Journal of the Korean Data & Information Science Society
2013, **24**(4), 825–833

http://dx.doi.org/10.7465/jkdi.2013.24.4.825
한국데이터정보과학회지

모의실험에 의한 온실가스 인벤토리 불확도 산정을 위한 지수분포 신뢰구간 추정방법[†]

이영섭[1] · 김희경[2] · 손덕규[3] · 이종식[4]

[123]동국대학교 통계학과 · [4]국립농업과학원

접수 2013년 5월 23일, 수정 2013년 7월 6일, 게재확정 2013년 7월 11일

요 약

온실가스 인벤토리 불확도 산정을 위해서는 인벤토리의 신뢰구간 추정이 필수적이다. 일반적으로 모수에 대한 신뢰구간 추정시에는 모집단이 정규분포를 따른다고 가정한다. 그러나 자료의 구조가 복잡해짐에 따라 정규분포가 아닌 비대칭형 자료, 즉 양의 왜도를 갖는 자료의 경우 기존의 정규분포를 가정한 신뢰구간 추정 방식은 적합하지 않다. 본 연구에서는 비대칭형 분포인 지수분포의 신뢰구간추정 방법으로 모수적인 방법과 비모수적인 방법에 대해 각각 비교분석하였다. 모의실험을 통한 신뢰구간 추정 결과를 바탕으로 범위확률, 신뢰구간 길이, 상대적 편의를 비교한 결과 모수적 방법 중에서 예상했던 대로 정확한 방법인 카이제곱방법이 신뢰계수와 유사한 범위확률을 보이고 상대적 편의도 작아 모수적 방법 중에서 신뢰구간 추정에 가장 적합한 것으로 나타났다. 마찬가지로 비모수적 방법 중에서는 표준화된 t-붓스트랩 방법이 가장 적합한 것으로 나타났다.

주요용어: 붓스트랩, 비대칭형 분포, 신뢰구간, 인벤토리 불확도, 지수분포.

1. 서 론

최근 환경문제로 인한 지구온난화가 국제적 이슈로 대두되고 있다. 기후변화 국제적 협의체인 IPCC (Intergovernmental Panel on Climate Change)는 2007년 4차 평가보고서에서 지난 100년간 (1906년~2006년) 지구 온도는 0.74℃ 상승하였으며, 특히 지난 50년간 온도 상승폭이 100년간에 비해 2배 가량 높다고 지적하였다. 지구온난화를 유발하는 가장 큰 원인중 하나로 온실가스를 들 수 있다 (Park 등, 2012). 유엔기후변화협약 (United Nations Framework Convention on Climate Change; UNFCCC)에서는 온실가스 감축의무 국가를 지정하고, 그 국가들로부터 온실가스 배출량과 흡수량의 인벤토리 (inventory)에 관한 보고를 요구하고 있다. 인벤토리에 관한 내용은 인벤토리의 양적평가와 그 인벤토리에서의 불확도 (uncertainty)를 포함한다. 온실가스 인벤토리에서의 불확도 산정은 통계적인 기법인 신뢰구간의 추정으로 가능하다 (Lee 등, 2012). IPCC 가이드라인 (IPCC, 2007)에 의하면 온실가스 배출 및 흡수의 불확도 산정에 95% 신뢰구간을 사용 할 것을 권장하고 있다. 온실가스 인벤토리 자료의 분포가 대칭형인 경우 기존의 정규분포를 가정한 신뢰구간 추정으로 불확도 산정이 가능하지만 비대칭형 자료, 즉 양의 왜도를 갖는 자료의 경우는 정규분포 가정의 신뢰구간 추정이 적합하지 않다.

---

† 본 논문은 농촌진흥청 공동연구사업 (과제번호: PJ00898605)의 지원에 의해 이루어진 것임.
1 교신저자: (100-715) 서울시 중구 필동로 1길 30, 동국대학교 통계학과, 교수. E-mail yung@dongguk.edu
2 (100-715) 서울시 중구 필동로 1길 30, 동국대학교 통계학과, 박사 후 연구원.
3 (100-715) 서울시 중구 필동로 1길 30, 동국대학교 통계학과, 석사과정.
4 (441-707) 경기도 수원시 권선구 수인로 126, 국립농업과학원기후변화생태과, 농업연구관.

온실가스 배출원별 배출 및 흡수량 자료는 배출원에 따라 비대칭형 자료 형태인 경우가 많다. 이러한 경우 몬테카를로 모의실험과 같은 비모수적 방법을 이용하여 불확도를 산정하였다. 또한 비대칭형 분포의 경우 대푯값으로 평균보다는 중위수를 이용하는 경우가 많지만, 불확도 평가 부분에 있어서 온실가스 배출량은 연간 평균값을 이용한다 (Eggleston과 Buendia, 2006). 따라서 본 연구에서는 양의 왜도를 갖는 자료의 경우 불확도 산정을 위해 모수적 방법으로 지수분포를 가정하여 평균에 대한 신뢰구간을 추정하고, 비모수적 방법으로 붓스트랩 방법을 적용하여 평균에 대한 신뢰구간을 추정하여 비교 분석하였다.

2. 지수분포 신뢰구간 추정 방법

지수분포는 비대칭형 분포로 확률변수 X의 값이 대부분 작은 값에 치우쳐있고, X의 값이 증가함에 따라 확률밀도가 0에 가까워지는 형태로 왜도가 매우 큰 분포 중에 하나이다. 지수분포의 확률밀도함수는 모수 $\theta > 0$에 대해 식 (2.1)과 같다.

$$f(x) = \frac{1}{\theta}e^{-x/\theta},\ 0 \le x < \infty \tag{2.1}$$

그리고 지수분포의 평균과 분산은 각각 식 (2.2)와 같다.

$$\begin{cases} EX = \theta \\ VarX = \theta^2 \end{cases} \tag{2.2}$$

2.1. 모수적 신뢰구간 추정 방법

지수분포의 평균 θ의 신뢰구간을 추정하기위한 방법으로는 최대우도 추정법 (maximum likelihood estimation)에 의한 전통적인 방법, Sprott방법, 카이제곱 방법 등 세 가지를 본 연구에서 고려하였다. 먼저, 최대우도 추정법을 이용한 전통적 방법에서는 θ의 MLE가 $\widehat{\theta} = \sum_{i=1}^{n} X_i/n = \overline{X}$이고, 여기서 n은 표본의 수를 나타낸다. n이 커지면 (대략 $n \ge 25$) $\widehat{\theta}$의 평균과 분산은 각각 점근적으로 식 (2.3)과 같게 된다 (Gross과 Clark, 1975).

$$E(\hat{\theta}) = \theta,\ Var\widehat{\theta} \doteq \frac{1}{-E\left(\frac{\partial^2}{\partial\theta^2}logL(\theta)\right)} = I\left(\theta\right)^{-1} = \frac{\theta^2}{n} \tag{2.3}$$

여기서 $I\left(\theta\right)^{-1}$는 피셔정보 (Fisher information)이며, $L(\cdot)$는 우도함수 (likelihood function)이다. 추정량 $\widehat{\theta}$은 근사적으로 평균이 θ이고 분산이 θ^2/n인 정규분포를 따르게 되며, 식 (2.4)와 같이 표현 할 수 있다. 따라서 모평균 θ에 대한 $(1-\alpha)\times 100\%$ 신뢰구간은 식 (2.5)와 같다.

$$\frac{\sqrt{n}(\widehat{\theta}-\theta)}{\widehat{\theta}} \sim N(0,1) \tag{2.4}$$

$$\theta \in \left\{\overline{X}\left(1-\frac{z_{1-\alpha/2}}{\sqrt{n}}\right),\ \overline{X}\left(1+\frac{z_{1-\alpha/2}}{\sqrt{n}}\right)\right\} \tag{2.5}$$

식 (2.5)에서 n이 매우 작으면 신뢰 하한 값이 음수를 가질 수도 있다. 이런 경우 신뢰 하한 값을 0으로 대체하거나 표본의 크기를 크게 해야 된다는 단점이 있다 (Lawless, 1982).

두 번째로 Sprott (1973)는 표본 수가 작은 경우 $\widehat{\theta}$보다 $(\widehat{\theta})^{-1/3}$이 정규분포에 훨씬 가까워진다는 사실에 착안해 신뢰구간 방법을 제안하였다. $(\widehat{\theta})^{-1/3}$의 분포는 식 (2.6)과 같고, 이를 표준화 하여 식

(2.7)과 같이 나타낼 수 있다.

$$(\widehat{\theta})^{-1/3} \sim N\left(\theta^{-1/3},\ \frac{\theta^{-1/3}}{3\sqrt{n}}\right) \tag{2.6}$$

$$\frac{(\widehat{\theta})^{-1/3} - \theta^{-1/3}}{\theta^{-1/3}/3\sqrt{n}} \sim N(0,1) \tag{2.7}$$

따라서 모평균 θ에 대한 $(1-\alpha)\times 100\%$ 신뢰구간은 식 (2.8)과 같이 나타낼 수 있다. Sprott 방법은 완전한 데이터뿐만 아니라 중도절단 데이터 (censored data)의 신뢰구간 추정도 가능하다 (Sundberg, 2001).

$$\theta \in \left(\frac{\overline{X}}{(1+z_{1-\alpha/2}/3\sqrt{n})^3},\ \frac{\overline{X}}{(1-z_{1-\alpha/2}/3\sqrt{n})^3}\right) \tag{2.8}$$

세 번째로 카이제곱 방법은 확률변수 $X_1, X_2, \cdots, X_n$의 평균이 θ인 지수분포를 따르면, $\sum_{i=1}^n X_i$는 모수 n과 θ를 갖는 감마분포를 따르게 된다. 감마분포와 카이제곱분포의 관계에 따라 식 (2.9)와 같이 표현 할 수 있다 (Epstein과 Sobel, 1953).

$$\frac{2}{\theta}\sum_{i=1}^{n} X_i \sim \Gamma\left(\frac{2n}{2}, 2\right) = \chi^2_{2n} \tag{2.9}$$

따라서 카이제곱분포를 이용하여 식 (2.10)과 같이 표현 할 수 있다.

$$p\left\{\chi^2_{\alpha/2,2n} < \frac{2}{\theta}\sum_{i=1}^{n} X_i < \chi^2_{1-\alpha/2,2n}\right\} = 1-\alpha \tag{2.10}$$

식 (2.10)을 이용하여 모평균 θ에 대한 $(1-\alpha)\times 100\%$ 신뢰구간을 구하면 식 (2.11)과 같다.

$$\theta \in \left\{\frac{2\bar{X}}{\chi^2_{1-\alpha/2,2n}},\ \frac{2\bar{X}}{\chi^2_{\alpha/2,2n}}\right\} \tag{2.11}$$

2.2. 비모수적 신뢰구간 추정 방법 : 붓스트랩 방법

앞에서 언급한 MLE를 이용한 방법, Sprott방법, 카이제곱방법은 지수분포를 가정한 모수적 방법이다. 이러한 특정 분포를 가정하지 않은 비모수적 방법으로는 붓스트랩 방법이 대표적이다. 붓스트랩 방법은 분포와 모수에 대한 가정 없이 주어진 표본으로부터 복원추출을 통해 모집단의 모수를 추정하는 장점을 가지고 있다 (Diciccio과 Efron, 1996; Sohn과 Shin, 2012). 붓스트랩 방법에는 붓스트랩 백분위수 방법, 일반적 붓스트랩 방법, 표준화된 t-붓스트랩 방법 등이 있다.

붓스트랩 백분위수 방법은 붓스트랩 표본의 평균값들의 백분위수를 이용하여 신뢰구간을 구하는 간단한 방법으로 그 절차는 다음과 같다 (Chung과 Han, 2013).

(단계1) 주어진 n개의 표본으로부터 n개의 표본을 복원추출 하여 평균을 구하고, 이를 $\overline{X}_b$라 하자. 이 과정을 B번 반복한다 $(b = 1, \cdots, B)$.

(단계2) 전체 B개의 붓스트랩 표본의 평균값들을 순서대로 정렬한다. 여기서 $\alpha/2 \times 100$백분위수와 $(1-\alpha/2)\times 100$백분위수가 모평균 θ에 대한 $(1-\alpha)\times 100\%$ 신뢰구간의 신뢰하한과 신뢰상한이 된다.

다음으로 일반적 붓스트랩 방법은 붓스트랩 평균과 붓스트랩 표준편차를 이용하여 신뢰구간을 추정하는 방법으로 다음과 같은 절차를 따른다 (Efron과 Tibshirani, 1993).

(단계1) 주어진 n개의 표본으로부터 n개의 표본을 복원추출 하여 평균을 구하고, 이를 $\overline{X}_b$라 하자. 이 과정을 B번 반복한다 ($b = 1, \cdots, B$).
(단계2) 붓스트랩 표본으로부터 전체 붓스트랩 평균 $\widehat{X}_B$와 표준편차 $\widehat{\sigma}_B$을 식 (2.12)와 같이 구한다.

$$\widehat{X}_B = \sum_{b=1}^{B} \overline{X}_b / B, \ \widehat{\sigma}_B = \sqrt{\frac{1}{B-1} \sum_{b=1}^{B} (\widehat{X}_B - \overline{X}_b)} \tag{2.12}$$

(단계3) $\widehat{X}_B$와 $\widehat{\sigma}_B$을 이용하여 식 (2.13)과 같이 정규분포를 따르는 추축량 (pivotal quantity)을 정의한다.

$$Z = \frac{\widehat{X}_B - \theta}{\widehat{\sigma}_B} \sim N(0,1) \tag{2.13}$$

(단계4) 식 (2.13)을 이용하여 모평균 θ에 대한 $(1-\alpha) \times 100\%$ 신뢰구간의 신뢰하한과 신뢰상한을 식 (2.14)와 같이 구할 수 있다.

$$\theta \in \left(\widehat{X}_B - z_{\alpha/2} \widehat{\sigma}_B, \ \widehat{X}_B + z_{\alpha/2} \widehat{\sigma}_B \right) \tag{2.14}$$

표준화된 t-붓스트랩 방법은 다음과 같은 절차를 걸쳐 구할 수 있다.

(단계1) 주어진 n개의 표본으로부터 n개의 표본을 복원추출 하여 평균을 구하고, 이를 $\overline{X}_b$라 하자. 이 과정을 B번 반복한다 ($b = 1, \cdots, B$).
(단계2) 주어진 n개의 표본으로부터 평균 $\overline{X}$와 표준편차 $\widehat{\sigma}$을 구하고, 붓스트랩 표본의 표준편차 $\widehat{\sigma_b^*}(b = 1, \cdots, B)$을 구한다. 이를 바탕으로 식 (2.15)와 같이 t_b^*값들을 구할 수 있다.

$$t_b^* = \frac{\overline{X_b} - \overline{X}}{\widehat{\sigma_b^*}}, \ b = 1, \cdots, B \tag{2.15}$$

(단계3) 단계2에서 구한 t_b^*값들을 $t_{(1)}^* \le t_{(2)}^* \le ... \le t_{(B)}^*$와 같이 순서대로 정렬한다. 여기서 $\alpha/2 \times 100$백분위수와 $(1-\alpha/2) \times 100$백분위수가 t_0^*과 t_1^*이라 하면, 모평균 θ에 대한 $(1-\alpha) \times 100\%$ 신뢰구간의 신뢰하한과 신뢰상한은 식 (2.16)과 같이 구할 수 있다.

$$\theta \in \left(\overline{X} - t_0^* \widehat{\sigma}, \ \overline{X} - t_1^* \widehat{\sigma} \right) \tag{2.16}$$

이러한 여러 가지 방법에 의한 모평균 θ의 신뢰구간 추정에 대해서 모의실험을 통하여 비교분석하고자 한다.

3. 모의실험

3.1. 모의실험설계

양의 왜도를 갖는 비대칭형 분포의 평균에 대해 신뢰구간을 추정하는데 있어 지수분포를 가정한 모수적 방법과 분포의 가정을 하지 않는 비모수적 방법을 적용한다. 모수적 방법으로는 MLE를 이용한 전통적 방법, Sprott 방법, 카이제곱 방법을 적용하고, 비모수적 방법으로는 붓스트랩 백분위수 방법, 일반적 붓스트랩 방법, 표준화된 t-붓스트랩 방법을 적용하여 모의실험을 하고자 한다.

표본의 크기는 소표본과 대표본인 경우의 결과를 비교하기 위하여 n =10, 50, 200, 500으로 하고, 비대칭형이며 양의 왜도를 갖는 지수분포에서 표본을 반복 추출하였다. 이때 지수분포의 모평균 θ를 2, 1, 2/3, 1/3, 1/5로 변화시켰으며, 붓스트랩 방법에 대해서는 B =1000번의 반복 추출을 하였다. 이러한 전체 과정을 5000번 반복 실시하였으며, 그 결과로 구해진 5000개의 신뢰구간들에 대해 범위확률

(coverage probability), 신뢰구간의 길이 (interval width), 상대적 편의 (relative bias)를 통해 결과를 비교 평가하였다. 범위확률은 모의실험 결과로 얻어진 5000개의 신뢰구간 중 모평균을 포함하고 있는 신뢰구간의 비율을 말하는 것으로 식 (3.1)과 같다 (Zhou, 1997).

$$\text{범위확률} = \frac{1}{5000}\sum_{j=1}^{5000} I(LC_j < \theta < UC_j) \tag{3.1}$$

식 (3.1)에서 $I(\cdot)$는 지시함수 (indicator function)이고, LC와 UC는 신뢰구간의 신뢰하한과 신뢰상한을 나타낸다. 계산된 범위확률 값이 신뢰구간의 신뢰계수에 가까울수록 적용된 추정방법이 적합하다고 할 수 있다. 예를 들면, 95%신뢰구간 추정에서는 신뢰계수가 0.95이고, 범위확률이 신뢰계수인 0.95에 가까울수록 적용된 추정방법이 적합하다고 할 수 있다.신뢰구간의 길이는 식 (3.2)와 같이 모의실험 결과로 얻어진 5000개의 신뢰구간들에 대한 신뢰구간 길이의 평균을 의미한다.

$$\text{신뢰구간의 길이} = \frac{1}{5000}\sum_{j=1}^{5000}(UC_j - LC_j) \tag{3.2}$$

동일한 범위확률을 가진다면 신뢰구간의 길이가 작을수록 적용된 추정방법이 적합하다고 할 수 있다.

상대적 편의는 추정된 신뢰구간 중 모수를 포함하지 않는 신뢰구간들의 편의정도를 표현해주는 척도이다. 만약 특정 추정방법을 사용하여 반복 실험한 결과로 얻어진 신뢰구간들 중 모수를 포함하지 않는 신뢰구간들 대부분이 θ보다 작은 값일 경우 그 추정방법은 음편의 (negative bias) 되었다고 할 수 있다. 반대로 모수를 포함하지 않는 신뢰구간들 중 대부분이 θ보다 큰 값일 경우 그 방법은 양편의 (positive bias) 되었다고 할 수 있다. 상대적 편의는 식 (3.3)과 같다.

$$\text{상대적 편의} = \frac{\left|\frac{1}{5000}\sum\limits_{j=1}^{5000} I(UC_j < \theta) - \frac{1}{5000}\sum\limits_{j=1}^{5000} I(LC_j > \theta)\right|}{\frac{1}{5000}\sum\limits_{j=1}^{5000} I(UC_j < \theta) + \frac{1}{5000}\sum\limits_{j=1}^{5000} I(LC_j > \theta)} \tag{3.3}$$

여기서 $1/5000\sum_{j=1}^{5000} I(UC_j < \theta)$는 5000번의 모의실험 중 모수 θ보다 작은 신뢰구간의 비율을 의미하고, $1/5000\sum_{j=1}^{5000} I(LC_j > \theta)$는 모수 θ보다 큰 신뢰구간의 비율을 의미한다. 상대적 편의는 0에서 1사이의 값을 가지며 1에 가까울수록 적용된 추정방법의 편의 정도가 큰 것을 의미하고, 0에 가까울수록 편의 되지 않음을 의미한다.

3.2. 모의실험결과

모수적방법인 MLE를 이용한 방법, Sprott방법, 카이제곱방법과 비모수적방법인 붓스트랩 백분위수 방법, 일반적 붓스트랩 방법, 표준화된 t-붓스트랩 방법을 적용하여 표본의 크기 $n = 10, 50, 200, 500$인 경우와 모수 $\theta = 2, 1, 2/3, 1/3, 1/5$인 경우 각각에 대해 95% 신뢰구간을 추정하는 모의실험 과정을 5000번 반복실시 하였다. 모수적 방법들을 적용하여 각각의 반복에서 구해진 5000개의 신뢰구간들에 대한 범위확률, 신뢰구간의 길이, 상대적 편의에 대한 결과가 Table 3.1에 나타나 있다. 반면 비모수적 방법들을 적용한 결과는 Table 3.2와 같다.

Table 3.1을 살펴보면 각각의 표본 크기와 θ값에 대해 예상했던 대로 카이제곱 방법에 의한 결과가 범위확률측면에서 가장 적합하고, 다음으로 Sprott방법이 적합한 것을 알 수 있다. 신뢰구간의 길이를 보면 MLE를 이용한 전통적 방법에 의한 결과가 가장 신뢰구간의 길이가 가장 짧은 것으로 나타났으나 각 방법에서 구해진 범위확률이 상이하므로 신뢰구간의 길이로 비교하는 것은 적합하지 않다. 상대적 편의를 기준으로 각 방법들을 비교해보면 대체적으로 카이제곱방법이 가장 적합하고, 다음으로 Sportt방법

이 적합한 것으로 나타났다. 반면 MLE를 이용한 전통적 방법은 다른 모수적 방법들에 비해서 상대적 편의가 큰 것으로 나타났다. 따라서 본 연구에서 적용한 모수적 방법들 중에서 카이제곱 방법을 적용한 신뢰구간 추정이 가장 적합하다고 할 수 있다.

비모수적 방법들을 적용한 결과로 Table 3.2를 살펴보면, 범위확률을 기준으로 하였을 경우 표본의 크기가 작을 때에는 표준화된 t-붓스트랩 방법이 대체로 적합하고 표본의 크기가 클 때에는 모든 방법들에 의한 범위확률이 신뢰계수 0.95에 가까운 값을 보이는 것을 알 수 있다. 신뢰구간의 길이를 살펴보면 소표본인 경우 표준화된 t-붓스트랩 방법에 의한 신뢰구간의 길이가 다른 방법들에 비해 길게 나타났으며, 대표본인 경우는 모든 방법들에 의한 결과가 대체로 비슷한 것으로 나타났다. 그러나 앞서 언급한 바와 같이 범위확률이 상이한 경우 신뢰구간의 길이를 비교하는 것은 적합하지 않다. 상대적 편의를 살펴보면 표준화된 t-붓스트랩 방법이 가장 적합하고, 다음으로 붓스트랩 백분위수 방법, 일반적인 붓스트랩 방법 순으로 나타났다.

Table 3.1 Coverage probability, interval width and relative bias of two-sided 95 percent confidence intervals for various parametric methods

n	θ	Method	Coverage probability	Interval width	Relative bias
		traditional method using MLE	0.9002	2.4809	1.0000
	2	Sprott method	0.9546	2.8738	0.2335
		chi-square method	0.9498	3.0005	0.0120
		traditional method using MLE	0.9046	1.2410	1.0000
	1	Sprott method	0.9516	1.4340	0.3058
		chi-square method	0.9470	1.4910	0.0340
		traditional method using MLE	0.9042	0.8264	0.9916
10	2/3	Sprott method	0.9464	0.9586	0.2090
		chi-square method	0.9458	1.0042	0.0554
		traditional method using MLE	0.9082	0.4151	1.0000
	1/3	Sprott method	0.9496	0.4758	0.4762
		chi-square method	0.9530	0.4997	0.0553
		traditional method using MLE	0.9070	0.2476	1.0000
	1/5	Sprott method	0.9538	0.2875	0.2208
		chi-square method	0.9534	0.2982	0.0043
		traditional method using MLE	0.9440	1.1103	0.7643
	2	Sprott method	0.9492	1.1448	0.1575
		chi-square method	0.9540	1.1510	0.0174
		traditional method using MLE	0.9416	0.5539	0.8014
	1	Sprott method	0.9472	0.5701	0.0909
		chi-square method	0.9520	0.5744	0.0417
		traditional method using MLE	0.9360	0.3685	0.8125
50	2/3	Sprott method	0.9480	0.3794	0.0615
		chi-square method	0.9458	0.3843	0.0185
		traditional method using MLE	0.9416	0.1854	0.7123
	1/3	Sprott method	0.9484	0.1901	0.2093
		chi-square method	0.9444	0.1918	0.0863
		traditional method using MLE	0.9332	0.1106	0.7545
	1/5	Sprott method	0.9502	0.1146	0.1406
		chi-square method	0.9446	0.1151	0.0469
		traditional method using MLE	0.9530	0.5544	0.4723
	2	Sprott method	0.9544	0.5594	0.0965
		chi-square method	0.9512	0.5590	0.0328
		traditional method using MLE	0.9432	0.2770	0.4859
	1	Sprott method	0.9456	0.2792	0.1103
		chi-square method	0.9542	0.2797	0.0480
		traditional method using MLE	0.9536	0.1849	0.5345
200	2/3	Sprott method	0.9534	0.1861	0.0472
		chi-square method	0.9536	0.1864	0.0690
		traditional method using MLE	0.9452	0.0925	0.4453
	1/3	Sprott method	0.9480	0.0931	0.0538
		chi-square method	0.9506	0.0934	0.0931
		traditional method using MLE	0.9546	0.0555	0.4273
	1/5	Sprott method	0.9496	0.0558	0.1667
		chi-square method	0.9522	0.0559	0.0711
		traditional method using MLE	0.9500	0.3503	0.3200
	2	Sprott method	0.9494	0.3521	0.0593
		chi-square method	0.9522	0.3514	0.0795
		traditional method using MLE	0.9518	0.1752	0.2448
	1	Sprott method	0.9498	0.1757	0.1394
		chi-square method	0.9514	0.1759	0.0453
		traditional method using MLE	0.9528	0.1168	0.2203
500	2/3	Sprott method	0.9492	0.1173	0.0157
		chi-square method	0.9480	0.1173	0.0154
		traditional method using MLE	0.9518	0.0584	0.4274
	1/3	Sprott method	0.9528	0.0586	0.0085
		chi-square method	0.9532	0.0586	0.0085
		traditional method using MLE	0.9476	0.0351	0.3053
	1/5	Sprott method	0.9554	0.0352	0.0045
		chi-square method	0.9486	0.0352	0.0973

따라서 본 연구에서 적용한 비모수적 방법들 중에서 표준화된 t-붓스트랩 방법을 적용한 신뢰구간 추정이 가장 적합하다고 할 수 있다.

Table 3.2 Coverage probability, interval width and relative bias of two-sided 95 percent confidence intervals for various nonparametric methods

n	θ	Method	Coverage probability	Interval width	Relative bias
		bootstrap percentile method	0.8624	2.1161	0.7180
	2	common bootstrap method	0.8576	2.1558	0.8230
		standardized t-bootstrap method	0.9420	3.6929	0.6966
		bootstrap percentile method	0.8670	1.0707	0.7173
	1	common bootstrap method	0.8482	1.0733	0.7945
		standardized t-bootstrap method	0.9360	1.8707	0.6313
		bootstrap percentile method	0.8648	0.7087	0.7485
10	2/3	common bootstrap method	0.8600	0.7249	0.8343
		standardized t-bootstrap method	0.9390	1.2173	0.7705
		bootstrap percentile method	0.8622	0.3530	0.7765
	1/3	common bootstrap method	0.8606	0.3615	0.8307
		standardized t-bootstrap method	0.9414	0.6171	0.6997
		bootstrap percentile method	0.8572	0.2135	0.7479
	1/5	common bootstrap method	0.8590	0.2165	0.8241
		standardized t-bootstrap method	0.9376	0.3665	0.7308
		bootstrap percentile method	0.9214	1.0664	0.5420
	2	common bootstrap method	0.9282	1.0778	0.6657
		standardized t-bootstrap method	0.9442	1.1899	0.1685
		bootstrap percentile method	0.9288	0.5345	0.5843
	1	common bootstrap method	0.9250	0.5376	0.6640
		standardized t-bootstrap method	0.9460	0.5931	0.1556
		bootstrap percentile method	0.9262	0.3562	0.4797
50	2/3	common bootstrap method	0.9240	0.3598	0.6421
		standardized t-bootstrap method	0.9464	0.3966	0.1343
		bootstrap percentile method	0.9216	0.1785	0.4184
	1/3	common bootstrap method	0.9220	0.1792	0.6769
		standardized t-bootstrap method	0.9456	0.1977	0.2647
		bootstrap percentile method	0.9300	0.1072	0.5200
	1/5	common bootstrap method	0.9188	0.1076	0.6453
		standardized t-bootstrap method	0.9420	0.1182	0.1517
		bootstrap percentile method	0.9450	0.5476	0.3309
	2	common bootstrap method	0.9456	0.5510	0.4265
		standardized t-bootstrap method	0.9498	0.5628	0.0040
		bootstrap percentile method	0.9420	0.2733	0.3241
	1	common bootstrap method	0.9460	0.2750	0.4222
		standardized t-bootstrap method	0.9446	0.2803	0.1625
		bootstrap percentile method	0.9440	0.1828	0.3357
200	2/3	common bootstrap method	0.9464	0.1833	0.4179
		standardized t-bootstrap method	0.9516	0.1868	0.0579
		bootstrap percentile method	0.9432	0.0914	0.2183
	1/3	common bootstrap method	0.9448	0.0918	0.4203
		standardized t-bootstrap method	0.9444	0.0936	0.0072
		bootstrap percentile method	0.9356	0.0547	0.2236
	1/5	common bootstrap method	0.9426	0.0551	0.3589
		standardized t-bootstrap method	0.9542	0.0563	0.1092
		bootstrap percentile method	0.9400	0.3485	0.1733
	2	common bootstrap method	0.9492	0.3493	0.2756
		standardized t-bootstrap method	0.9450	0.3513	0.0473
		bootstrap percentile method	0.9438	0.1743	0.2456
	1	common bootstrap method	0.9502	0.1750	0.2289
		standardized t-bootstrap method	0.9518	0.1758	0.1369
		bootstrap percentile method	0.9450	0.1158	0.1927
500	2/3	common bootstrap method	0.9492	0.1166	0.2598
		standardized t-bootstrap method	0.9432	0.1172	0.0352
		bootstrap percentile method	0.9432	0.0579	0.1901
	1/3	common bootstrap method	0.9438	0.0583	0.3381
		standardized t-bootstrap method	0.9494	0.0585	0.0356
		bootstrap percentile method	0.9484	0.0348	0.1163
	1/5	common bootstrap method	0.9450	0.0349	0.3018
		standardized t-bootstrap method	0.9496	0.0351	0.0635

4. 결론 및 향후과제

양의 왜도를 갖는 자료의 경우 불확도 산정을 위해 지수분포를 가정하고 모수적인 방법인 MLE를 이용한 방법, Sprott방법, 카이제곱방법을 이용하여 평균에 대한 신뢰구간을 추정 하였다. 그리고 비모수적 방법인 붓스트랩 백분위수 방법, 일반적 붓스트랩 방법, 표준화된 t-붓스트랩 방법을 이용하여 평균에 대한 신뢰구간을 추정하였다. 지수분포를 가정했으므로 모수적 방법, 특히 카이제곱방법의 결과가 신뢰구간 추정에 적합하다는 건 당연한 사실이다. 그럼에도 불구하고 본 연구에서는 표본수가 적을 때

나 많을 때 그리고 모평균 θ값을 변화시켜 신뢰구간이 어떻게 변하고 있으며, 어느 방법이 더 좋은가를 비교분석하였다. 실제 자료를 분석하는 것은 많은 제약이 따르므로 비대칭인 지수분포를 가정하고 확률변수를 생성한 것이다. 그 결과 온실가스 인벤토리 구축에 있어서 온실가스 배출량이 극심한 비대칭이고 지수분포의 형태를 띤다고 할 경우, 모수적 방법 중에서는 예상했던 대로 정확한방법인 카이제곱방법이 가장 적합한 것으로 나타났으며, 비모수적 방법 중 표준화된 t-붓스트랩 방법이 가장 적합한 것으로 나타났다.

비대칭형분포의 평균에 대한 신뢰구간을 추정하는데 있어 본 연구에서 적용한 붓스트랩 방법 외에 t-pivot방법, BCA (bias-corrected and accelerated)방법 등과 같이 더 다양한 붓스트랩 방법들을 이용하여 신뢰구간 추정·비교 할 수 있을 것이다. 모수적 신뢰구간을 추정 시 지수분포를 가정하였는데, 지수분포뿐만 아니라 와이블분포, 감마분포, 파레토분포등과 같이 다른 비대칭형 분포를 가정한 경우에 대해서도 신뢰구간 추정 방법을 비교 분석하고자 하며 향후 정확한 실증분석을 위해 실제 자료를 적용시켜 모수적 또는 비모수적 붓스트랩 방법들을 비교 연구하겠다.

References

Chung, H. C. and Han, C. P. (2013). Conditional bootstrap confidence intervals for classification error rate when a block of observations is missing. *Journal of the Korean Data & Information Science Society*, **24**, 189-200.

Diciccio, T. J. and Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, **11**, 189-228.

Efron, B. and Tibshirani, R. (1993). *An instruction to the bootstrap*, Chapman & Hall/CRC, New York.

Eggleston, S. and Buendia, L. (2006). *2006 IPCC guidelines for national greenhouse gas inventories*, Institute for Global Environmental Strategies, Kamiyamaguchi, Hayama, Kanagawa, Japan.

Epstein, B. and Sobel, M. (1953). Life testing. *Journal of the American Statistical Association*, **48**, 486-502.

Gross, A. J. and Clark, V. A. (1975). *Survival distributions: Reliability applications in the biomedical sciences*, John Wiley & Sons, New York.

IPCC (2007). *IPCC guidelines for national greenhouse gas inventories*, The Intergovernmental Panel on Climate Change, Geneva, Switzerland.

Lawless, J. F. (1982). *Statistical models and methods for lifetime data*, Wiley, New York.

Lee, Y. S., Son, D. K. and Jung, S. Y. (2012). Estimation of confidence interval of exponential distribution for the greenhouse gas inventory uncertainty. *Proceedings of the Autumn Conference of Korean Data & Information Science Society*, 27-28.

Park C. G., Soh, J. and Lee, Y. S. (2012). Estimation methods of fuel consumption using distance traveled: Focused on Monte Carlo method. *Journal of the Korean Data & Information Science Society*, **23**, 247-256.

Sohn, K. C. and Shin, I. H. (2012). Check for regression coefficient using jackknife and bootstrap methods in clinical data. *Journal of the Korean Data & Information Science Society*, **23**, 643-648.

Sprott, D. A. (1973). Normal likelihoods and their relation to large sample theory of estimation. *Biometrika*, **60**, 457-465.

Sundberg, R. (2001). Comparison of confidence procedures for type I censored exponential lifetimes. *Lifetime Data Analysis*, **7**, 393-413.

Zhou, X. H. (1997). Confidence intervals for the log-normal mean. *Statistic in Medicine*, **16**, 783-790.

Journal of the Korean Data & Information Science Society
2013, **24**(4), 825–833

http://dx.doi.org/10.7465/jkdi.2013.24.4.825
한국데이터정보과학회지

Estimation of confidence interval in exponential distribution for the greenhouse gas inventory uncertainty by the simulation study†

Yung-Seop Lee[1] · Hee-Kyung Kim[2] · Duck Kyu Son[3] · Jong-Sik Lee[4]

[123]Department of Statistics, Dongguk University
[4]Division of Climate Change Agroecology National Academy of Agricultural Science

Received 23 May 2013, revised 6 July 2013, accepted 11 July 2013

Abstract

An estimation of confidence intervals is essential to calculate uncertainty for greenhouse gases inventory. It is generally assumed that the population has a normal distribution for the confidence interval of parameters. However, in case data distribution is asymmetric, like nonnormal distribution or positively skewness distribution, the traditional estimation method of confidence intervals is not adequate. This study compares two estimation methods of confidence interval; parametric and non-parametric method for exponential distribution as an asymmetric distribution. In simulation study, coverage probability, confidence interval length, and relative bias for the evaluation of the computed confidence intervals. As a result, the chi-square method and the standardized t-bootstrap method are better methods in parametric methods and non-parametric methods respectively.

Keywords: Asymmetric distribution, bootstrap, confidence interval, exponential distribution, inventory uncertainty.

---

† This work was carried out with the support of "Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ00898605)" Rural Development Administration, Republic of Korea.

[1] Corresponding author: Professor, Department of Statistics, Dongguk University- Seoul, Seoul 100-715, Korea. Email: yung@dongguk.edu

[2] Post doctor, Department of Statistics, Dongguk University, Seoul 100-715, Korea.

[3] Graduate student, Department of Statistics, Dongguk University, Seoul 100-715, Korea.

[4] Senior researcher, Division of Climate Change Agroecology, National Academy of Agricultural Science, Suwon 441-707, Korea.