

Estimation of the Mean and Variance for Normal Distributions whose Both Sides are Truncated

Chong Sun Hong¹⁾ and Yun Young Choi²⁾

Abstract

In order to estimate the mean and variance for a Normal distribution which is truncated at both right and left sides, maximum likelihood estimators based on the entire sample from the original distribution are compared with the sample mean and variance of the censored sample which is the data remaining after truncation using simulation. We found that, surprisingly, the mean squared error of the mean based on the censored data is smaller than that of the full sample estimators.

Keywords : Bias, Censored-sample estimators, Full-sample estimators, MSE.

1. 서론

다양한 방법으로 절단된 정규분포(truncated normal distribution)의 평균과 분산이 필요한 경우가 많다. 예를 들어 운전면허시험 응시자들 전체의 점수보다 운전면허에 합격한 사람들의 점수에 더 관심이 있을 수도 있다. 만약 원시 모집단인 전체 응시자들의 점수자료가 정규분포를 따른다면 합격한 사람들의 점수자료는 절단된 정규분포를 가지는데, 이는 합격점수가 주어졌을 때 합격점수 이상의 점수들이 임의로 선택된 조건부 분포와 같다.

Hald(1952)를 비롯한 많은 사람들이 절단된 정규분포의 누적분포함수와 밀도함수를 유도해 냈고, Kececioglu(1991)는 절단분포가 원시 모집단의 부분으로 표현될 때 누적분포함수에 대한 간단한 식을 유도해 냈다. Barr와 Sherrill(1999)은 원시 정규분포의 모수와 누적분포함수의 표현으로 절단된 정규분포의 평균과 분산을 유도하였으며, 여기서 분산은 자유도가 3인 카이제곱분포로 표현이 가능하여 평균과 분산의 수치적 값들을 쉽게 계산할 수 있었다. Johnson과 Kotz(1970)는 평균과 분산을 유도해냈지만 분산에 대한 유도식은 시뮬레이션하기에 복잡하고 Barr와 Sherrill(1999)의 연구 결과보다 장황하다. 그 외에도 Fisher(1931)는 평균과 분산을 'Hh함수'를 이용하여 표현했다.

절단된 분포의 자료(censored data)에 기초한 원시 모집단의 모수들의 추정에 관한 오랜 문헌들이 많다. Gupta(1952), Halperin(1952)등의 사람들과 함께 Cohen은 많은 문헌연구(1949, 1950,

1) Professor, Department of Statistics, Sungkyunkwan University, Seoul 110-745 KOREA.

E-mail : cshong@skku.ac.kr

2) Researcher, Research Institute of Applied Statistics, Sungkyunkwan University, Seoul 110-745 KOREA.

E-mail : chic602@hanmail.net

1959, 1961)들을 발표하였으며, 특히 Cohen(1991, Ch. 2)의 연구는 매우 훌륭하여 많은 논문에 인용되었다. Johnson과 Kotz(1970)는 다양하게 자료를 절단하는 경우에 대하여 최대가능도 추정량들에 대한 연구결과를 제시하였다.

그러나 실제 몇몇의 적용에서는 그 반대에 관심이 있다: 원시 모집단의 전체표본을 확보하여 이 자료를 이용할 수 있을 때(어떤 경우에는 원시분포의 모수가 알려져 있는 경우도 있다), 절단 후에 절단된 집단의 평균과 분산을 추정 또는 계산하길 원하는 경우도 있다. 이런 경우에 절단된 분포의 평균과 분산을 추정하는 방법으로 Barr와 Sherrill(1999)은 다음과 같이 두 가지 방법을 고려했다. 첫 번째 방법은 최대가능도 추정량의 불변성(invariance)의 성질에 의해 전체표본자료에 기초한 표본평균과 표본분산 추정량을 절단된 분포의 평균과 분산에 대체하여 구하고, 다른 방법으로는 절단 영역에 속하는 관측값을 제거하여 절단된 모집단에 대한 확률표본으로 간주하여 절단자료로부터 직접 표본평균과 표본분산을 구한다. 더 적은 자료에 기초했기 때문에 절단된 표본에 기초한 추정량이 전체 표본에 기초한 추정량보다 덜 효율적이라고 추측할 지 모르지만 다음에서 보여질 것처럼 평균 추정의 경우 절단된 표본에 기초한 추정량이 더 효율적임이 밝혔다.

본 논문에서는 Barr와 Sherrill(1999)의 연구를 확장하여 정규분포가 양쪽 방향으로 절단된 경우에 평균과 분산을 추정하기 위하여 시뮬레이션을 통하여 위에서 언급한 두 가지 방법으로 구한 추정량들을 비교하여 효율성을 연구하고자 한다. 그리고 Barr와 Sherrill(1999)의 연구에서 충분히 설명하지 못한 효율성에 대하여도 본 연구에서는 설명하였으며, 양쪽 절단의 여러 경우에 대하여 비교 연구를 하였다.

정규분포의 양쪽 절단의 예는 어느 대학 또는 전공학과에 합격한 수험생들의 점수들에 대한 모형을 생각할 수 있다. 수험생들은 자신의 수학능력평가시험 점수에 맞추어 진학을 할 것이고 이에 따라 각 대학마다 합격한 수험생들의 상한점수와 하한점수가 존재하며, 복수지원이 가능해지면서 상한점수와 하한점수의 구분은 더욱 명확하게 되었다. 그러므로 수험생들의 점수가 정규분포를 따른다면, 어느 한 대학에 합격한 수험생들의 점수들의 분포는 양쪽이 절단된 정규분포의 형태를 갖게 된다.

2. 양쪽 절단의 평균과 분산의 유도

Barr와 Sherrill(1999)은 한쪽 방향(왼쪽 또는 오른쪽)으로 절단된 정규분포의 평균과 분산을 표준정규누적분포함수와 자유도가 3인 카이제곱분포의 적분형태를 이용하여 구하였으며, 비표준정규 분포의 평균과 분산에 대하여도 언급하였다. 본 연구에서는 양쪽 방향으로 절단된 정규분포의 경우에 평균과 분산을 다음과 같이 유도하였다.

Z 가 절단점 t_1 과 t_2 ($t_1 < t_2$)의 양쪽이 절단된 표준정규 확률변수라 하면(절단영역은 $(-\infty, t_1] \cup [t_2, \infty)$), Z 의 밀도함수는 다음과 같다.

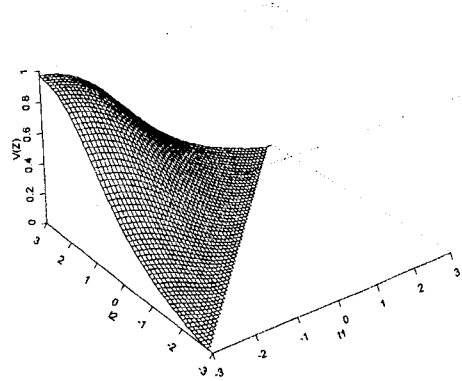
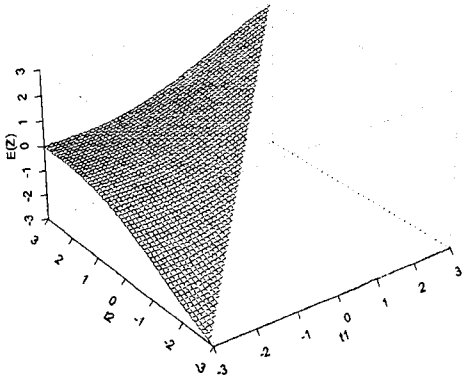
$$f(z) = c(t_1, t_2) e^{-\frac{1}{2}z^2}, \quad t_1 < z < t_2,$$

여기서 $c(t_1, t_2) = 1/[\sqrt{2\pi}\{\Phi(t_2) - \Phi(t_1)\}]$ 이고 $\Phi(t)$ 는 표준정규 누적분포함수이다.

Z 의 평균을 유도하면 다음과 같다.

$$E(Z) = c(t_1, t_2) \int_{t_1}^{t_2} z e^{-\frac{1}{2}z^2} dz = c(t_1, t_2) \left\{ e^{-\frac{1}{2}t_1^2} - e^{-\frac{1}{2}t_2^2} \right\}. \quad (1)$$

[그림 1]은 (1)식에서 정의되었으며 절단점 t_1 과 t_2 ($t_1 < t_2$)의 함수로 표현되는 양쪽방향으로 절단된 표준정규분포의 평균을 이차원 그래프로 나타내었다. [그림 1]을 통하여 왼쪽 절단점 t_1 이 감소할수록 평균은 감소하고, 오른쪽 절단점 t_2 가 증가할수록 평균은 증가한다는 것을 알 수 있다.



[그림 1] 양쪽방향으로 절단된 표준정규분포의 평균 [그림 2] 양쪽방향으로 절단된 표준정규분포의 분산

Z 의 분산을 얻기 위해 우선 $E(Z^2)$ 을 유도하면 다음과 같다.

(1) $t_1, t_2 \geq 0$ 인 경우

$$\begin{aligned} E(Z^2) &= c(t_1, t_2) \int_{t_1}^{t_2} z^2 e^{-\frac{1}{2}z^2} dz = \frac{c(t_1, t_2)}{2} \int_{t_1}^{t_2} z e^{-\frac{1}{2}z^2} 2z dz \\ &= \frac{c(t_1, t_2)}{2} \sqrt{2\pi} \int_{t_1^2}^{t_2^2} \frac{1}{2^{3/2} \Gamma(3/2)} u^{\frac{3}{2}-1} e^{-\frac{u}{2}} du \\ &= c(t_1, t_2) \sqrt{\frac{\pi}{2}} \{ C_3(t_2^2) - C_3(t_1^2) \}, \end{aligned}$$

여기서 $C_3(\cdot)$ 는 자유도가 3인 카이제곱 분포함수를 뜻한다.

(2) $t_1 < 0 \leq t_2$ 인 경우

$$\begin{aligned} E(Z^2) &= c(t_1, t_2) \left\{ \int_{t_1}^0 z^2 e^{-\frac{1}{2}z^2} dz + \int_0^{t_2} z^2 e^{-\frac{1}{2}z^2} dz \right\} \\ &= \frac{c(t_1, t_2)}{2} \left\{ \int_{t_1}^0 z e^{-\frac{1}{2}z^2} 2z dz + \int_0^{t_2} z e^{-\frac{1}{2}z^2} 2z dz \right\} \\ &= \frac{c(t_1, t_2)}{2} \sqrt{2\pi} \left\{ \int_0^{t_1^2} \frac{1}{2^{3/2}\Gamma(3/2)} u^{\frac{3}{2}-1} e^{-\frac{u}{2}} du + \int_0^{t_2^2} \frac{1}{2^{3/2}\Gamma(3/2)} u^{\frac{3}{2}-1} e^{-\frac{u}{2}} du \right\} \\ &= c(t_1, t_2) \sqrt{\frac{\pi}{2}} \{ C_3(t_1^2) + C_3(t_2^2) \}, \end{aligned}$$

(3) $t_1, t_2 < 0$ 인 경우

$$\begin{aligned} E(Z^2) &= c(t_1, t_2) \int_{t_1}^{t_2} z^2 e^{-\frac{1}{2}z^2} dz = c(t_1, t_2) \int_{-t_2}^{-t_1} z^2 e^{-\frac{1}{2}z^2} dz \\ &= c(t_1, t_2) \sqrt{\frac{\pi}{2}} \{ C_3(t_1^2) - C_3(t_2^2) \}. \end{aligned}$$

위의 식을 이용하면 Z 의 분산은 다음과 같다.

$$V(Z) = c(t_1, t_2) \left[\sqrt{\frac{\pi}{2}} C_3(t_1, t_2) - c(t_1, t_2) \left\{ e^{-\frac{1}{2}t_1^2} - e^{-\frac{1}{2}t_2^2} \right\}^2 \right], \tag{2}$$

여기서 $C_3(t_1, t_2)$ 는 다음과 같다.

$$\begin{aligned} C_3(t_1, t_2) &= C_3(t_2^2) - C_3(t_1^2), & t_1, t_2 \geq 0 \text{인 경우,} \\ C_3(t_1, t_2) &= C_3(t_1^2) + C_3(t_2^2), & t_1 < 0 \leq t_2 \text{인 경우,} \\ C_3(t_1, t_2) &= C_3(t_1^2) - C_3(t_2^2), & t_1, t_2 < 0 \text{인 경우.} \end{aligned}$$

(2)식에서 정의된 양쪽방향으로 절단된 표준정규분포의 분산을 절단점 t_1 과 t_2 ($t_1 < t_2$)에 대하여 그래프로 표현한 그림이 [그림 2]이다. [그림 2]를 살펴보면, 왼쪽 절단점 t_1 이 감소할수록 분산은 증가하고, 그리고 오른쪽 절단점 t_2 가 증가할수록 분산이 증가한다는 것을 파악할 수 있다. 따라서 t_1 과 t_2 의 거리가 멀수록 분산이 증가한다는 것을 발견하였다.

실제로 절단된 분포의 다양한 적용에 있어서 표준정규분포보다는 오히려 비표준 정규분포의 경우가 많다. 그러므로 이런 경우에는 비표준 정규분포를 표준화시켜주어야 한다. 만약 X 가 a 의 왼쪽과 b 의 오른쪽이 절단된, 분포가 $N(\mu, \sigma^2)$ 인 확률변수라면 Z 는 $t_1 = (a - \mu)/\sigma$ 의 왼쪽과 $t_2 = (b - \mu)/\sigma$ 의 오른쪽이 절단된 표준정규확률변수이다. 그러므로 비표준 정규분포의 평균과 분산을 표준정규분포의 평균과 분산으로 표현하면 다음과 같다.

$$E(X) = \sigma E(Z) + \mu \tag{3}$$

$$V(X) = \sigma^2 V(Z) \tag{4}$$

여기서 $E(Z)$ 와 $V(Z)$ 는 각각 (1)식과 (2)식에서 주어졌다.

3. 절단분포의 평균과 분산의 추정

Barr와 Sherrill(1999)은 μ 와 σ 가 미지의 모수인 모집단 $N(\mu, \sigma^2)$ 로부터 확률표본이 주어졌을 때, 한쪽 방향으로 절단된 분포의 평균과 분산을 추정하는 방법으로 두 가지를 제시했는데 한 가지는 최대가능도 추정량의 불변성에 의해 절단된 평균과 분산의 추정량이 식(3)과 (4)에서 μ 와 σ^2 에 대응하는 각각의 최대가능도 추정량 \bar{X} 와 $(n-1)S^2/n$ 으로 대체되어 얻어질 수 있다.(이 방법으로 얻어진 추정량을 '전체표본 추정량'이라 하고 이 방법을 '전체표본 추정방법'이라 부른다.) 다른 방법은 절단하고자 하는 영역에 속하는 표본 관측값을 제거하여 절단된 모집단으로부터 추출된 확률표본으로 간주한다. 그리고 이 절단표본자료로부터 직접 표본평균과 표본분산을 계산하여 추정하는 것이다.(이 방법으로 얻어진 추정량을 '절단표본 추정량'이라 하고 이 방법을 '절단표본 추정방법'이라고 부른다.)

본 논문에서는 양쪽방향으로 절단된 분포의 평균과 분산을 추정하기 위하여 시뮬레이션을 이용하여 두 가지 추정방법들을 비교하였고, 그 결과들을 [표 1]부터 [표 4]까지에 보여주었다. 우선 표를 살펴보면 표본크기 n 의 확률표본을 표준정규분포에서 10,000개 생성하고, 절단 영역([표 1]과 [표 3]의 절단 영역은 $(-\infty, t]$, [표 2]와 [표 4]의 절단 영역은 $(-\infty, t_1] \cup [t_2, \infty)$ 이다.)에 대해 절단된 정규분포의 평균과 분산을 추정하였다. 절단의 정도가 크고 표본크기가 작은 경우 절단 후에 자료가 하나만 남거나 하나도 남지 않게 되는 경우가 발생하는데 이런 경우는 분산을 계산할 수 없게 된다.

따라서 절단된 후에 한 개 이하의 값들이 남아있을 경우, 전체표본 추정량으로 그 값을 대체하는 방법('대체방법'이라 하고 [표 1]과 [표 2]에 나타나 있다.)과 자료의 손실을 가져오기는 하지만 순수한 추정량으로 두 추정방법을 비교하기 위하여 그에 해당되는 자료를 제거하는 방법('제거방법'이라 하고 [표 3]과 [표 4]에 나타나 있다.)을 사용하였다. 절단된 후에 표본크기가 하나 이하인 경우에 전체표본 추정값으로 대체하는데 이런 경우가 발생할 비율을 '대체율'이라 정의하고 표의 맨 오른쪽 열에 나타나 있다.

Barr와 Sherrill(1999)에서는 왼쪽 절단의 경우에 대하여 평균과 분산을 추정할 때 절단표본 추정방법이 전체표본 추정방법보다 효율적임을 주장하였다. 또한 각 추정량들의 값이 왜 다른지에 대한 설명이 충분하지 않았다. 본 연구에서는 양쪽 절단의 경우와 절단된 자료가 한개 이하인 경우, 이에 해당되는 자료를 제거하는 방법을 추가하였다. 또한 평균의 경우, 절단의 정도뿐만 아니라 절단 후의 분포가 대칭인 경우도 고려하였으며 평균과 분산을 추정할 때 어떤 추정방법이 더 효율적인지에 대한 견해도 조금 다르다.

[표 1]과 [표 2]는 최대가능도 추정량으로 대체되어 얻어진 결과이다. 표의 오른쪽 열의 작은 대체율을 가진 경우 이 추정량들은 본질적으로 순수한 절단표본 추정량이며, 특히 이에 상응하는 전체표본 추정량과 비교하는 것에 관심이 있다. 그래서 대체율이 0보다 큰 경우에 대해서 자료의 손실을 가져오지만 순수한 추정량으로 비교하기 위하여 해당 자료를 제거한 결과들을 [표 3]과 [표 4]에서 보여주었다.

우선 평균추정량들에 대한 편의를 [표 1]과 [표 2]에서 살펴보면 대체율이 큰 경우를 제외하고는 절단표본 평균추정량의 편의는 거의 0에 가까운데 비해 전체표본 평균추정량의 편의는 절단 후의 분포가 비대칭이 됨에 따라, 표본크기 n 이 작을수록 점점 커지는 것을 알 수 있다. 이것은 편의가 다음과 같이 정의되고

$$b(\hat{\mu}) = E\left[c(t_1, t_2)(e^{-\frac{1}{2}t_1^2} - e^{-\frac{1}{2}t_2^2})(\tilde{\sigma} - 1) + \bar{X} \right],$$

여기서 $\tilde{\sigma}$ 는 σ^2 의 최대가능도 추정량으로 $\tilde{\sigma} = (n-1)s^2/n$ 이다. 이때 $\tilde{\sigma}$ 은 σ^2 의 편의추정량이며 모 분산 1보다 작기 때문에 편의가 발생한다고 설명된다. 또한 분포가 비대칭이 됨에 따라 (1)식에서 정의된 $E(Z)$ 의 값이 점점 커지는 영향을 받기 때문이다.

예를 들어 절단 후의 분포가 대칭인 [표 2]의 구간 $(-2, 2)$ 인 경우와 $(-1, 1)$ 인 경우를 살펴보면 편의가 거의 0에 가까움을 알 수 있다. 반면에 절단표본 평균추정량의 편의는 우리가 알고있는 것처럼 절단분포에 대한 \bar{X}_r 가 불편추정량이므로 대체율이 큰 경우를 제외하고는 0에 가까운 값을 가지는 것이다.

대체율이 큰 경우 절단표본 평균추정량의 편의가 존재하는 것은 기대한 것처럼 절단표본 평균추정량이 전체표본 평균추정량으로 대체 되었기 때문인데, 대체율이 0보다 큰 경우에 대해서 그에 해당하는 표본을 제거하는 방법을 이용하여 얻어진 결과인 [표 3]과 [표 4]를 보면 전체표본 평균추정량의 편의는 별로 차이가 없지만 절단표본 평균추정량의 편의는 거의 0에 가까워졌음을 알 수 있다.

평균추정량들에 대한 평균제곱오차(MSE)를 살펴보면 특히 대체율이 작은 경우 전체표본 평균추정량의 평균제곱오차가 절단표본 평균추정량의 평균제곱오차보다 훨씬 크다는 것을 알 수 있는데, 이는 전체표본 평균추정량의 분산이 절단 후의 분포가 비대칭이 됨에 따라 급속하게 커지는 $E(Z)$ 의 값에 영향을 받기 때문이다.

절단 후의 분포가 대칭인 경우인 [표 2]의 구간 $(-2, 2)$ 인 경우와 $(-1, 1)$ 인 경우를 보면 두 구간의 평균제곱오차가 같다.

이와 같이 절단 후의 분포가 대칭인 경우는 절단표본 평균추정량의 평균제곱오차와 별로 차이가 없음을 알 수 있다. 평균제곱오차들의 값이 별로 차이가 나지 않는 경우는 대체율이 크다는 것을 알 수 있는데 이는 기대한 것처럼 절단표본 평균추정량이 전체표본 평균추정량으로 대체 되었기 때문이다. [표 3]과 [표 4]의 결과를 살펴보면 전체표본 평균추정량의 평균제곱오차는 별로 차이가 없지만 절단표본 평균추정량의 평균제곱오차는 현저하게 작아졌음을 알 수 있다.

참고로 추정량의 분산과 편의의 제곱을 더한 형태인 평균제곱오차들을 설명하는데 있어서 추정량의 분산으로만 설명한 것은 전체표본 평균추정량의 경우 편의가 있긴 하지만 제곱한 값은 아주 작은 값이 되기 때문에 평균제곱오차에 많은 영향은 주지 못하기 때문이다.

분산추정량들에 대한 편의를 살펴보면 대체율이 큰 경우를 제외하고는 절단표본 분산추정량의 편의는 거의 0에 가깝는데 비해 전체표본 분산추정량의 편의는 절단의 정도가 커짐에 따라(절단 영역이 작아짐) 점점 작아지기는 하지만 절단표본 분산들의 편의보다 0에서 멀다는 것을 알 수 있다. 이런 현상은 다음과 같이 정의된 분산의 편의에서 살펴볼 수 있다.

$$b(\hat{\sigma}^2) = E[(\tilde{\sigma}^2 - 1)V(Z)].$$

평균의 편의에서와 마찬가지로 최대가능도 추정량 $\hat{\mu}$ 은 편의추정량이기 때문에 편의가 표본크기 n 이 작은 경우에 커진다. 그러나 절단의 정도가 커질수록 편의가 감소하는데 이것은 절단의 정도가 커지면 (2)식에서 정의된 $V(Z)$ 의 값이 점점 작아지기 때문이다.

반면에 절단표본 분산들의 편의는 우리가 알고 있는 것처럼 절단분포에 대한 S_1^2 이 불편 추정량이므로 대체율이 큰 경우를 제외하고는 0에 가까운 값을 가지는 것이다. 대체율이 큰 경우 절단표본 분산추정량의 편의가 존재하는 것은 기대한 것처럼 절단표본 분산추정량이 전체표본 분산추정량으로 대체되었기 때문인데, [표 3]과 [표 4]를 보면 전체표본 분산추정량의 편의는 별로 차이가 없지만 절단표본 분산추정량의 편의는 0에 가까워졌음을 알 수 있다.

분산추정량들에 대한 평균제곱오차를 살펴보면 절단의 정도가 커짐에 따라 전체표본 분산추정량의 평균제곱오차와 절단표본 분산추정량의 평균제곱오차가 차이가 많이 나는 것을 알 수 있는데 이는 전체표본 분산추정량의 분산이 절단의 정도가 커짐에 따라 급속하게 작아지는 $V(Z)$ 의 값에 영향을 받기 때문이다.

평균제곱오차들의 값이 많이 차이 나지 않는 경우는 대체율이 크다는 것을 알 수 있는데 이는 기대한 것처럼 절단표본 분산추정량이 전체표본 분산추정량으로 대체 되었기 때문이다. [표 3]과 [표 4]의 결과를 살펴보면 전체표본 분산추정량의 평균제곱오차는 별로 차이가 없지만 특히, 대체율이 큰 경우에는 절단표본 분산추정량의 평균제곱오차는 조금 더 커졌음을 발견할 수 있다. 이는 제거방법으로 인하여 표본의 크기가 작아지기 때문에 절단 분산추정량의 평균제곱오차가 커지게 되는 것으로 설명된다.

4. 결론

한쪽 또는 양쪽 절단 후의 분포가 비대칭인 경우 평균 추정의 경우 전체표본 평균추정량의 평균제곱오차가 절단표본 평균추정량보다 더 크다. 또한 평균제곱오차 값에 큰 영향은 미치지 못하지만 절단표본 평균추정량은 불편인데 비해 전체표본 평균추정량은 편의가 존재한다.

특히 양쪽 절단 후의 분포가 대칭인 경우, 평균제곱오차가 비대칭인 경우에 비해 약간 작기는 하지만 절단표본 평균추정량의 평균제곱오차에 비하면 여전히 큰 값이다.

그러므로 양쪽 방향으로 절단된 분포의 평균을 추정하는 경우에는 절단표본 추정방법이 훨씬 더 효율적임을 확인할 수 있다. 그러나 분산의 경우 1보다 작은 $V(Z)$ 의 영향으로 전체표본 분산추정량의 평균제곱오차가 절단표본 분산추정량의 평균제곱오차보다 작아진다. 전체표본 분산추정량의 편의는 평균의 경우와 마찬가지로 평균제곱오차에 크게 영향을 미치지 못하는 못한다.

따라서 양쪽 방향으로 절단된 분포의 분산을 추정하는 경우 특히 표본크기가 작은 경우에 대하여 전체표본에 기초한 최대가능도 추정량을 사용하여 분산을 추정하는 것이 바람직하다.

[표 1] 왼쪽 절단분포의 평균과 분산에 대한 추정량 : 대체방법

t	P[X > t]	n	평균				분산				대체율
			전체표본		절단표본		전체표본		절단표본		
			bias	MSE	bias	MSE	bias	MSE	bias	MSE	
-1	0.84	10	-0.03	0.10	-0.00	0.07	-0.06	0.07	0.00	0.11	0.000
		20	-0.02	0.05	-0.00	0.04	-0.03	0.04	-0.00	0.05	0.000
		36	-0.01	0.03	-0.00	0.02	-0.02	0.02	-0.00	0.03	0.000
		50	-0.00	0.02	0.00	0.01	-0.02	0.02	-0.00	0.02	0.000
		100	-0.00	0.01	0.00	0.01	-0.01	0.01	0.00	0.01	0.000
-0.5	0.69	10	-0.03	0.11	0.00	0.08	-0.05	0.05	-0.00	0.10	0.000
		20	-0.02	0.06	0.00	0.04	-0.02	0.02	-0.00	0.04	0.000
		36	-0.01	0.03	0.00	0.02	-0.01	0.01	0.00	0.02	0.000
		50	-0.01	0.02	0.00	0.01	-0.01	0.01	-0.00	0.02	0.000
		100	-0.01	0.01	-0.00	0.01	-0.01	0.00	-0.00	0.01	0.000
0	0.50	10	-0.06	0.14	-0.01	0.09	-0.04	0.02	-0.00	0.10	0.011
		20	-0.03	0.07	0.00	0.04	-0.02	0.01	0.00	0.05	0.001
		36	-0.02	0.04	0.00	0.02	-0.01	0.01	-0.00	0.02	0.000
		50	-0.00	0.03	0.00	0.01	-0.01	0.01	0.00	0.02	0.000
		100	-0.00	0.01	0.00	0.00	-0.00	0.00	0.00	0.01	0.000
.5	0.31	10	-0.09	0.18	-0.09	0.13	-0.03	0.01	-0.00	0.10	0.137
		20	-0.05	0.08	-0.01	0.05	-0.01	0.01	-0.00	0.05	0.006
		36	-0.02	0.05	0.00	0.03	-0.01	0.00	0.00	0.03	0.000
		50	-0.02	0.03	0.00	0.02	-0.01	0.00	0.00	0.02	0.000
		100	-0.01	0.02	0.00	0.01	-0.00	0.00	-0.00	0.01	0.000
1	0.16	10	-0.12	0.22	-0.20	0.19	-0.02	0.01	-0.01	0.05	0.511
		20	-0.05	0.11	-0.07	0.09	-0.01	0.00	0.00	0.06	0.146
		36	-0.04	0.06	0.00	0.04	-0.01	0.00	-0.00	0.04	0.015
		50	-0.03	0.04	0.00	0.03	-0.00	0.00	0.00	0.03	0.002
		100	-0.01	0.02	0.00	0.01	-0.00	0.00	-0.00	0.01	0.000
2	0.16	10	-0.18	0.41	-0.20	0.38	-0.01	0.00	-0.01	0.00	0.981
		20	-0.09	0.20	-0.13	0.17	-0.01	0.00	-0.01	0.00	0.919
		36	-0.05	0.11	-0.11	0.09	-0.00	0.00	-0.00	0.01	0.806
		50	-0.04	0.08	-0.09	0.07	-0.00	0.00	-0.00	0.01	0.686
		100	-0.01	0.04	-0.04	0.05	-0.00	0.00	0.00	0.02	0.327

t = 절단점
 μ_t = 절단된 표준정규분포의 평균
 σ_t^2 = 절단된 표준정규분포의 분산

n = 표본크기
 bias = 추정량의 표본평균 - 모수값
 MSE = 추정량의 표본분산 + (bias)²

[표 2] 양쪽 절단분포의 평균과 분산에 대한 추정량 : 대체방법

t_1	t_2	$P[t_1 < X < t_2]$	n	평균				분산				대체율
				전체표본		절단표본		전체표본		절단표본		
				bias	MSE	bias	MSE	bias	MSE	bias	MSE	
-2	2	0.95	10	0.00	0.10	-0.00	0.08	-0.08	0.11	-0.00	0.10	0.000
		$\mu_t \approx 0.00$	20	0.00	0.05	0.00	0.04	-0.04	0.06	-0.00	0.05	0.000
		$\sigma_t^2 \approx 0.774$	36	-0.00	0.03	-0.00	0.02	-0.02	0.03	0.00	0.02	0.000
			50	-0.00	0.02	-0.00	0.02	-0.01	0.02	-0.00	0.02	0.000
			100	-0.00	0.01	0.00	0.01	-0.01	0.01	0.00	0.01	0.000
-1	2	0.82	10	-0.02	0.11	-0.00	0.07	-0.05	0.05	0.00	0.05	0.000
		$\mu_t \approx 0.230$	20	-0.01	0.05	-0.00	0.03	-0.03	0.03	0.00	0.02	0.000
		$\sigma_t^2 \approx 0.520$	36	-0.01	0.03	-0.00	0.02	-0.01	0.01	0.00	0.01	0.000
			50	-0.00	0.02	0.00	0.01	-0.01	0.01	-0.00	0.01	0.000
			100	-0.00	0.01	-0.00	0.01	-0.01	0.01	-0.00	0.00	0.000
-1	1	0.68	10	0.00	0.10	-0.00	0.05	-0.03	0.02	-0.00	0.02	0.001
		$\mu_t \approx 0.000$	20	0.00	0.05	0.00	0.02	-0.01	0.01	-0.00	0.01	0.000
		$\sigma_t^2 \approx 0.291$	36	-0.00	0.03	-0.00	0.01	-0.01	0.00	0.00	0.00	0.000
			50	-0.00	0.02	-0.00	0.01	-0.01	0.00	-0.00	0.00	0.000
			100	-0.00	0.01	-0.00	0.00	-0.00	0.00	-0.00	0.00	0.000
0	2	0.48	10	-0.06	0.13	-0.01	0.07	-0.03	0.01	-0.01	0.03	0.017
		$\mu_t \approx 0.723$	20	-0.03	0.06	0.00	0.03	-0.01	0.01	0.00	0.01	0.000
		$\sigma_t^2 \approx 0.251$	36	-0.02	0.04	0.00	0.01	-0.01	0.00	0.00	0.01	0.000
			50	-0.01	0.03	0.00	0.01	-0.00	0.00	-0.00	0.00	0.000
			100	-0.01	0.01	-0.00	0.00	-0.00	0.00	-0.00	0.00	0.000
0	1	0.34	10	-0.06	0.13	-0.01	0.07	-0.03	0.01	0.00	0.03	0.018
		$\mu_t \approx 0.466$	20	-0.02	0.06	-0.00	0.01	-0.00	0.00	-0.00	0.00	0.002
		$\sigma_t^2 \approx 0.080$	36	-0.01	0.03	-0.00	0.01	-0.00	0.00	0.00	0.00	0.000
			50	-0.01	0.02	0.00	0.00	-0.00	0.00	0.00	0.00	0.000
			100	-0.00	0.01	0.00	0.00	-0.00	0.00	0.00	0.00	0.000
1	2	0.14	10	-0.11	0.20	-0.17	0.17	-0.01	0.00	-0.01	0.00	0.600
		$\mu_t \approx 1.383$	20	-0.05	0.10	-0.06	0.06	-0.00	0.00	-0.00	0.00	0.221
		$\sigma_t^2 \approx 0.073$	36	-0.02	0.05	-0.01	0.02	-0.00	0.00	0.00	0.00	0.032
			50	-0.02	0.04	-0.00	0.01	-0.00	0.00	-0.00	0.00	0.005
			100	-0.01	0.02	0.00	0.01	-0.00	0.00	0.00	0.00	0.000

t = 절단점
 μ_t = 절단된 표준정규분포의 평균
 σ_t^2 = 절단된 표준정규분포의 분산

n = 표본크기
 bias = 추정량의 표본평균 - 모수값
 MSE = 추정량의 표본분산 + (bias)²

[표 3] 왼쪽 절단분포의 평균과 분산에 대한 추정량 : 제거방법

t	P[X > t]	n	평균				분산				대체율
			전체표본 추정법		절단표본 추정법		전체표본 추정법		절단표본 추정법		
			bias	MSE	bias	MSE	bias	MSE	bias	MSE	
0	0.50	10	-0.06	0.13	0.00	0.08	-0.04	0.02	-0.00	0.11	0.011
		20	-0.03	0.07	-0.00	0.04	-0.02	0.01	0.00	0.05	0.001
.5	0.31	10	-0.09	0.17	-0.00	0.09	-0.03	0.01	0.00	0.11	0.137
		20	-0.04	0.08	-0.00	0.05	-0.01	0.01	0.00	0.05	0.006
1	0.16	10	-0.12	0.23	0.00	0.08	-0.02	0.01	-0.00	0.09	0.511
		20	-0.06	0.11	0.00	0.06	-0.01	0.00	-0.00	0.06	0.146
		36	-0.03	0.06	0.00	0.04	-0.01	0.00	-0.00	0.04	0.015
		50	-0.02	0.04	0.00	0.03	-0.00	0.00	0.00	0.03	0.002
2	0.02	10	-0.18	0.41	-0.01	0.05	-0.01	0.00	0.00	0.06	0.981
		20	-0.09	0.19	0.02	0.06	-0.01	0.00	0.01	0.04	0.919
		36	-0.06	0.11	0.01	0.06	-0.00	0.00	-0.00	0.04	0.806
		50	-0.03	0.08	0.00	0.05	-0.00	0.00	-0.00	0.04	0.686
		100	-0.02	0.04	-0.00	0.04	-0.00	0.00	0.00	0.03	0.327

[표 4] 양쪽 절단분포의 평균과 분산에 대한 추정량 : 제거방법

t ₁	t ₂	P[t ₁ < X < t ₂]	n	평균				분산				대체율
				전체표본 추정법		절단표본 추정법		전체표본 추정법		절단표본 추정법		
				bias	MSE	bias	MSE	bias	MSE	bias	MSE	
-1	1	0.68	10	0.00	0.10	0.00	0.05	-0.03	0.02	-0.00	0.02	0.001
0	2	0.48	10	-0.05	0.13	0.00	0.06	-0.02	0.01	0.00	0.03	0.017
0	1	0.34	10	-0.03	0.11	-0.00	0.02	-0.01	0.00	-0.00	0.00	0.018
			20	-0.02	0.06	-0.00	0.01	-0.00	0.00	0.00	0.00	0.002
1	2	0.14	10	-0.10	0.20	-0.00	0.03	-0.01	0.00	-0.00	0.01	0.600
			20	-0.05	0.10	0.00	0.03	-0.00	0.00	0.00	0.00	0.221
			36	-0.03	0.06	0.00	0.02	-0.00	0.00	-0.00	0.00	0.032
			50	-0.02	0.04	0.00	0.01	-0.00	0.00	0.00	0.00	0.005

참고문헌

- [1] Barr, D. and Sherrill, E. (1999). Mean and Variance of truncated Normal Distribution, The American Statistician, 53, 4, 357-361.
- [2] Cohen, A. (1949). On Estimating the Mean and Standard Deviation of Truncated Normal Distributions, Journal of the American Statistical Association, 44, 518-525
- [3] Cohen, A. (1950). Estimating the Mean and Variance of Normal Populations from Singly Truncated and Doubly Truncated Samples, Annals of the Mathematical Statistics, 21, 557-569.
- [4] Cohen, A. (1959). Simplified Estimation for the Normal Distribution when Samples are Singly Censored or Truncated, Technometrics, 1, 217-237.
- [5] Cohen, A. (1961). Tables for Maximum Likelihood Estimation: Singly Truncated and Singly Censored Samples, Technometrics, 3, 433-438.
- [6] Cohen, A. (1991). Truncated and Censored Samples: Theory and Application, New York: Marcel Dekker.
- [7] Fisher, R. (1931). The Truncated Normal Distribution, British Association for the Advancement of Science, 5.
- [8] Gupta, A. (1952). Estimation of the Mean and Standard Deviation of the Normal Population from a Censored Sample, Biometrika, 39, 260-273.
- [9] Hald, A. (1952). Statistical Theory with Engineering Applications.
- [10] Halperin, M. (1952). Estimation in the Truncated Normal Distribution, Journal of the American Statistical Association, 47, 457-465.
- [12] Johnson, N., and Kotz, S. (1970). Continuous Univariate Distribution-1.
- [13] Keceioglu, D. (1991). Reliability Engineering Handbook(Vol. 1), Englewood Cliffs.

[2002년 1월 접수, 2002년 3월 채택]