数理解析研究所講究録
第682巻 1989年 74-80

二変量指数分布における二段階標本抽出法について

広島大理学部　百武 弘登（Hiroto Hyakutake）

1. 序

指数分布は信頼性理論において簡単なモデルとして広く使われている。Marshall and Olkin (1967, JASA) は2つの構成要素をもつあるシステムが3つのタイプのPoissonショックにより故障が生じるという仮定により二変量指数分布を導き、生存関数を用いることにより提案した。その分布が位置母数をもつならば、二変量指数分布は次の同時生存関数をもつものとして定義できるであろう。

$$P(X>x, Y>y)=\exp\{-\lambda_1(x-\mu_1)-\lambda_2(y-\mu_2)-\lambda_0\max(x-\mu_1, y-\mu_2)\} \tag{1.1}$$

ただし $x>\mu_1$, $y>\mu_2$, $-\infty<\mu_1, \mu_2<\infty$, $\lambda_i>0$ $(i=0,1,2)$ とする。また (X,Y) は生存関数 (1.1) をもつ二変量指数分布に従う確率ベクトルである。

ここでは大きさと信頼係数を与えたときの位置母数 (μ_1, μ_2) の信頼領域の構成を考える。特に尺度母数 $(\lambda_1, \lambda_2, \lambda_0)$ が未知

のときに対しては二段階標本抽出法を提案する。二段階標本抽出法はStein (1945, AMS) が正規分布の平均の信頼区間の構成に対して導出したもので、その後 Healy (1956, AMS) により多変量正規分布の平均の同時信頼区間の構成へと拡張されている。また一変量指数分布に対する二段階標本抽出法は Ghurye (1958, AMS)により与えられている。

2. 大きさを与えた信頼領域

$\{(X_i, Y_i), i=1,2,\cdots\}$ を(1.1)をもつ二変量指数分布に従う独立な確率ベクトルの列とする。母集団から n 個の観測値をとり μ_1 と μ_2 の推定量として、それぞれ

$$X_{n,1} = \min(X_1, \cdots, X_n)$$

$$Y_{n,1} = \min(Y_1, \cdots, Y_n)$$

とする。このとき $(X_{n,1}, Y_{n,1})$ もまた位置母数 (μ_1, μ_2)、尺度母数 $(n\lambda_1, n\lambda_2, n\lambda_0)$ の二変量指数分布に従うことが容易にわかる。ここで次を満足するような標本数 n を決めたいとする。

$$P(X_{n,1}-l \leq \mu_1 \leq X_{n,1},\ Y_{n,1}-l \leq \mu_2 \leq Y_{n,1}) \geq 1-\alpha \qquad (2.1)$$

ただし $l>0$ と $\alpha\ (0<\alpha<1)$ はあらかじめ与えられているものとする。簡単のため(2.1)の左辺を $P(A_{n,1})$ とおく。(2.1)の左辺についてBonferroniの不等式を使えば

$$P(A_{n,1}) \geq 1-P(X_{n,1} > \mu_1 + l) - P(Y_{n,1} > \mu_2 + l) \qquad (2.2)$$

が得られる。二変量指数分布の周辺分布は指数分布であることから、

$$P(X_{n,1} > \mu_1 + l) = \int_{\mu_1 + l}^{\infty} n(\lambda_1 + \lambda_0) e^{-n(\lambda_1+\lambda_0)(x-\mu_1)} dx$$
$$= \int_{2n(\lambda_1+\lambda_0)l}^{\infty} g_2(t)\,dt$$

同様に

$$P(Y_{n,1} > \mu_2 + l) = \int_{2n(\lambda_2+\lambda_0)l}^{\infty} g_2(t)\,dt$$

がわかる。ただし $g_q(\cdot)$ は自由度 q の χ^2 分布の密度関数である。以上のことから

$$P(A_{n,1}) \geq 1 - 2\int_{2n(\lambda_{min}+\lambda_0)l}^{\infty} g_2(t)\,dt$$

が得られる。ただし、$\lambda_{min} = \min(\lambda_1, \lambda_2)$ である。これより(2.1)を満たすために必要とされる標本数は

$$n \geq \frac{1}{(\lambda_{min} + \lambda_0)\,l} \cdot \frac{1}{2}\chi_2^2(\alpha/2) \tag{2.3}$$

であることがわかる。ただし $\chi_2^2(\alpha/2)$ は自由度2の χ^2 分布の上側 $100(\alpha/2)\%$ 点である。標本数 n としては、(2.3)を満たす最小の整数を選べば良い。次の表は $\alpha = 0.05$ のとき l と $(\lambda_1, \lambda_2, \lambda_0)$ をいくつか与えてやったときの必要な最小標本数 n とそのときの(2.1)の左辺 $P(A_{n,1})$ を具体的に計算した数値例である。また表には示していないが(2.2)におけるようなBonferroniの不等式による影響はそれほど大きくないこともわかっている。

$(\lambda_1,\lambda_2,\lambda_0)=(.1,.1,.1)$

ℓ	n	$P(A_{n,1})$
.1	185	.9544
.2	93	.9553
.3	62	.9553
.4	47	.9569
.5	37	.9544

$(\lambda_1,\lambda_2,\lambda_0)=(.1,.2,.1)$

ℓ	n	$P(A_{n,1})$
.1	185	.9720
.2	93	.9725
.3	62	.9725
.4	47	.9737
.5	37	.9720

$(\lambda_1,\lambda_2,\lambda_0)=(.1,.5,.1)$

ℓ	n	$P(A_{n,1})$
.1	185	.9752
.2	93	.9757
.3	62	.9757
.4	47	.9767
.5	37	.9752

$(\lambda_1,\lambda_2,\lambda_0)=(.1,.1,.2)$

ℓ	n	$P(A_{n,1})$
.1	123	.9573
.2	62	.9585
.3	41	.9573
.4	31	.9585
.5	25	.9597

3. 二段階標本抽出法

位置－尺度分布族において尺度母数が未知であるなら標本数を固定（つまり一段階での標本抽出）したとき位置母数に対する大きさを与えた信頼領域は構成できないことがよく知られている。そこで(2-1)を満足するような信頼領域を構成するには少なくとも二段階の標本抽出法が必要になってくる。

まず第一段階の標本として$n_0(\geq 2)$個の観測値を(1.1)の生存関数をもつ二変量指数母集団からとってくる。ここでこのn_0個の標本をもとに次を計算する。

$$S_1=\frac{1}{n_0-1}\sum_{i=1}^{n_0}(X_i-X_{n_0,1}) \quad , \quad S_2=\frac{1}{n_0-1}\sum_{i=1}^{n_0}(Y_i-Y_{n_0,1})$$

このとき次のようにNを定義する。

$$N = \max\{ n_0,\ [cS_1]+1\ ,[cS_2]+1 \} \qquad (3.1)$$

ただし$[b]$はbを越えない最大整数でcは与えられた正定数である。cについては後で述べる。次に第二段階の標本として新しく$N-n_0$個の観測値を同じ二変量指数母集団からとってくる。ここで

$$X_{N,1} = \min(X_1, \cdots, X_{n_0}, X_{n_0+1}, \cdots, X_N)$$

$$Y_{N,1} = \min(Y_1, \cdots, Y_{n_0}, Y_{n_0+1}, \cdots, Y_N)$$

としてやると(2.2)と(3.1)から次が示せる。

$$P(A_{N,1}) \geq 1 - P(X_{N,1} > \mu_1 + l) - P(Y_{N,1} > \mu_2 + l)$$

$$= 1 - P\{N(X_{N,1} - \mu_1) > Nl\} - P\{N(Y_{N,1} - \mu_2) > Nl\}$$

$$\geq 1 - P\{N(X_{N,1} - \mu_1)/S_1 > cl\} - P\{N(Y_{N,1} - \mu_2)/S_2 > cl\}$$

また$N(X_{N,1}-\mu_1)/S_1$と$N(Y_{N,1}-\mu_2)/S_2$は自由度$(2, 2(n_0-1))$のF分布に従っていることから

$$P(A_{N,1}) \geq 1 - 2\int_{cl}^{\infty} f_{2,2(n_0-1)}(t)\,dt$$

が得られる。ただし$f_{2,2(n_0-1)}(\cdot)$は自由度$(2, 2(n_0-1))$のF分布の密度関数である。このF分布の上側$100(\alpha/2)$%点を$F_{2,2(n_0-1)}(\alpha/2)$とおいてやると、(3.1)において定数cは

$$c \geq \frac{1}{l} F_{2,2(n_0-1)}(\alpha/2) \qquad (3.2)$$

を満たすようにしてやれば(2.1)を満足する大きさを与えたときの信頼領域が構成できることがわかる。

上で述べた二段階標本抽出法において $c=F_{2,2(n_0-1)}(\alpha/2)/\ell$ とおき、第一段階の標本数 n_0 が次を満たすようにしてやると、

$$\lim_{\ell\to 0} n_0 = \infty \quad , \quad \lim_{\ell\to 0} (\ell n_0) = 0$$

このとき、

$$\lim_{\ell\to 0} S_i = \frac{1}{\lambda_i+\lambda_0} \quad \text{a.s.} \quad (i=1,2)$$

がわかる。これより、

$$\lim_{\ell\to 0} \max(S_1, S_2) = \frac{1}{\lambda_{min}+\lambda_0} \quad \text{a.s.}$$

が得られる。また(3.1)から

$$\max(cS_1, cS_2) \leq N \leq \max(cS_1, cS_2) + n_0$$

が成り立つので、2節の $(\lambda_1, \lambda_2, \lambda_0)$ が既知の場合の標本数を

$$n = \frac{1}{(\lambda_{min}+\lambda_0)\ell}\cdot\frac{1}{2}\chi_2^2(\alpha/2)$$

としてやれば、

$$\lim_{\ell\to 0} \frac{N}{n} = 1 \quad \text{a.s.}$$

$$\lim_{\ell\to 0} \frac{E(N)}{n} = 1$$

が得られる。この性質を二段階標本抽出法が2節の既知の場合の方法に対して漸近有効と呼ぶ。

参考文献

Ghurye, S.G. (1958). Note on sufficient statistics and two-stage procedures. *Ann. Math. Statist.*, 29, 155-166.

Healy, W. C. Jr. (1956). Two-sample procedures in simultaneous estimation. *Ann. Math. Statist.*, 27, 687-702.

Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution. *J. Amer. Statist. Asoc.*, 62, 30-44.

Stein, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.*, 16, 243-258.