한국통계학회논문집 제4권 2호
1997년 8월 pp. 557 - 563

이변량 1종 중단된 Marshall-Olkin 모형에서 동일성과 독립성 검정

김 희 제[1)], 조 장 식[2)]

요 약

Marshall-Olkin의 이변량 지수모형을 따르는 두개의 부품으로 이루어진 시스템에서, 두 부품의 수명들이 이변량 1종 중단된 자료로 관찰되는 경우, 모수에 대한 최우추정량을 구한다. 그리고 두 부품의 수명에 대한 근사적 독립성과 동일성 검정법을 제안하고 몬테칼로 모의실험을 통하여 검정력을 비교하였다.

1. 서 론

두개의 부품으로 구성되어 있는 어떤 시스템을 생각할 때 두 부품의 수명시간을 (X, Y)로 둔다면, X 와 Y는 일반적으로 서로 상관관계가 있는 종속적인 확률변수가 되는 경우가 많이 있다. 예를들어 사람의 양쪽 눈, 양쪽 콩팥 등 쌍으로 이루어진 시스템을 생각하면 각 쌍들의 수명은 서로 상관관계가 있다. 이와같이 서로 상관관계가 있는 두 부품으로 이루어진 어떤 시스템의 수명시간에 대한 모형으로서 Marshall과 Olkin(1967)은 이변량 지수분포를 제안하면서 그 분포의 여러가지 중요한 성질을 밝혔다. 확률변수 (X, Y)가 다음과 같은 결합생존함수로 주어진다면 모수 $\lambda_1, \lambda_2, \lambda_3$ 를 갖는 Marshall과 Olkin의 이변량 지수 분포를 따른다고 한다.

$$\begin{aligned}\overline{F}(x, y) &= P(X > x\ ,\ \ Y > y)\ ,\quad x, y \geq 0 \\ &= \exp\{-\lambda_1 x - \lambda_2 y - \lambda_3 \max(x, y)\},\quad x, y \geq 0, \qquad (1.1)\end{aligned}$$

여기서 $\lambda_1,\ \lambda_2,\ \lambda_3 > 0$ 이다. $\lambda = \lambda_1 + \lambda_2 + \lambda_3$이라 둔다면 λ_3/λ 는 이변량 지수분포에서 X 와 Y의 상관계수가 되고 $P(X = Y)$와도 같음은 잘 알려져 있다. 그래서 모수 $\lambda_3 = 0$인 것은 두 부품이 서로 독립이다는 것과 같다. 또한 $\lambda_1 = \lambda_2$인 것은 X와 Y의 분포가 동일하다는 것과 같다.

완전한 자료가 관찰되는 경우, Arnold(1968)는 이 모수들에 대한 점추정치를 계산하였고, Bemis, Bain과 Higgins(1972) 그리고 Bhattacharyya와 Johnson(1973)은 $\lambda_1 = \lambda_2$인 가정하에서의

---

1) (608-736) 부산광역시 남구 대연동 110 경성대학교 전산통계학과 교수.
2) (608-736) 부산광역시 남구 대연동 110 경성대학교 전산통계학과 전임강사.

독립성 검정방법을 연구하였다. 그리고 Hanagal과 Kale(1991)은 $\lambda_1=\lambda_2$인 가정이 없는 상태에서 근사적 정규성을 이용하여 몇가지 독립성 및 동일성검정법에 관한 연구를 하였다. 한편 두 부품에 대한 중단시간이 동일한 일변량 중단된 자료(univariate censored data)로 관찰되는 경우 그들은 (1992) 모수에 대한 최우추정량을 구하고, 근사적 독립성 및 동일성 검정법을 제안하였다.

그러나 현실적으로, 부품들의 수명이 실험자의 의도에 의해서 또는 실험환경의 제약에 의해서 두 부품에 대한 중단시간이 다른 이변량 1종 중단된 자료(bivariate type 1 censored data)로 관찰되는 경우가 많이 발생한다. 예를 들어 두 부품의 수명이 동일한 분포를 갖지 않는 경우 두 부품의 중단시간을 일변량 중단모형보다 이변량 중단모형으로 하는 것이 타당하다.

본 논문에서는 Marshall-Olkin의 이변량 지수모형을 따르는 두 부품의 수명이 이변량 1종 중단된 자료로 관찰되는 경우, 모수들에 대한 최우추정량을 구하고 그 추정량의 근사적 정규성을 이용하여 모형에 대한 독립성 및 동일성 검정법을 제안한다.

2. Marshall-Olkin 모형의 개요 및 기호

(X, Y)를 Marshall과 Olkin의 이변량 지수모형을 따르는 두 부품의 수명시간이라고 하자. 그러면 두 부품의 결합 생존함수 $\overline{F}(x,y)$는 (1.1)과 같으며, 결합 확률밀도함수 $f(x,y)$는 다음과 같이 주어진다.

$$f(x,y) = \begin{cases} \lambda_1(\lambda_2+\lambda_3)\cdot\exp(-\lambda_1 x-(\lambda_2+\lambda_3)y), & y>x \\ \lambda_2(\lambda_1+\lambda_3)\cdot\exp(-(\lambda_1+\lambda_3)x-\lambda_2 y), & x>y \\ \lambda_3\cdot\exp(-\lambda x), & x=y. \end{cases} \tag{2.1}$$

조건부 생존함수는 각각 다음과 같다.

$$\begin{aligned} \overline{F}_{X\mid Y=y}(x) &= P(X>x\mid Y=y) \\ &= \begin{cases} \exp(-\lambda_1 x), & y>x \\ \lambda_2(\lambda_2+\lambda_3)^{-1}\exp(-(\lambda_1+\lambda_3)x+\lambda_3 y), & x\geq y, \end{cases} \end{aligned} \tag{2.2}$$

$$\begin{aligned} \overline{F}_{Y\mid X=x}(y) &= P(Y>y\mid X=x) \\ &= \begin{cases} \exp(-\lambda_2 y), & x>y \\ \lambda_1(\lambda_1+\lambda_3)^{-1}\exp(-\lambda_3(y-x)-\lambda_2 y), & x\leq y. \end{cases} \end{aligned} \tag{2.3}$$

그리고 본 논문에서 사용되는 기호들을 소개하면 다음과 같다.

t_{x_i}, $i=1,2,\cdots,n$: X에 대한 i번째 관찰치의 중단시간.

t_{y_i}, $i=1,2,\cdots,n$: Y에 대한 i번째 관찰치의 중단시간.

$G_{1i}=I(X_i>t_{x_i})$, $G_{1i}^{*}=1-G_{1i}$, $i=1,2,\cdots,n$.

$G_{2i}=I(Y_i>t_{y_i})$, $G_{2i}^{*}=1-G_{2i}$, $i=1,2,\cdots,n$.

$R_{1i}=I(X_i < Y_i)$, $R^*_{1i}=1-R_{1i}$, $i=1,2,\cdots,n$.

$R_{2i}=I(X_i > Y_i)$, $R^*_{2i}=1-R_{2i}$, $i=1,2,\cdots,n$.

$R_{3i}=I(X_i = Y_i)$, $R^*_{3i}=1-R_{3i}$, $i=1,2,\cdots,n$.

$\underline{\lambda}=(\lambda_1,\lambda_2,\lambda_3)$.

$\xrightarrow{d}$: 분포수렴.

$\phi(\cdot)$: 검정함수

3. 모수에 대한 최우추정량

이변량 1종 중단된 자료가 관찰되는 경우, 각각의 부품들에 대해서 다음과 같은 세가지 경우가 발생할 수 있다.

(1) 두개의 부품들이 모두 관찰되는 경우($G^*_{1i}G^*_{2i}=1$).

(2) 하나의 부품만 관찰되고 다른 하나의 부품은 중단되는 경우($G_{1i}G^*_{2i}+G^*_{1i}G_{2i}=1$).

(3) 두개의 부품들이 모두 중단되는 경우($G_{1i}G_{2i}=1$).

그러면 부품들의 i번째 수명시간 (x_i,y_i)은 다음과 같이 관찰된다.

$$(x_i,y_i) = \begin{cases} (x_i,y_i), & x_i < t_{x_i}\ ,\ y_i < t_{y_i} \\ (t_{x_i},y_i), & x_i > t_{x_i}\ ,\ y_i < t_{y_i} \\ (x_i,t_{y_i}), & x_i < t_{x_i}\ ,\ y_i > t_{y_i} \\ (t_{x_i},t_{y_i}), & x_i > t_{x_i}\ ,\ y_i > t_{y_i}\ , \end{cases} \tag{3.1}$$

여기서 중단시간 t_{x_i}와 t_{y_i} $,i=1,2,\cdots,n$는 미리 정해진 값이며, 같은 값으로 하는 경우가 많다. 특히, $t_{x_i}=t_{y_i}$인 경우는 일변량 1종 중단모형이 된다.

$\underline{\lambda}$에 대한 우도함수는 다음과 같이 계산되어진다.

$$\begin{aligned} L(\underline{\lambda}) = & \prod_{i=1}^{n}\{[f(x_i,y_i)]^{G^*_{1i}G^*_{2i}}\cdot[\overline{F}(x_i,y_i)]^{G_{1i}G_{2i}}\cdot \\ & [\overline{F}_{X|Y=y}(x_i)f_Y(y_i)]^{G_{1i}G^*_{2i}}\cdot[\overline{F}_{Y|X=x}(y_i)f_X(x_i)]^{G^*_{1i}G_{2i}}\}^{(R_{1i}+R_{2i}+R_{3i})} \\ = & \lambda_1^{n_1}\lambda_2^{n_2}\lambda_3^{n_3}(\lambda_1+\lambda_3)^{n_4}(\lambda_2+\lambda_3)^{n_5}\exp[-\lambda_1 x_s-\lambda_2 y_s-\lambda_3(x_s+y_s-t_s)]. \end{aligned} \tag{3.2}$$

여기서 $n_1=\sum_{i=1}^{n}(R_{1i}G^*_{1i}G^*_{2i}+R^*_{2i}G^*_{1i}G_{2i})$, $n_2=\sum_{i=1}^{n}(R_{2i}G^*_{1i}G^*_{2i}+R^*_{1i}G_{1i}G^*_{2i})$,

$n_3=\sum_{i=1}^{n}R_{3i}G^*_{1i}G^*_{2i}$, $n_4=\sum_{i=1}^{n}R_{2i}G^*_{1i}$, $n_5=\sum_{i=1}^{n}(R_{1i}G^*_{1i}G^*_{2i}+R_{1i}G_{1i}G^*_{2i})$,

$$x_s=\sum_{i=1}^{n}x_i,\ \ y_s=\sum_{i=1}^{n}y_i,\ \ t_s=\sum_{i=1}^{n}\min(x_i,y_i).$$

그리고 $n_1, n_2, \cdots, n_5$는 확률변수이며 이들의 기대값은 다음과 같이 계산되어 진다.

$$\begin{aligned}E(n_1)=\sum_{i=1}^{n}\{&\lambda_1/\lambda-\lambda_1\exp(-\lambda t_{x_i})/\lambda+\exp(-\lambda t_{y_i})-\exp(-(\lambda_2+\lambda_3)t_{y_i})\\&+(1-\exp(-\lambda_1 t_{x_i}))\cdot\exp(-(\lambda_2+\lambda_3)t_{y_i})\cdot I(t_{x_i}<t_{y_i})\\&+\lambda_3(\exp(-\lambda t_{y_i})-\exp(-\lambda t_{x_i}))/\lambda\cdot I(t_{y_i}<t_{x_i})\},\end{aligned}$$

$$\begin{aligned}E(n_2)=\sum_{i=1}^{n}\{&\lambda_2/\lambda-\lambda_2\exp(-\lambda t_{y_i})/\lambda+\exp(-(\lambda_1+\lambda_3)t_{x_i}-\lambda_2 t_{y_i})-\exp(-(\lambda_1+\lambda_3)t_{x_i})\\&+(1-\exp(-\lambda_2 t_{y_i}))\cdot\exp(-(\lambda_1+\lambda_3)t_{x_i})\cdot I(t_{y_i}<t_{x_i})\\&+\lambda_3(\exp(-\lambda t_{x_i})-\exp(-\lambda t_{y_i})))/\lambda\cdot I(t_{x_i}<t_{y_i})\},\end{aligned}$$

$$E(n_3)=\sum_{i=1}^{n}\{(\lambda_3-\lambda_3\exp(-\lambda\min(t_{x_i},t_{y_i})))/\lambda\},$$

$$\begin{aligned}E(n_4)=\sum_{i=1}^{n}\{&\lambda_2/\lambda-\lambda_2\exp(-\lambda t_{y_i})/\lambda+\exp(-(\lambda_1+\lambda_3)t_{x_i}-\lambda_2 t_{y_i})-\exp(-(\lambda_1+\lambda_3)t_{x_i})\\&+[\exp(-\lambda_2 t_{y_i})\cdot[1-\exp(-(\lambda_1+\lambda_3)t_{x_i})]\\&+(\lambda_1+\lambda_3)\cdot(\exp(-\lambda t_{x_i})-1)/\lambda]\cdot I(t_{x_i}>t_{y_i})\},\end{aligned}$$

$$\begin{aligned}E(n_5)=\sum_{i=1}^{n}\{&\lambda_1/\lambda-\lambda_1\exp(-\lambda t_{x_i})/\lambda+\exp(-\lambda t_{y_i})-\exp(-(\lambda_2+\lambda_3)t_{y_i})\\&+\lambda_1\exp(-\lambda t_{x_i})/\lambda-\exp(-(\lambda_2+\lambda_3)t_{y_i}-\lambda_1 t_{x_i})\}.\end{aligned}$$

또한 로그-우도함수는 다음과 같이 주어진다.

$$\begin{aligned}\log L(\underline{\lambda})=\ &n_1\log\lambda_1+n_2\log\lambda_2+n_3\log\lambda_3+n_4\log(\lambda_1+\lambda_3)+n_5\log(\lambda_2+\lambda_3)\\&-\lambda_1 x_s-\lambda_2 y_s-\lambda_3(x_s+y_s-t_s).\end{aligned}\tag{3.3}$$

따라서 (3.3)의 로그-우도함수를 모수들에 대해서 일차 편미분한 우도방정식은 다음과 같다.

$$\frac{\partial}{\partial\lambda_1}\log L(\underline{\lambda})=\frac{n_1}{\lambda_1}+\frac{n_4}{\lambda_1+\lambda_3}-x_s=0.\tag{3.4}$$

$$\frac{\partial}{\partial\lambda_2}\log L(\underline{\lambda})=\frac{n_2}{\lambda_2}+\frac{n_5}{\lambda_2+\lambda_3}-y_s=0.\tag{3.5}$$

$$\frac{\partial}{\partial\lambda_3}\log L(\underline{\lambda})=\frac{n_3}{\lambda_3}+\frac{n_4}{\lambda_1+\lambda_3}+\frac{n_5}{\lambda_2+\lambda_3}-(x_s+y_s-t_s)=0.\tag{3.6}$$

모수 $\lambda_1, \lambda_2, \lambda_3$에 대한 최우추정량($\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3$)은 위의 우도방정식 (3.4)-(3.6)을 뉴턴-랍슨과 같은 반복적 방법에 의해 얻을 수 있다.

따라서 $\sqrt{n}\ (\hat{\underline{\lambda}}-\underline{\lambda})$의 분포는 표본의 크기가 커짐에 따라 근사적으로 평균벡터가 영이고 분산-공분산행렬이 $I^{-1}(\underline{\lambda})$인 다변량 정규분포를 따름을 알 수 있다. 즉,

$$\sqrt{n}\ (\hat{\underline{\lambda}}-\underline{\lambda}) \xrightarrow{d} N(\underline{0}\ ,\ I^{-1}(\underline{\lambda})), \tag{3.7}$$

여기서 $I(\underline{\lambda})=E\left[-\frac{\partial^2}{\partial\lambda_i\partial\lambda_j}\log L(\underline{\lambda})\right]=((I_{ij}));i,j=1,2,3$이며

$$I_{11}=E(n_1)/\lambda_1^2+E(n_4)/(\lambda_1+\lambda_3)^2,\ \ I_{12}=0,\ \ I_{13}=E(n_4)/(\lambda_1+\lambda_3)^2,$$
$$I_{22}=E(n_2)/\lambda_2^2+E(n_5)/(\lambda_2+\lambda_3)^2, I_{23}=E(n_5)/(\lambda_2+\lambda_3)^2,$$
$$I_{33}=E(n_3)/\lambda_3^2+E(n_4)/(\lambda_1+\lambda_3)^2+E(n_5)/(\lambda_2+\lambda_3)^2.$$

4. 동일성과 독립성 검정

우선 두 부품의 수명이 동일한 분포를 갖는가에 대한 근사적 검정법을 생각하자.

3장에서 (3.7)의 근사적 정규성을 이용하여 동일성 검정을 할 수 있다. 즉, 귀무가설 $H_0:\lambda_1=\lambda_2$ 하에서 $\sqrt{n}\ (\hat{\lambda}_1-\hat{\lambda}_2)$는 평균이 0이고 분산이 $I^{11}+I^{22}-2I^{12}$인 근사적 정규분포를 따른다는 것을 알 수 있다. 여기서 I^{ij}, $i,j=1,2,3$는 행렬 $I^{-1}(\underline{\lambda})$의 (i,j)번째 원소이다.

따라서 동일성 검정의 가설 $H_0:\lambda_1=\lambda_2$ 대 $H_1:\lambda_1>\lambda_2$에 대한 검정법을 다음과 같이 제안한다.

$$\phi(x,y)=\begin{cases}1, & \sqrt{n}(\hat{\lambda}_1-\hat{\lambda}_2)/\sqrt{\hat{I}^{11}+\hat{I}^{22}-2\hat{I}^{12}}\ >\ z_\alpha\\ 0, & \text{그외.}\end{cases} \tag{4.1}$$

여기서 z_α는 표준정규분포의 오른쪽 꼬리면적이 α가 되는 값이며, $\hat{I}^{ij}$, $i,j=1,2,3$는 I^{ij}에서 λ_i 대신에 $\hat{\lambda}_i$를 대입한 값이다.

같은 방법으로 가설 $H_0:\lambda_1=\lambda_2$대 $H_1:\lambda_1<\lambda_2$에 대한 검정법을 다음과 같이 제안한다.

$$\phi(x,y)=\begin{cases}1, & \sqrt{n}(\hat{\lambda}_1-\hat{\lambda}_2)/\sqrt{\hat{I}^{11}+\hat{I}^{22}-2\hat{I}^{12}}\ <\ -z_\alpha\\ 0, & \text{그외.}\end{cases} \tag{4.2}$$

한편 두 부품의 수명에 대한 독립성 검정을 생각해 보자. (3.7)의 근사적 정규성을 이용하여, $\sqrt{n}\ (\hat{\lambda}_3-\lambda_3)$는 평균이 0이고 분산이 I^{33}인 근사적 정규분포를 따름을 알 수 있다. 따라서 가설 $H_0:\lambda_3=0$대 $H_1:\lambda_3>0$에 대한 검정법을 다음과 같이 제안한다.

$$\phi(x,y)=\begin{cases}1, & \sqrt{n}\cdot\hat{\lambda}_3/\sqrt{\hat{I}^{33}}\ >\ z_\alpha\\ 0, & \text{그외.}\end{cases} \tag{4.3}$$

5. 모의실험 및 결론

이 절에서는 앞에서 제안한 근사적 동일성과 독립성 검정법의 검정력을 몬테칼로 모의실험을 통하여 계산하고자 한다. 그러나 동일성과 독립성 검정의 검정력이 거의 같은 경향을 보이기 때문에, 여기서는 동일성 검정의 검정력만 나타내기로 한다. 우선 Marshall-Olkin의 이변량 지수분포의 난수는 Friday와 Patil(1977)이 제안한 방법으로 생성하였으며, 사용된 모수는 $\lambda_1 = 0.08$과 $\lambda_3 = 0.03$는 고정시켜 놓고, λ_2의 값을 0.1, 0.13, 0.16으로 변화시켜 가면서 동일성 검정 $H_0: \lambda_1 = \lambda_2$ 대 $H_1: \lambda_1 < \lambda_2$에 대한 검정력을 유의수준 0.05와 0.01에 대하여 계산하였다. 그리고 각 모수들에 대하여 표본의 수는 30, 50, 100개를 생성하였으며, 각각의 표본의 수에서 500번 반복하여 검정력을 계산하였다. 이변량 1종 중단시간은 모든 i에 대해서 $(t_{x_i}, t_{y_i}) = (\infty, \infty)$, (25.0, 17.50), (16.50, 9.50)으로 사용하였다. 여기에 대한 결과는 <표 1>에 나타나 있다.

<표 1>에 의하면 각각의 이변량 중단시간에 대하여, λ_2의 값이 증가할수록 모든 경우에 검정력이 증가함을 알 수 있다. 또한 각 표본의 크기에 대하여, 중단시간이 증가할수록(중단비율이 낮을수록) 검정력이 증가함을 알 수 있다. 그리고 모든 경우에 표본의 크기가 증가할수록 동일성 검정법의 검정력이 증가함을 알 수 있다.

위에서 살펴본 바와 같이 중단시간이 동일한 $t_{x_i} = t_{y_i}$인 경우(일변량 중단된 자료를 갖는경우)의 확장된 경우로써, 이변량 제 1종 중단된 자료를 갖는 모형의 동일성 및 독립성 검정법이 검정력면에서 타당성이 있음을 알 수 있다. 그리고 t_{x_i}와 t_{y_i}가 각각 확률변수인 이변량 임의중도절단자료(bivariate random censored data) 또는 제 2종 중단자료(type II censored data)등 다른 형태의 중단자료를 갖는 경우로 확장할 수 있으며, 또한 Marshall-Olkin모형 뿐만 아니라 다른 종류의 이변량 지수모형에 대해서도 적용하여 비교하는 것도 좋을것이라 생각한다.

<표 1> 동일성 검정 $H_0:\lambda_1=\lambda_2$ 대 $H_1:\lambda_1<\lambda_2$에 대한 검정력

λ_2의 값	표본의 크기	유의수준	이 변 량 중 단 시 간		
			(∞,∞)	(25.0, 17.5)	(16.5, 9.50)
0.10	30	0.01	0.5900	0.5650	0.4950
		0.05	0.6500	0.6200	0.5450
	50	0.01	0.7150	0.6900	0.6050
		0.05	0.7500	0.6950	0.6550
	100	0.01	0.8750	0.8550	0.7900
		0.05	0.9050	0.8650	0.8350
0.13	30	0.01	0.8750	0.8400	0.8100
		0.05	0.8900	0.8850	0.8250
	50	0.01	0.9250	0.9250	0.8800
		0.05	0.9450	0.9300	0.9150
	100	0.01	1.0000	1.0000	0.9950
		0.05	1.0000	1.0000	0.9950
0.16	30	0.01	0.9600	0.9500	0.9350
		0.05	0.9750	0.9550	0.9450
	50	0.01	0.9850	0.9800	0.9750
		0.05	0.9900	0.9850	0.9850
	100	0.01	1.0000	1.0000	1.0000
		0.05	1.0000	1.0000	1.0000

참고문헌

[1] Arnold, B.C. (1968), Parameter Estimation for a Multivariate Exponential Distribution, *Journal of the American Statistical Association*, 62, 848-852.

[2] Bemis, B.M., Bain, L.T. and Higgins, J.J. (1972), Estimation and Hypothesis Testing for the Parameters of a Bivariate Exponential Distribution, *Journal of the American Statistical Association*, 67, 927-929.

[3] Bhattacharyya, G.K. and Johnson, R.A. (1973), On Test of Independence in a Bivariate Exponential Distribution, *Journal of the American Statistical Association*, 68, 704-706.

[4] Friday, D.S. and Patil, G.P. (1977), A Bivariate Exponential Model with Applications to Reliability and Computer Generation of Random Variables, *The Theory and Applications of Reliability*, ed. Tsokos, C.P. abd Shimi, I.N., Academic, 527-549.

[5] Hanagal, D.D. and Kale, B.K. (1991), Large Sample Tests for Testing Symmetry and Independence in Some Bivariate Exponential Models, *Communication in Statistics-Theory and Methods*, 21(9), 2625-2643.

[6] Hanagal, D.D. and Kale, B.K. (1992), Some Inference Results in Bivariate Exponential Distributions Based on Censored Samples, *Communication in Statistics-Theory and Methods*, 21(5), 1273-1295.

[7] Marshall, A.W. and Olkin, I. (1967), A Multivariate Exponential Distribution, *Journal of the American Statistical Association*, 62, 30-44.