

지수 분포를 따르는 점진 제1종 구간 중도절단표본에서 모수 추정

신혜정¹ · 이광호² · 조영석³

¹²영남대학교 통계학과

³부산대학교 통계학과

접수 2010년 8월 8일, 수정 2010년 9월 20일, 게재확정 2010년 9월 23일

요약

본 논문은 점진 제1종 구간 중도절단표본과 점진 제2종 중도절단표본에서 모수를 추정하는 방법을 소개하고, 점진 제2종 중도절단표본에서 모수를 추정하는 방법을 활용하고자 점진 제1종 구간 중도절단표본에서 얻은 자료를 변환하여 모수를 추정하는 새로운 방법을 제안하였다. 척도모수가 미지인 지수 분포를 따르는 점진 제1종 구간 중도절단표본을 이용하여 점진 제2종 중도절단표본의 최대우도추정량을 사용하여 모수를 추정하였고, 모의실험을 통하여 두 방법에서 구한 추정량을 비교한 결과 본 논문에서 새로 제시한 방법을 이용하여 구한 모수의 추정량이 평균제곱오차 측면에서 더 우수한 추정량임이 나타났다.

주요용어: 구간 중도절단, 점진 제1종 구간 중도절단, 점진 제2종 중도절단, 지수분포, 최대우도추정량.

1. 서론

생존 분석에서 사용되는 많은 자료들은 중도절단된 형태로 많이 관측이 된다. 중도절단된 자료의 일반적인 형태가 구간 중도절단이다. 구간 중도절단(interval censored)은 정치, 경제, 산업, 의학, 사회과학 등의 자료에서 많이 발생하는데 특히, 의학 분야에서 임상실험과 건강연구의 장기적 변화를 다룬 연구나 공학실험에서 지속적으로 관찰할 수 있다. 구간 중도절단은 사건이 어느 구간 안에 발생한 것은 알려져 있지만 그 구간 안에서 정확한 발생 시점을 알지 못하는 경우를 말하는데, 이러한 구간 중도절단된 표본 연구는 모수적, 비모수적인 다양한 방법들이 있고, Lodon (1988), Kalbfleisch와 Prentice (2002), Lawless (2003), Sun (2006) 등이 다양한 방법들을 연구하였다.

제1종 중도절단표본 (Type I censored Sample)은 n 개의 개체를 미리 정해놓은 시간 (T)까지 실험하면서 개체들의 사건 발생 시간을 관측하여 얻게 되는 표본이고, 점진 제1종 구간 중도절단표본 (Progressively Type I Interval Censored Sample)은 n 개의 개체들 중에서 미리 정해진 시간의 구간들에서 사건이 발생된 개체 수를 측정하고, 각각의 구간에서 남아있는 생존개체들 중에서 일부를 제거하면서 얻게 되는 표본이다. 이러한 점진 제1종 중도절단표본은 Cohen (1963)이 논의하였고, 점진 제1종 구간 중도절단표본에서 지수 분포를 바탕으로 한 추정문제는 Aggarwala (2001)가 연구하였다. 로그정규

¹ (712-749) 경상북도 경산시 대동 214-1, 영남대학교 통계학과, 겸임교수.

² (712-749) 경상북도 경산시 대동 214-1, 영남대학교 통계학과, 교수.

³ 교신저자: (609-735) 부산광역시 금정구 부산광역시 금정구 부산대학교 63번길 2 (장전동), 부산대학교 통계학과, 부교수. E-mail: choys@pusan.ac.kr

분포는 Amin (2008)이 연구하였고, 지수와이블족에 대해서는 Ashour와 Afify (2007)가 연구하였으며, 와이블분포를 바탕으로 한 연구는 Ng와 Wang (2009)이 하였다. 또한 Chen과 Lio (2010)는 일반지수 분포에 대하여 연구하였다.

제2종 중도절단표본 (Type II censored Sample)은 n 개의 개체 중에서 미리 정해 놓은 m 번째 사건발생 시간이 관측되면 남아있는 $n - m$ 개의 개체는 버리고 실험을 중지하여 얻는 표본이며, 양측 개체를 버리는 연구로는 Kang 등 (1999)이 랄리분포에 대하여 연구하였고, Kang 등 (2003)은 지수분포의 적합도검정을 연구하였으며, Kang 등 (2005)은 이중지수분포의 모수를 추정하였다. 최근에 양측 개체와 중간에 일부를 버리는 연구로는 Cho 등 (2009)이 지수극단분포의 모수추정에 관하여 연구를 하였다.

점진 제2종 중도절단표본 (Progressively Type II censored Sample)은 제2종 중도절단표본의 일반적인 형태로 각각의 사건발생 시간이 관측될 때 마다 남아있는 개체 중에서 일부의 개체를 제거하여 자료를 얻는다. Balakrishnan과 Sandu (1996)는 이러한 점진 제2종 중도절단표본에서 모수를 추정하는 방법을 제시하였으며 이후에도 더 많은 연구들이 활발히 이루어지고 있다.

본 논문에서는 점진 제1종 구간 중도절단표본과 점진 제2종 중도절단표본에서 모수를 추정하는 방법을 소개하고, 점진 제2종 중도절단표본에서 모수를 추정하는 방법을 활용하고자 점진 제1종 구간 중도절단표본에서 얻은 자료를 변환하여 지수분포에서 모수를 추정하는 새로운 방법을 제안하고 추정량의 효율성을 비교해 보고자 한다.

2장에서는 점진 제1종 구간 중도절단표본에 대하여 설명하고, 지수분포에서 Aggarwala (2001)가 제시한 최대우도추정량을 소개한다. 3장에서는 점진 제2종 중도절단표본에 대하여 설명하고, 지수분포에서 Balakrishnan과 Sandu (1996)가 제시한 최대우도추정량을 소개한다. 4장에서는 지수분포를 따르는 점진 제1종 구간 중도절단표본에서 새로운 모수 추정방법을 설명하고, 이 방법을 이용한 모수를 추정한다. 5장에서는 모의실험을 통하여 추정량의 효율성을 비교하고, 6장에서 결론을 간략하게 언급한다.

2. 점진 제1종 구간 중도절단표본

일반적으로 제1종 또는 제2종 중도절단표본은 실험 개체를 실험이 종료되기 전에는 제거 하지 않는다. 그러나 점진 제1종 중도절단표본은 정해 놓은 시간까지 계속 임상 실험한다고 할 때, 실험 개체를 실험 종료 전에 제거 하는 방법이다.

Aggarwala (2001)에 의하면 점진 제1종 구간 중도절단표본은 다음과 같은 방법으로 얻어질 수 있다. 관찰하고자 하는 모든 n 개 개체의 사건이 발생하는 관찰 시작시점은 $T_0 = 0$ 이고, 끝나는 시점은 T_m 이며, 미리 정해놓은 시간 T_1, T_2, \dots, T_m (m 은 고정된 수)은 나누어진 m 개 구간에서 각각 사건이 발생하는 개체를 관찰하는 시점이고, R_1, R_2, \dots, R_m 은 각각의 T_1, T_2, \dots, T_m 시간에서 남아있는 개체 중에서 제거하는 개체들의 개수이다. 총 n 개 개체 중에서 첫 번째 구간 $(0, T_1]$ 에서 사건 발생 개체 수는 X_1 이고, 이 구간에서 남아 있는 개체 $n - X_1$ 개 중에서 제거하는 개체 수는 R_1 이다. 두 번째 구간 $(T_1, T_2]$ 에서 사건 발생 개체 수는 X_2 이고, 이 구간에서 남아 있는 개체는 $n - X_1 - X_2 - R_1$ 개 중에서 제거하는 개체 수는 R_2 개 이다. 이렇게 하여 m 번째 구간 $(T_{m-1}, T_m]$ 에서 사건 발생 개체 수는 X_m 이고, 남아 있는 개체 $n - \sum_{i=1}^m X_i - \sum_{i=1}^{m-1} R_i = R_m$ 개를 모두 제거한다. 이렇게 하여 모든 구간에서 제거되는 생존 개체들의 개수는 총 $\sum_{i=1}^m R_i$ 개이며, T_i 시간에 제거되는 생존 개체들의 개수 R_i 는 $\min(R_i, T_i$ 시간에 생존한 남은 개체들의 수)이다. 여기서 $i = 1, 2, \dots, m - 1$ 이며, R_m 는 T_m 시간에 생존한 모든 개체들의 수이다. 따라서 확률변수 X_1, X_2, \dots, X_m 은 $(0, T_1], (T_1, T_2], \dots, (T_{m-1}, T_m]$ 구간에서 각각 사건 발생 개체들의 수이며, 총 개체 수 $n = \sum_{i=1}^m (X_i + R_i)$ 로 나타난다. 이렇게 얻은 표본을 점진 제1종 구간 중도절단표본이라 하고, 점진 제1종 구간 중도절단에 대한 개요는 다음 그림 2.1과 같다.

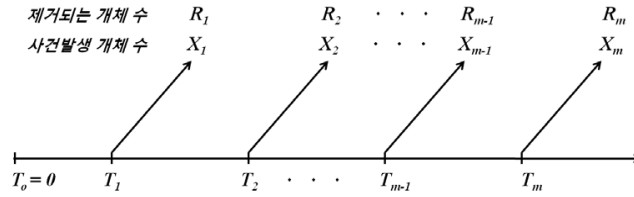


그림 2.1 점진 제 1종 구간 중도절단에 대한 개요

사건 발생 시간 (T)이 척도모수 (θ)를 가지는 지수분포 $Exp(0, \theta)$ 을 따른다고 가정하면, 확률분포함수 $f(t; \theta)$ 와 누적확률분포함수 $F(t; \theta)$ 는 다음과 같다.

$$f(t; \theta) = \frac{1}{\theta}e^{-t/\theta}, t > 0, \theta > 0, F(t; \theta) = 1 - e^{-t/\theta}, t > 0, \theta > 0.$$

척도모수 (θ)를 가지는 지수분포를 바탕으로 척도 모수 θ 의 최대우도추정량 (MLE)을 구해보면, 우도함수는

$$\begin{aligned} L(\theta) &\propto [F(T_1)]^{X_1} [1 - F(T_1)]^{R_1} [F(T_2) - F(T_1)]^{X_2} [1 - F(T_2)]^{R_2} \\ &\times \dots \times [F(T_m) - F(T_{m-1})]^{X_m} [1 - F(T_m)]^{R_m} \\ &= \prod_{i=1}^m [F(T_i) - F(T_{i-1})]^{X_i} [1 - F(T_i)]^{R_i} \\ &= \prod_{i=1}^m [e^{-T_{i-1}/\theta} - e^{-T_i/\theta}]^{X_i} e^{-(X_i T_i + R_i T_i)/\theta}. \end{aligned}$$

로그 우도함수는

$$\ln L(\theta) = \text{constant} + \sum_{i=1}^m X_i \ln [1 - e^{-(T_i - T_{i-1})/\theta}] - \frac{1}{\theta} \sum_{i=2}^m X_i T_i - \frac{1}{\theta} \sum_{i=1}^m R_i T_i.$$

θ 에 대한 MLE를 구하기 위하여 $\ln L(\theta)$ 을 θ 에 대하여 편미분을 하면

$$\frac{\partial \ln L(\theta)}{\partial \theta} = -\frac{1}{\theta^2} \sum_{i=1}^m \frac{X_i (T_i - T_{i-1}) e^{-(T_i - T_{i-1})/\theta}}{1 - e^{-(T_i - T_{i-1})/\theta}} + \frac{1}{\theta^2} \sum_{i=2}^m X_i T_i + \frac{1}{\theta^2} \sum_{i=1}^m R_i T_i = 0.$$

이 경우 정확한 θ 에 대한 MLE를 구하기 힘들기 때문에 각 구간이 등간격인 특별한 경우 즉, 주기적으로 절단 (censoring)이 발생하는 경우라 할 수 있으며 $T_i = i \times t$ 로 나타낼 수 있다. 이러한 경우,

$$\frac{\partial \ln L(\theta)}{\partial \theta} = -\frac{1}{\theta^2} \sum_{i=1}^m \frac{X_i t e^{-t/\theta}}{1 - e^{-t/\theta}} + \frac{1}{\theta^2} \sum_{i=2}^m t(i-1)X_i + \frac{1}{\theta^2} \sum_{i=1}^m R_i t = 0$$

을 이용하여 Aggarwala (2001)가 다음과 같이 θ 의 MLE를 제시하였다.

$$\hat{\theta}_{I,MLE} = \frac{t}{\ln [1 + \sum_{i=1}^m X_i / (\sum_{i=2}^m (i-1)X_i + \sum_{i=1}^m iR_i)]}.$$

3. 점진 제2종 중도절단표본

제2종 중도절단표본 (type II censored sample)은 총 n 개의 개체 중에서 m 개의 사건이 발생될 때까지 실험을 계속하여 얻어지는 표본으로 실험 중간에 실험 개체를 제거하지 않는다. 점진 제2종 중도절단표본은 제2종 중도절단표본의 일반적인 형태로 총 n 개의 개체 중에서 첫 번째 사건이 발생되면 남은 $n - 1$ 개의 개체 중에서 정해 놓은 IR_1 개의 개체를 제거하고, 두 번째 사건이 발생되면 남은 $n - 2 - IR_1$ 개의 개체 중에서 정해놓은 IR_2 개의 개체를 제거한다. 이렇게 하여 m 번째 사건이 발생되면 남은 $IR_m = n - m - IR_1 - IR_2 - \dots - IR_{m-1} - 1$ 개의 개체를 모두 제거하는 방법으로 사건 발생시간을 관측한다. 관측된 사건 발생 시간은 순서통계량 $Y_{1:m,n}, Y_{2:m,n}, \dots, Y_{m:m,n}$ 으로 나타내고, 이때, $IR_1 = IR_2 = \dots = IR_{m-1} = 0$ 이면 제2종 중도절단표본과 같은 형태가 된다. 그림 2.2는 점진 제2종 중도절단에 대한 개요이다.

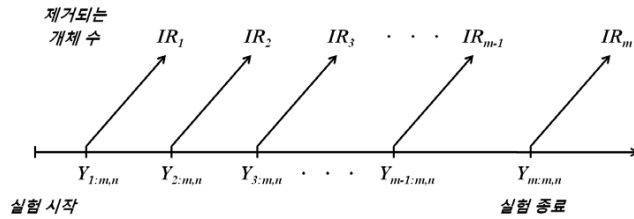


그림 3.1 점진 제 2종 중도절단에 대한 개요

사건 발생시간 Y 들이 누적분포함수 $F(y)$ 와 확률분포함수 $f(y)$ 을 가지는 연속인 모집단으로부터의 확률표본이라면, $Y_{1:m,n}, Y_{2:m,n}, \dots, Y_{m:m,n}$ 의 결합확률밀도함수는 다음과 같다.

$$L(\theta) = f_{Y_{1:m,n}, \dots, Y_{m:m,n}}(y_1, \dots, y_m) = c \prod_{i=1}^m f(y_i) [1 - F(y_i)]^{IR_i},$$

$$y_1 \leq y_2 \leq \dots \leq y_m$$

여기서 $c = n(n - IR_1 - 1) \dots (n - IR_1 - IR_2 - \dots - IR_{m-1} - m + 1)$ 이다.

점진 제2종 중도절단에서 사건발생 시간 Y 들이 척도모수 (θ)를 가지는 지수분포를 따른다고 가정하면, Balakrishnan과 Sandu (1996)가 제시한 척도모수에 대한 MLE는 다음과 같다.

$$\hat{\theta}_{II,MLE} = \frac{1}{m} \left[\sum_{j=2}^m (IR_j + 1) (Y_{j:m,n} - Y_{1:m,n}) + nY_{1:m,n} \right].$$

4. 새로운 모수 추정방법

점진 제1종 구간 중도절단은 총 n 개의 개체들 중 미리 정해놓은 m 개의 구간 $(0, T_1], (T_1, T_2], \dots, (T_{m-1}, T_m]$ 에서 각각 사건이 발생된 개체 수 X_1, X_2, \dots, X_m 을 관찰하고, 각각의 T_1, T_2, \dots, T_m 시간에서 남아있는 개체들 중에서 각각 R_1, R_2, \dots, R_m 개를 제거하는 방법이다. 이 자료를 바탕으로 $IM = X_1 + X_2 + \dots + X_m$ 개체의 사건 발생이 되는 정확한 시간 Y 들과 각 시점에서 제거 개수 IR 들을 생성할 수 있다면, 점진 제2종 중도절단표본에서 모수를 추정하는 방법을 적용할 수 있을 것으로 생각된다.

정확한 시간 Y 들과 각 시점에서 제거 개수 IR 들을 생성하려고 점진 제1종 구간 중도절단표본에서 사건 발생 개체 수 X_1, X_2, \dots, X_m 을 이용하여 각 구간에서 사건 발생 시간 Y 값들을 생성한다. 첫 번째 구간 $(0, T_1]$ 에서 사건발생이 관측된 개체 수 X_1 개의 사건 발생 시간 Y 값들을 같은 간격을 가지도록 순서대로 생성하고, 두 번째 구간 $(T_1, T_2]$ 에서도 관측된 개체 수 X_2 개의 사건 발생 시간 Y 값들을 생성하는 방식으로 관측된 모든 구간에서 사건발생 시간 Y 값들을 생성한다. 이렇게 생성된 사건 발생 시간들 $Y_1, Y_2, \dots, Y_{X_1}, Y_{X_1+1}, \dots, Y_{IM}$ 은 총 n 개의 개체 중에서 IM 개의 사건이 발생할 때까지 실험을 하게 되는 점진 제2종 중도절단표본과 같은 형태가 된다.

그리고 각각의 Y 시점에서 제거되는 생존 개체들은 $IR_1, IR_2, \dots, IR_{X_1}, IR_{X_1+1}, \dots, IR_{IM}$ 로 나타낸다. 점진 제1종 구간 중도절단에서는 $(0, T_1], (T_1, T_2], \dots, (T_{m-1}, T_m]$ 구간에서 남아있는 생존 개체들 중에서 각각 R_1, R_2, \dots, R_m 의 생존 개체를 제거하게 되지만, 본 논문에서 제시한 방법은 첫 번째 구간 $(0, T_1]$ 에서 Y_1, Y_2, \dots, Y_{X_1} 을 생성하게 되고, 생성된 Y 들의 각 시점에서 남아있는 생존 개체들 중 제거하는 생존 개체들의 수는 각각 $IR_1, IR_2, \dots, IR_{X_1}$ 가 된다. 이때 Y_1, Y_2, \dots, Y_{X_1} 의 어떤 사건 발생 시점에서 R_1 개의 생존 개체를 제거해야 할지를 생각해 봐야 한다. 각각의 생성된 Y 에서 생존 개체를 제거하는 방법으로 여러 가지 경우를 생각해 볼 수 있는데 본 논문에서는 구간 $(T_0, T_1]$ 의 마지막 사건 발생 시점 Y_{X_1} 값에서 생존 개체 R_1 을 제거하는 방법을 사용하였다.

$$\text{즉, } Y_1 = T_0 + \frac{T_1 - T_0}{X_1 + 1}, IR_1 = 0, Y_2 = T_0 + 2 \frac{T_1 - T_0}{X_1 + 1}, IR_2 = 0, \dots,$$

$$Y_{X_1} = T_0 + X_1 \frac{T_1 - T_0}{X_1 + 1}, IR_{X_1} = R_1.$$

$$\text{두 번째 구간 } (T_1, T_2] \text{에서 } Y_{X_1+1} = T_1 + \frac{T_2 - T_1}{X_2 + 1}, IR_{X_1+1} = 0.$$

$$Y_{X_1+2} = T_1 + 2 \frac{T_2 - T_1}{X_2 + 1}, IR_{X_1+2} = 0, \dots, Y_{X_1+X_2} = T_1 + X_2 \frac{T_2 - T_1}{X_2 + 1}, IR_{X_1+X_2} = R_2.$$

다음 구간에서도 같은 방법으로 생성하고, 마지막 구간 $(T_{m-1}, T_m]$ 에서

$$Y_{X_1+X_2+\dots+X_{m-1}+1} = T_{m-1} + \frac{T_m - T_{m-1}}{X_m + 1}, IR_{X_1+X_2+\dots+X_{m-1}+1} = 0,$$

$$Y_{X_1+X_2+\dots+X_{m-1}+2} = T_{m-1} + 2 \frac{T_m - T_{m-1}}{X_m + 1}, IR_{X_1+X_2+\dots+X_{m-1}+2} = 0, \dots,$$

$$Y_{X_1+X_2+\dots+X_{m-1}+X_m} = T_{m-1} + X_m \frac{T_m - T_{m-1}}{X_m + 1}, IR_{X_1+X_2+\dots+X_{m-1}+X_m} = R_m.$$

만약, 어떤 구간에서 사건발생 $X_i = 0$ 개라면 그 구간에서 제거하는 R_i 는 다음 구간에 사건 발생 관측이 존재한다면, 그 구간에서 제거한 것으로 생각한다. 마지막 구간까지 관측이 존재하지 않는다면, $IR_{X_1+X_2+\dots+X_{i-1}} = R_i + R_i + 1 + \dots + R_m$ 로 계산하고, $n = IM + IR_1 + \dots + IR_{IM}$ 된다.

이와 같은 방법으로 자료를 변환하여 Y 와 IR 들을 생성하고, 3장에서 소개한 점진 제2종 중도절단 표본에서 최대우도추정량을 이용하여, 척도모수 (θ)를 가지는 지수분포 $Exp(0, \theta)$ 를 따르는 점진 제1종 구간 중도절단표본에서 척도모수에 대한 새로운 추정량을 다음과 같이 제시한다.

$$\hat{\theta}_{II} = \frac{1}{IM} \left[\sum_{j=2}^{IM} (IR_j + 1) (Y_{j:IM,n} - Y_{1:IM,n}) + nY_{1:IM,n} \right].$$

5. 모의실험

개체 n 개의 관찰 시작시점은 $T_0 = 0$ 이고, 끝나는 시점은 T_m 이며, 미리 정해놓은 시간 T_1, T_2, \dots, T_m (m 은 고정된 수)으로 나누어진 m 개의 각 구간에서 제거 할 개체들의 비율 p_1, p_2, \dots, p_{m-1} , ($p_m =$

1)으로부터 지수분포를 따르는 점진 제1종 구간 중도절단표본 X_1, X_2, \dots, X_m 을 생성하는 알고리즘 (Aggarwala, 2001)을 소개하면 다음과 같다.

$X_1 \sim BIN[n, F(T_1)]$ 이고, $i = 2, 3, \dots, m$ 에 대하여

$$\begin{aligned} & X_i | X_{i-1}, \dots, X_1, R_{i-1}, \dots, R_1 \\ & \sim BIN \left[n - \sum_{j=1}^{i-1} (X_j + R_j), \frac{F(T_i) - F(T_{i-1})}{1 - \sum_{j=1}^{i-1} [F(T_j) - F(T_{j-1})]} \right] \\ & = BIN \left[n - \sum_{j=1}^{i-1} (X_j + R_j), \frac{F(T_i) - F(T_{i-1})}{1 - F(T_{i-1})} \right]. \end{aligned}$$

m 개의 이항확률변수를 발생시키는 알고리즘은 다음 단계와 같이 수행한다.

단계 1. 각각의 변수를 $i = 0, xsum = 0, rsum = 0$ 로 초기화 한다.

단계 2. 다음 i 로

단계 3. $i = m + 1$ 이면 알고리즘을 끝낸다.

단계 4. $BIN[n - xsum - rsum, F(T_i) - F(T_{i-1}) / [1 - F(T_{i-1})]]$ 을 따르는 이항확률변수 X_i 을 발생 시킨다.

단계 5. $R_i = floor [p_i(n - xsum - rsum - X_i)]$ 을 계산한다.

단계 6. $xsum = xsum + X_i, rsum = rsum + R_i$ 을 계산한다.

단계 7. 단계 2.로 돌아간다.

여기에서 $p_1, p_2, \dots, p_{m-1}, (p_m = 1)$ 과 $(R_m = n - \sum_{i=1}^m X_i - \sum_{i=1}^m R_i)$ 은 고정된 값이다.

지수분포를 바탕으로 점진 제1종 구간 중도절단표본에서 Aggarwala (2001)가 추정한 $\hat{\theta}_{I,MLE}$ 와 4장에서 제안한 새로운 방법을 이용하여 추정한 $\hat{\theta}_{II}$ 의 평균과 평균제곱오차 (MSE)를 비교하기 위하여 위에 제시한 알고리즘을 이용하여 모의실험을 하였다.

개체 $n = 20, 30, 50$, 구간 $m = 10$, 간격 $t = 1$, 척도모수 $\theta = 3$ 으로 두고, 각 구간에서 제거할 개체들의 비율을 $\mathbf{p}_A = (0.5, 0.5, 0.5, 0.5, 0, 0, 0, 0, 0, 1)$, $\mathbf{p}_B = (0.25, 0.25, 0.25, 0.25, 0, 0, 0, 0, 0, 1)$ 으로 했을 때, 1000번 반복하여 모의실험을 실시한 평균과 평균제곱오차를 다음 표5.1에 제시하였다.

표 5.1 평균과 평균제곱오차

n		$\hat{\theta}_{I,MLE}$		$\hat{\theta}_{II}$	
		평균	평균제곱오차	평균	평균제곱오차
20	\mathbf{p}_A	3.227339	1.294448	2.742067	0.648314
	\mathbf{p}_B	3.098608	0.756374	2.895952	0.535993
30	\mathbf{p}_A	3.145513	0.795417	2.798690	0.500820
	\mathbf{p}_B	3.071997	0.475510	2.925752	0.373977
50	\mathbf{p}_A	3.087147	0.430683	2.876513	0.317347
	\mathbf{p}_B	3.029961	0.281448	2.935406	0.240818

위 표5.1의 결과를 보면 점진 제1종 구간 중도절단에서 $\hat{\theta}_{I,MLE}$ 의 평균제곱오차보다 본 논문에서 새로 제시한 $\hat{\theta}_{II}$ 의 평균제곱오차가 더 작게 나타나 $\hat{\theta}_{II}$ 가 평균제곱오차측면에서 일치성을 가지며 더 좋은 추정량임을 알 수 있고, 편의 측면에서도 일치성을 가지는 추정량임을 알 수 있다.

6. 결론

본 논문에서는 점진 제2종 중도절단표본에서 모수를 추정하는 방법을 활용하고자 점진 제1종 구간 중도절단표본에서 얻은 자료를 변환하여 모수를 추정하는 새로운 방법을 제안하였다. 모의실험을 통하여

본 논문에서 새로 제시한 방법을 이용하여 추정된 모수의 추정량이 Aggarwala (2001)가 제시한 추정량보다 평균제곱오차 측면에서 일치성을 가지며 더 좋은 것으로 나타났고, 편의 측면에서도 일치성을 가지는 추정량임을 알 수 있었다. 새로 제안한 방법에서 자료를 생성하는 방법은 사건 발생 시간 Y 값들의 생성간격을 구간 내에서 등 간격으로 하였지만, 앞으로 사건 발생 시간 Y 값들의 생성 간격을 등 간격이 아닌 다른 방법으로 해볼 수도 있을 것이다. 그리고 생존 개체 제거 시점도 각 구간의 마지막에 생성된 사건 발생 시간 Y 로 제한하였다. 하지만 제거 시점도 다르게 변환시켜 연구를 해 볼 수도 있을 것이다. 또한 주어지는 구간의 간격도 연속으로 순서화된 등 간격이 아닌 길이가 다르게 나타나는 구간의 경우들에 대해서도 연구를 해 볼 수도 있을 것이다.

참고문헌

- Aggarwala, R. (2001). Progressive interval censoring: Some mathematical results with applications to inference. *Communications in Statistics-Theory and Methods*, **30**, 1921-1935.
- Amin, Z. H. (2008). A note on the parameter estimation for the lognormal distribution based on progressively Type I interval censored samples. *Model Assisted Statistics Application*, **3**, 169-176.
- Ashour, S. K. and Afify, W. M. (2007). Statistical analysis exponentiated Weibull Family under Type I progressive interval censoring with random removal. *Journal of Applied Sciences Research*, **3**, 1851-1863.
- Balakrishnan, N. and Sandu, R. A. (1996). Best linear unbiased and maximum likelihood estimation for exponential distributions under general progressive TypeII censored sample. *Sanky ā : The Indian Journal of Statistics*, **58**, 1-9.
- Chen, D. G. and Lio, Y. L. (2010). Parameter estimation for generalized exponential distribution under progressive Type-I interval censoring. *Computational Statistics and Data Analysis*, **54**, 1581-1591.
- Cho, Y. S., Kang, S. B. and Han, J. T. (2009). The exponentiated extreme value distribution. *Journal of the Korean Data & Information Science Society*, **20**, 719-731.
- Cohen, A. C. (1963). Progressively censoring censored samples in life testing. *Technometrics*, **5**, 327-339.
- Kalbfleisch, J. D. and Prentice, R. L. (2002). *The statistical analysis of failure time data*, Second edition, John Wiley, New York.
- Kang, S. B., Cho, Y. S. and Choi, S. Y. (2003). Tests for the exponential distribution based on Type-II censored samples. *Journal of the Korean Data & Information Science Society*, **14**, 367-346.
- Kang, S. B., Cho, Y. S. and Han, J. T. (2005). Estimation for the double exponential distribution based on Type-II censored samples. *Journal of the Korean Data & Information Science Society*, **16**, 115-126.
- Kang, S. B., Cho, Y. S. and Hwang, K. M. (1999). AMLE for the Rayleigh Distribution with Type-II censoring. *Journal of the Korean Data & Information Science Society*, **10**, 405-413.
- Lawless, J. F. (2003). *Statistical models and methods for lifetime data*, John Wiley, New York.
- Lodon, D. (1988). *Survival Models*, Actex Publications, Connecticut.
- Ng, T. H. K. and Wang, Z. (2009). Statistical estimation for the parameters of Weibull distribution based on progressively Type-I interval censored sample. *Journal of Statistical Computation and Simulation*, **79**, 145-159.
- Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data*, Springer Verlag, New York.

Parameter estimation for exponential distribution under progressive type I interval censoring

Hyejung Shin¹ · Kwangho Lee² · Youngseuk Cho³

¹²Department of Statistics, Yeungnam University

³Department of Statistics, Pusan National University

Received 8 August 2010, revised 20 September 2010, accepted 23 September 2010

Abstract

In this paper, we introduce a method of parameter estimation of progressive Type I interval censored sample and progressive type II censored sample. We propose a new parameter estimation method, that is converting the data which obtained by progressive type I interval censored, those data be used to estimate of the parameter in progressive type II censored sample. We used exponential distribution with unknown scale parameter, the maximum likelihood estimator of the parameter calculates from the two methods. A simulation is conducted to compare two kinds of methods, it is found that the proposed method obtains a better estimate than progressive Type I interval censoring method in terms of mean square error.

Keywords: Exponential distribution, interval censoring, maximum likelihood estimation, progressive type I interval censoring, progressive type II censoring.

¹ Adjunct professor, Department of Statistics, Yeongnam University, Gyeongsan 712-749, Korea.

² Professor, Department of Statistics, Yeongnam University, Gyeongsan 712-749, Korea.

³ Corresponding author: Associate professor, Department of Statistics, Pusan National University, Busan 609-735, Korea. E-mail: choys@pusan.ac.kr