을 것이다.

3절에서 유도한 스튜던트화 분산의 누율에 대한 정보를 이용하여, 식(2.3)의 안부점근사를 실시한 결과는 표 4.1과 표 4.2와 같다. 모집단의 분포는 편의상 표준정규분포를 고려하였으며, 표본의 크기는 10과 30인 경우를 고려하였다. 각 표에 대한 설명은 다음과 같다.

Simulated Exact : 반복횟수 100,000번의 모의실험을 수행한 결과.
Saddle I: 안부점근사((식(2.3))의 첫 항을 이용한 결과.
Saddle II : 안부점근사의 결과

위의 모의실험 결과에서 알 수 있듯이 안부점근사의 결과는 기존의 정규근사에 비해 상당히 정확한 결과를 제공하며, 표본의 수가 커질수록 근사의 정도가 높아짐을 알 수 있다. 안부점근사의 결과가 모의실험을 통한 정확한 값(Simulated Exact) 에 비해 다소 작은 값으로 추정(underestimate)이 되고 있으나 일반적인 사항은 아니라고 말할 수 있다. 또한 위의 결과는 편의상 통계적 추론에 자주 사용되는 우측 꼬리확률의 영역에서만 모의실험을 수행하였으며, 제시된 안부점근사의 결과(Saddle II, 식(2.3))는 Jensen(1992)이 유도한 다음의 근사식 (식(2.3)과 근사의 정도가 동일함을 밝힘.)

$$Pr(V_n \le v) = \Phi(w + \frac{1}{w}log(\frac{\zeta}{w})) + O(n^{-3/2}) \tag{4.1}$$

표 4.1: 스튜던트화 분산에 대한 안부점근사(표본의 크기 : n=10인 경우)

Approximations to $\Pr(V_n \le t)$			
t	Simulated Exact	Saddle I	Saddle II
1.0	0.894	0.793	0.841
1.2	0.919	0.816	0.861
1.4	0.940	0.838	0.880
1.6	0.954	0.858	0.896
1.8	0.967	0.875	0.910
2.0	0.974	0.891	0.923
2.2	0.981	0.905	0.935
2.4	0.986	0.918	0.945
2.6	0.990	0.929	0.953
2.8	0.993	0.938	0.961
3.0	0.994	0.947	0.968
3.2	0.997	0.954	0.974
3.4	0.997	0.961	0.979
3.6	0.998	0.966	0.984
3.8	0.998	0.971	0.988
4.0	0.999	0.976	0.991

을 이용한 결과와 거의 일치함을 확인하였다. (모의실험 결과 소수 아래 두자리까지 완전히 일치하는 관계로 결과는 생략하였음.) 마지막으로 근사식 (4.1)이 가지는 장점가운데 하나는 식(2.3)의 안부점 근사의 결과가 음의 값이 나올 수 있는 구조적 결함에 대한 보완이 될 수 있다는 측면에서 안부점근사에 대한 하나의 대체식으로 활용될 수 있을 것이다.

5. 결론

본 논문에서는 평균벡터의 평활함수모형에 대한 안부점근사법을 제안하였다. 고려된 평활함수모형으로는 스튜던트화 평균을 고려하였다. 통계량의 처음 4차까지의 누율에 기초한 이 근사법은 평활함수모형에 포함되는 많은 종류의 통계량의 분포에 대해 정확도가 뛰어난 근사를 제공하기 때문에 정밀한 추론이 요구되는 많은 통계적 문제에 효과적으로 사용할 수 있다. 다만 제시된 방법의 효과적인 사용을 위해서는 통계량의 누율에 대해 요구되는 차수까지의 전개가 필요하여 이에 대한 텐서(tensor)기법 등의 효과적인 계산과정이 요구된다고 할 수 있다.

표 4.2: 스튜던트화 분산에 대한 안부점근사(표본의 크기 : n=30인 경우)

Approximations to $\Pr(V_n \leq t)$			
t	Simulated Exact	Saddle I	Saddle II
1.0	0.872	0.829	0.874
1.2	0.903	0.858	0.898
1.4	0.929	0.883	0.918
1.6	0.949	0.904	0.935
1.8	0.964	0.921	0.950
2.0	0.974	0.936	0.961
2.2	0.981	0.948	0.970
2.4	0.987	0.958	0.978
2.6	0.992	0.967	0.985
2.8	0.994	0.973	0.990
3.0	0.996	0.979	0.994
3.2	0.997	0.983	0.997
3.4	0.999	0.987	0.999
3.6	0.999	0.990	0.999
3.8	0.999	0.992	0.999
4.0	1.000	0.994	0.999

부록

여기에서는 식 (3.4)에서 제시된 스튜던트화 분산에 대한 처음 4차까지의 누율에 대한 자세한 계산과정을 수록하였다. 각 식의 계산과정에서 나타나는 기호 $[i](i:)$은 유사한 항의 개수가 i개임을 의미하며, McCullagh(1987)의 누율계산법에서 사용된 표현과 동일하므로 이를 참고하면 좋을 것이다.

A. 식(3.3)의 V_n에 대한 확률적 전개식으로부터 $\mu'_{1n}(V_n) = E(V_n)$을 구하면 다음과 같다.

$$
\begin{aligned}
E(V_n) &= \frac{1}{2}n^{-1/2}a^{-3/2}[-2a - E\{(W^2-1)(W^4-\nu_4)\} \\
&\quad +4\nu_3 E\{(W^2-1)W\} + 2E\{(W^2-1)^2\}] + O(n^{-3/2}) \\
&= \frac{1}{2}n^{-1/2}a^{-3/2}\{-2a - (\nu_6-\nu_4) + 4\nu_3\cdot\nu_3 + 2\cdot a\} + O(n^{-3/2}) \\
&= \frac{1}{2}n^{-1/2}a^{-3/2}(4\nu_3^2 + \nu_4 - \nu_6) + O(n^{-3/2}).
\end{aligned}
$$

B. 식(3.3)으로부터 $\mu'_{2n}(V_n) = E(V_n^2)$을 구하면 다음과 같다.

$$\begin{aligned}
V_n^2 &= a^{-1}Z_2^2 + n^{-1/2}\{-2a^{-1}Z_1^2Z_2 - a^{-2}Z_2^2(Z_4 - 4\nu_3Z_1 - 2Z_2)\} \\
&\quad + n^{-1}\{Z_2^2a^{-2}(4Z_3Z_1 - 8Z_1^2 + Z_2^2) + 2Z_2Z_1^2a^{-2}(Z_4 - 4\nu_3Z_1 - 2Z_2) \\
&\quad + Z_2^2a^{-3}(Z_4 - 4\nu_3Z_1 - 2Z_2)^2 + Z_1^4a^{-1}\} + O_p(n^{-3/2}).
\end{aligned}$$

$$\begin{aligned}
E(V_n^2) &= 1 + n^{-1/2}\cdot n^{-1/2}\cdot a^{-2}\{-2a(\nu_4 - 1) - (\nu_8 - \nu_4^2 - 2\nu_6 + 2\nu_4) \\
&\quad + 4\nu_3(\nu_5 - \nu_3) + 2(\nu_6 - 3\nu_4 + 2)\} \\
&\quad + n^{-1}\cdot a^{-3}[4a\{a\cdot\nu_4 + (\nu_5 - \nu_3)\cdot\nu_3[2]\} \\
&\quad - 8a(a\cdot 1 + \nu_3\cdot\nu_3[2]) + a\cdot a\cdot a[3] + 2a\{1\cdot(\nu_6 - \nu 4) + \nu_3\cdot\nu_5[2]\} \\
&\quad - 8a\nu_3(1\cdot\nu_3[3]) - 4a(1\cdot a + \nu_3\cdot\nu_3[2]) \\
&\quad + a\cdot Var\{W^4 - \nu_4 - 4\nu_3W - 2(W^2 - 1)\} \\
&\quad + Cov^2\{W^2 - 1, W^4 - \nu_4 - 4\nu_3W - 2(W^2 - 1)\}[2] + a\cdot a[3]] + O(n^{-2}) \\
&= 1 + n^{-1}\{a^{-2}(2\nu_6 + 8\nu_3\nu_5 - 17\nu_4 + 5\nu_4^2 - 32\nu_3^2 + 10) \\
&\quad + 2a^{-3}(\nu_6 - 2\nu_4 - 4\nu_3^2)^2\} + O(n^{-2}).
\end{aligned}$$

C. 식(3.3)으로부터 $\mu'_{3n}(V_n) = E(V_n^3)$을 구하면 다음과 같다.

$$\begin{aligned}
V_n^3 &= a^{-3/2}Z_2^3 + n^{-1/2}\left\{-3a^{-3/2}Z_2^2Z_1^2 - \frac{3}{2}a^{-5/2}Z_2^3(Z_4 - 4\nu_3Z_1 - 2Z_2)\right\} \\
&\quad + O_p(n^{-1}).
\end{aligned}$$

$$\begin{aligned}
E(V_n^3) &= n^{-1/2}a^{-3/2}(\nu_6 - 3\nu_4 + 2) \quad [\because (W^2 - 1)^3 = W^6 - 3W^4 + 3W^2 - 1] \\
&\quad + n^{-1/2}[-3a^{-3/2}(a\cdot 1 + \nu_3\cdot\nu_3[2]) \\
&\quad - \frac{3}{2}a^{-5/2}\{a(\nu_6 - \nu_4)[3] - 4\nu_3a\cdot\nu_3[3] - 2a\cdot a[3]\}] + O(n^{-3/2}).
\end{aligned}$$

D. 식(3.3)으로부터 $\mu'_{4n}(V_n) = E(V_n^4)$을 구하면 다음과 같다.

$$\begin{aligned}
V_n^4 &= a^{-2}Z_2^4 + n^{-1/2}\{-4a^{-2}Z_2^3Z_1^2 - 2a^{-3}Z_2^4(Z_4 - 4\nu_3Z_1 - 2Z_2)\} \\
&\quad + n^{-1}\{2a^{-3}Z_2^4(4Z_3Z_1 - 8Z_1^2 + Z_2^2) + 8a^{-3}Z_2^3Z_1^2(Z_4 - 4\nu_3Z_1 - 2Z_2) \\
&\quad + 3a^{-4}Z_2^4(Z_4 - 4\nu_3Z_1 - 2Z_2)^2 + 6a^{-2}Z_2^2Z_1^4\} + O_p(n^{-3/2}).
\end{aligned}$$

$$\begin{aligned}
E(V_n^4) \;=\; & a^{-2}\{3a^2 + n^{-1}Cum(W^2-1, W^2-1, W^2-1, W^2-1)\} \\
& +n^{-1/2}n^{-1/2}a^{-3}[-4a\{1\cdot(\nu_6 - 3\nu_4 + 2)[1] + \nu_3(\nu_5 - 2\nu_3)[6] + a\cdot a[3]\} \\
& -2\{(\nu_6 - \nu_4)\cdot(\nu_6 - 3\nu_4 + 2)[4] + a\cdot(\nu_8 - \nu_4^2 - 2\nu_6 + 2\nu_4)[6]\} \\
& +8\nu_3\{\nu_3\cdot(\nu_6 - 3\nu_4 + 2)[4] + a(\nu_5 - 2\nu_3)[6]\} + 4a\cdot(\nu_6 - 3\nu_4 + 2)[10]] \\
& +n^{-1}a^{-4}[8a\{\nu_4\cdot a\cdot a[3] + \nu_3\cdot(\nu_5 - \nu_3)\cdot a[12]\} \\
& -16a(1\cdot a\cdot a[3] + \nu_3\cdot\nu_3\cdot a[12]) + 2a(a\cdot a\cdot a[15]) \\
& +8a\{a\cdot(\nu_6 - \nu_4)\cdot 1[3] + a\cdot\nu_3\cdot\nu_5[6] + \nu_3\cdot\nu_3\cdot(\nu_6 - \nu_4)[6]\} \\
& -32a\nu_3(a\cdot\nu_3\cdot 1[9] + \nu_3\cdot\nu_3\cdot\nu_3[6]) - 16a(1\cdot a\cdot a[3] + \nu_3\cdot\nu_3\cdot a[12]) \\
& +3\{Var(W^4 - \nu_4 - 4\nu_3 W - 2(W^2-1))\cdot a\cdot a[3] \\
& +Cov^2(W^4 - \nu_4 - 4\nu_3 W - 2(W^2-1), W^2-1)\cdot a[12]\} \\
& +6a^2\{a\cdot 1\cdot 1[3] + \nu_3\cdot\nu_3\cdot 1[12]\}] + O(n^{-2}).
\end{aligned}$$

위 식에서 다음의 결과를 이용하면 식(3.4)의 결과를 쉽게 얻을 수 있다.

$$\begin{aligned}
Cum(W^2-1, W^2-1, W^2-1, W^2-1) &= \kappa_4(W^2-1) = \mu_4' - 3(\mu_2')^2 \\
&= E(W^2-1)^4 - 3(E(W^2-1)^2)^2 \\
&= -3\nu_4^2 + \nu_8 - 4\nu_6 + 12\nu_4 - 6.
\end{aligned}$$

$$\begin{aligned}
Var(W^4 - \nu_4 - 4\nu_3 W - 2(W^2-1)) &= E(W^4 - \nu_4 - 4\nu_3 W - 2W^2 + 2)^2 \\
&= \nu_8 - \nu_4^2 + 32\nu_3^2 + 8\nu_4 - 8\nu_3\nu_5 - 4\nu_6 - 4.
\end{aligned}$$

$$\begin{aligned}
Cov^2(W^4 - \nu_4 - 4\nu_3 W - 2(W^2-1) \quad &, \quad W^2-1) \\
&= E(W^4 - \nu_4 - 4\nu_3 W - 2W^2 + 2)(W^2-1) \\
&= (\nu_6 - 3\nu_4 - 4\nu_3^2 + 2)^2.
\end{aligned}$$

참고문헌

[1] Barndorff-Nielsen, O.E. and Cox, D.R. (1979). Edgeworth and saddlepoint approximations with statistical applications, *Journal of Royal Statistical Society*, Series B, **41**(3), 279-312.

[2] Daniels, H.E. (1954). Saddlepoint approximations in statistics, *Annals of Mathematical Statistics*, **25**, 631-650.

[3] Daniels, H.E. (1987). Tail probability approximation, *International Statistical Review*, **55**(1), 37-48.

[4] Easton, G.S. and Ronchetti, E. (1986). General saddlepoint approximations with applications to L statistics, *Journal of American Statistical Association*, **81**(394), 420-430.

[5] Field, C.A. and Ronchetti, E. (1990). *Small Sample Asymptotics*. Institute of Mathematical Statistics, Hayward, California.

[6] Hall, P. (1992).*The Bootstrap and Edgeworth Expansion*, Springer-Verlag, New York.

[7] Jensen, J.L. (1992). The modified signed likelihood statistic and saddlepoint approximations, *Biometrika*, **79**(4), 693-703.

[8] Kendall, M.G. and Stuart, A. (1977). *The Advanced Theory of Statistics*, Vol 1. London, Griffin.

[9] Lugannani, R. and Rice, S. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables, *Advances in Applied Probability*, **12**, 475-490.

[10] McCullagh, P. (1987). *Tensor Methods in Statistics*, Chapman and Hall, London, New York.

[11] Na, J.H. (1998). Saddlepoint approximation to the distribution of general statistic, *The Korean Journal of Applied Statistics*, **11**(2), 287-302.

[12] Pollard, D. (1985). New ways to prove central limit theorems, *Econometric Theory*, **1**, 295-314.

[13] Ritcey, J.A. (1985). *Calculating Radar Detection Probabilities by Contour Integration*, Ph. D. Thesis, University of California, San Diego.

[2001년 2월 접수, 2001년 6월 채택]

Saddlepoint Approximation to the Smooth Functions of Means Model *

Jong-Hwa Na[1)] Ju-Sung Kim [2)]

ABSTRACT

We propose a saddlepoint approximation to the smooth functions of mean vectors. The approximation is based on the first four cumulants of the statistics. Stochastic expansions of the cumulants of the statistics are also needed to adopt the approximation. Simulation studies are carried out for studentized variance. The suggested approximations are very accurate for moderate or small sample sizes.

Keywords: Smooth function model; Saddlepoint approximation; Cumulant and moment; Studentized variance; Tail probability; Stochastic expansion.

* This work was supported by Korea Research Foundation Grant.(KRF-99-003-D00059)

1) Associate Professor, Department of Statistics, Chungbuk National University.
E-mail: cherin@cbucc.chungbuk.ac.kr

2) Professor, Department of Statistics,Chungbuk National University.
E-mail: kimjs@cbucc.chungbuk.ac.kr