Journal of the Korean
Data & Information Science Society
2006, Vol. 17, No. 2, pp. 487 ~ 492

Jackknife Estimators in the Left Truncated Exponential Model

Kil-Ho Cho[1] · Jang Sik Cho[2]

Abstract

Jackknife estimators for parameters in the left truncated exponential model are presented. And we show that the generalized jackknife estimators are more efficient than others in terms of the bias and the mean squared error.

Keywords : 고차일반화잭나이프추정량, 일반화잭나이프추정량, 잭나이프추정량, 최우추정량

1. 머리말

통계학에서 사용되는 통계모형은 한계가 없는 확률변수에 대한 모형들이 많다. 그러나 많은 현실 상황에서는 자료 값들이 어떤 값 이상으로 관찰되는 경우가 흔하다. 이럴 경우 관찰된 자료에 대해서 왼쪽 절단된 분포(left truncated distribution)를 적용시키는 것이 더 타당 할 것이다.

일반적으로 절단된 분포를 적용시키는 경우에 절단점에 대한 최우추정량은 편의추정량인 경우가 많다. 이에 편의를 줄일 수 있는 잭나이프 추정법을 사용한다면 편의가 줄어든 더 좋은 추정량을 얻을 수 있다.

잭나이프 추정량은 Quenoulli(1956)가 추정량의 편의를 줄이기 위한 목적으로 제안하였으며, Tukey(1958)가 근사적인 신뢰구간을 추정하기 위하여 일반화된 방법으로 발전시켰다.

점추정에 관한 잭나이프 방법은 Chakrabraty와 Rao(1968), Rao와 Rao(1970), 조길호 등(1998), Rao와 Shao(1999), 강신수(2005) 등이 연구했으며, 구간추정에 관한 잭나이프 방법은 Miller(1974), Efron과 Stein(1981) 등이 적용했고, 선형모형에서의 잭나이프 방법은 Hinkley(1977), Simonoff와 Tsai(1986) 등이 연구하였다.

---

1) 대구광역시 북구 산격동 1370번지 경북대학교 통계학과 교수
E-mail : khcho@knu.ac.kr
2) 부산광역시 경성대학교 정보통계학과 부교수
E-Mail : jscho@ks.ac.kr

본 연구에서는 왼쪽 절단된 지수모형에서 편의를 줄이기 위한 잭나이프 추정량들을 Robson과 Whitlock(1964)이 제안한 이론을 이용하여 구하고, 그 추정량들의 편의와 평균제곱오차의 측면에서 기존의 추정량과 서로 비교 · 분석하고자 한다.

2. 최우추정량과 잭나이프 추정량

확률변수 $X_i(i=1, 2, \cdots, n)$들이 다음과 같은 확률밀도함수를 갖는 지수모형의 확률표본이라 하자.

$$f(x)=\begin{cases}\frac{1}{\sigma}\exp[-\frac{(x-\theta)}{\sigma}], & x>\theta \\ 0, & \text{그밖에}\end{cases} \tag{1}$$

모수 θ와 σ의 최우추정량은

$$\hat{\theta}_1=\hat{\theta}=min(X_{1,}X_2, \cdots, X_n)=X_{(1)}, \tag{2}$$

$$\hat{\sigma}_1=\hat{\sigma}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\hat{\theta})=\overline{X}-X_{(1)} \tag{3}$$

이다. 이 두 추정량들은 편의추정량이다. 그러므로 편의를 줄일 가능성이 있는 잭나이프 방법을 사용할 만하다.

$X_1, X_2, \cdots, X_n$을 확률밀도함수 식 (1)을 가지는 확률표본이라고 하자.

여기서 X_i를 제외한 나머지 표본에서 정의된 추정량 $\hat{\theta}^{-i}$는

$$\hat{\theta}^{-i}=\begin{cases}X_{(1)}, & \text{if } X_i\neq X_{(1)} \\ X_{(2)}, & \text{if } X_i=X_{(1)}\end{cases}$$

이므로 n개의 가능한 $\hat{\theta}^{-i}$들의 평균은

$$\hat{\theta}_2=\overline{\hat{\theta}^{-i}}=\frac{1}{n}\sum_{i=1}^{n}\hat{\theta}^{-i}=\frac{1}{n}[(n-1)X_{(1)}+X_{(2)}] \tag{4}$$

이다. 그리고 X_i와 X_j를 제외한 나머지 표본에서 정의된 추정량 $\hat{\theta}^{-i,j}$는

$$\hat{\theta}^{-i,j}=\begin{cases}X_{(1)}, & \text{if } X_i\neq X_{(1)} \text{ and } X_j\neq X_{(1)} \\ X_{(2)}, & \text{if } X_i=X_{(1)} \text{ and } X_j\neq X_{(1)}, X_{(2)} \\ X_{(3)}, & \text{if } X_i=X_{(1)} \text{ and } X_j=X_{(2)}\end{cases}$$

이므로 $\binom{n}{2}$ 개의 $\hat{\theta}^{-i,j}$들의 평균은

$$\hat{\theta}_3 = \overline{\hat{\theta}^{-i,j}} == \frac{2}{n(n-1)}\left[\frac{(n-1)(n-2)}{2}X_{(1)} + (n-2)X_{(2)} + X_{(3)}\right] \tag{5}$$

이다. 마찬가지 방법으로 σ에 대해서 구하면 다음과 같다.

$$\hat{\sigma}_2 = \overline{X} - \frac{n-1}{n}X_{(1)} - \frac{1}{n}X_{(2)} \tag{6}$$

$$\hat{\sigma}_3 = \overline{X} - \frac{n-2}{n}X_{(1)} - \frac{2(n-2)}{n(n-1)}X_{(2)} - \frac{2}{n(n-1)}X_{(3)} \tag{7}$$

이제 θ와 σ에 대해서 일반화잭나이프추정량과 고차일반화잭나이프추정량을 구하고자 한다. 먼저 θ에 대한 일반화잭나이프추정량 $G(\hat{\theta})$와 고차일반화잭나이프추정량 $G^{(2)}(\hat{\theta})$를 구하기 위해서 식 (2), (4), (5)의 기대값을 구하면

$$E(\hat{\theta}_1) = \theta + \frac{1}{n+1}H_{F_\theta}^{(1)}(0) + \frac{1}{(n+1)(n+2)}H_{F_\theta}^{(2)}(0) + \frac{1}{(n+1)(n+2)(n+3)}H_{F_\theta}^{(3)}(0) + \cdots,$$

$$E(\hat{\theta}_2) = \theta + \frac{1}{n}H_{F_\theta}^{(1)}(0) + \frac{1}{n(n+1)}H_{F_\theta}^{(2)}(0) + \frac{1}{n(n+1)(n+2)}H_{F_\theta}^{(3)}(0) + \cdots,$$

$$E(\hat{\theta}_3) = \theta + \frac{1}{n-1}H_{F_\theta}^{(1)}(0) + \frac{1}{n(n-1)}H_{F_\theta}^{(2)}(0) + \frac{1}{n(n-1)(n+1)}H_{F_\theta}^{(3)}(0) + \cdots$$

이다. 여기서,

$$H_{F_\theta}^{(1)}(0) = \frac{\exp(-\frac{\theta}{\sigma})}{(\frac{1}{\sigma}+1)\exp(-\frac{\theta}{\sigma}) - 1},$$

$$H_{F_\theta}^{(2)}(0) = \frac{\frac{1}{\sigma^2}\exp(-\frac{3\theta}{\sigma})}{[(\frac{1}{\sigma}+1)\exp(-\frac{\theta}{\sigma}) - 1]^3},$$

$$H_{F_\theta}^{(3)}(0) = \frac{\frac{1}{\sigma^3}\exp(-\frac{4\theta}{\sigma})[\frac{2}{\sigma}\exp(-\frac{\theta}{\sigma}) - \exp(-\frac{\theta}{\sigma}) + 1]}{[(\frac{1}{\sigma}+1)\exp(-\frac{\theta}{\sigma}) - 1]^5}$$

이다. 따라서 θ에 대한 일반화잭나이프추정량 $G(\hat{\theta})$는

$$G(\hat{\theta}) = \frac{\hat{\theta}_1 - R\hat{\theta}_2}{1-R} = 2X_{(1)} - X_{(2)}. \tag{8}$$

여기서, $R = \dfrac{n}{n+1}$ 이다.

또한 θ에 대한 고차일반화잭나이프추정량 $G^{(2)}(\hat{\theta})$를 구하면

$$G^{(2)}(\hat{\theta}) = 3X_{(1)} - 3X_{(2)} + X_{(3)} \tag{9}$$

이다.

θ에 대한 최우추정량과 잭나이프추정량들의 편의를 계산하면

$$\begin{gathered} E(\hat{\theta}) = \theta + O(n^{-1}) \\ E(G(\hat{\theta})) = \theta + O(n^{-2}) \\ E(G^{(2)}(\hat{\theta})) = \theta + O(n^{-3}) \end{gathered}$$

을 얻을 수 있다.

다음으로 σ에 대한 일반화잭나이프추정량 $G(\hat{\sigma})$와 고차일반화잭나이프추정량 $G^{(2)}(\hat{\sigma})$를 구하기 위해서 식 (3), (6), (7)의 기대값을 구하면

$$\begin{aligned} E(\widehat{\sigma_1}) = \sigma - \frac{1}{n+1} H_{F_\theta}^{(1)}(0) - \frac{1}{(n+1)(n+2)} H_{F_\theta}^{(2)}(0) \\ - \frac{1}{(n+1)(n+2)(n+3)} H_{F_\theta}^{(3)}(0) - \cdots , \end{aligned}$$

$$E(\widehat{\sigma_2}) = \sigma - \frac{1}{n} H_{F_\theta}^{(1)}(0) + \frac{1}{n(n+1)} H_{F_\theta}^{(2)}(0) - \frac{1}{n(n+1)(n+2)} H_{F_\theta}^{(3)}(0) - \cdots ,$$

$$E(\widehat{\sigma_3}) = \sigma - \frac{1}{n-1} H_{F_\theta}^{(1)}(0) + \frac{1}{n(n-1)} H_{F_\theta}^{(2)}(0) - \frac{1}{n(n-1)(n+1)} H_{F_\theta}^{(3)}(0) - \cdots$$

이다. 따라서 σ에 대한 일반화잭나이프추정량 $G(\hat{\sigma})$는

$$G(\hat{\sigma}) = \overline{X} - 2X_{(1)} + X_{(2)} \tag{10}$$

이고, σ에 대한 고차일반화잭나이프추정량 $G^{(2)}(\hat{\sigma})$는

$$G^{(2)}(\hat{\sigma}) = \overline{X} - 3X_{(1)} + 3X_{(2)} - X_{(3)} \tag{11}$$

이다.

σ에 대한 최우추정량과 잭나이프추정량들의 편의를 계산하면

$$E(\hat{\sigma}) = \sigma + O(n^{-1})$$
$$E(G(\hat{\sigma})) = \sigma + O(n^{-2})$$
$$E(G^{(2)}(\hat{\sigma})) = \sigma + O(n^{-3})$$

이다.

3. 모의실험 및 결론

제안된 잭나이프추정량들과 최우추정량들의 유효성을 편의와 평균제곱오차의 측면에서 서로 비교하고자 한다.

IMSL 부프로그램을 이용하여 독립적으로 난수를 1,000번 반복 생성하였고, 표본의 크기는 10, 20, 30, 50, 100으로 하고, 모수 값은 $\theta = 1, 2, 3$로 $\sigma = 1, 2, 5$로 하였다.

각 추정량 $\hat{\theta}$, $G(\hat{\theta})$, $G^{(2)}(\hat{\theta})$, $\hat{\sigma}$, $G(\hat{\sigma})$, $G^{(2)}(\hat{\sigma})$에 대한 추정된 편의와 평균제곱오차는 <표 1>과 같다. 표에서 나타난 결과의 경향은 모수 값의 변화에 영향을 받지 않았기에 <표 1>에는 $\theta = 2$와 $\sigma = 2$인 경우만 나타내었다.

<표 1> 추정량들의 편의(Bias)와 평균제곱오차(MSE)

n		$\hat{\theta}$	$G(\hat{\theta})$	$G^{(2)}(\hat{\theta})$	$\hat{\sigma}$	$G(\hat{\sigma})$	$G^{(2)}(\hat{\sigma})$
10	Bias	0.19251	-0.02581	0.01345	-0.20235	0.01596	-0.02330
	MSE	0.07333	0.07486	0.30177	0.37690	0.37796	0.68206
20	Bias	0.09938	-0.00678	-0.00709	-0.11607	-0.00992	-0.00961
	MSE	0.01938	0.01954	0.06824	0.19324	0.19472	0.28334
30	Bias	0.06863	0.00352	0.00771	-0.07272	-0.00760	-0.01179
	MSE	0.00919	0.00874	0.02518	0.12324	0.12752	0.13991
50	Bias	0.03939	-0.00146	0.00203	-0.03303	0.00782	0.00433
	MSE	0.00317	0.00313	0.01092	0.08854	0.08230	0.09831
100	Bias	0.02090	0.00111	0.00150	-0.01927	0.00051	0.00012
	MSE	0.00088	0.00086	0.00256	0.03870	0.03703	0.04003

<표 1>로부터 n에 관계없이 모든 잭나이프추정량들의 편의가 최우추정량의 편의보다 적게 나타났다. 반면 평균제곱오차는 n이 작을 때($n = 10, 20$)는 최우추정량 $\hat{\theta}$, $\hat{\sigma}$이 $G(\hat{\theta})$, $G(\hat{\sigma})$보다 조금 적고, n이 커져 갈수록 $\hat{\theta}$, $\hat{\sigma}$은 $G(\hat{\theta})$, $G(\hat{\sigma})$와 비슷해지는 반면 $G^{(2)}(\hat{\theta})$, $G^{(2)}(\hat{\sigma})$보다는 더 적게 나타났다.

따라서 왼쪽 절단된 지수모형인 경우 편의와 평균제곱오차 측면에서 일반화잭나이프추정량의 사용을 제안한다.

참고문헌

1. 강신수(2005). Coparison of EM with jackknife standard errors and multiple imputation standard errors, *한국데이터정보과학회지*, 16(4), 1079-1086.
2. 조길호, 김용구, 정성화(1998). Jackknife estimation for mean in exponential model with grouped and censored data, *한국통계학회논문집*, 5(3), 869-878.
3. Chakrabraty, R. P. and Rao J. N. K. (1968). The bias and stability of the jackknife variance estimator in ratio estimation, *JASA*, 63, 748.
4. Efron, B. and Stein, C. (1981). The jackknife estimate of variance, *Ann. Statist.*, 9, 586-596.
5. Hinkley, D. (1977). Jackknifing in unbalanced situation, *Technometrics*, 19, 85-292.
6. Miller, R. G. (1974). The jackknife - a review, *Biometrika,* 61, 1-15.
7. Quenouille, M. (1956). Notes on bias in estimation, *Biometrika,* 43, 353-360.
8. Rao, J. N. K. and Rao P. S. R. S. (1970). Some small results for ratio estimators, (abstract), *Ann. Math. Statist.*, 41, 1141-1142.
9. Rao, J. N. K. and Shao J. (1999). Modified balanced repeated replication for complex survey, *Biometrika,* 86, 403-415.
10. Robson, D. S. and Whitlock, J. H. (1964). Estimation of a truncation point, *Biometrika,* 51, 33-39.
11. Simonoff, J. S. and Tsai, C. (1986). Jackknife-based estimators and confidence regions in nonlinear regression, *Technometrics*, 28, 103-112.
12. Tukey, J. (1958). Bias and confidence in not quite large samples, (abstract), *Ann. Math. Statist.*, 29, 614.

[2006년 3월 접수, 2006년 4월 채택]