Bull. Korean Math. Soc.
Vol. 10 (1973), pp. 69-75

最適多重層化二段抽出法에 관하여

鄭 漢 永

1. 序 言

1968年에 실시한 國富統計調査에 있어서 法人 및 個人企業部門에서는 특히 많은 問題點을 내포하고 있으므로 그 標本設計에서도 基礎的, 理論的 再檢討를 필요로 하게 되었다. 本論文에서는 현재 經濟統計調査에서 사용되는 地域別, 階級別 層化無作爲二段抽出法에 最適抽出比를 적용하는 등 많은 改善點을 提示하고자 한다.

2. 一般的, 理論的觀點

(i) 母集團의 構造, 標本의 構成

母集團은 s개의 地域階層과 t개의 階級에 의하여 아래 표와 같이 二重으로 層化되어 있다고 본다. 그 各層 A_{hk}에 있어서 h는 이 層이 屬하는 地域特性符號를, k는 이 層이 屬하는 階級特性符號를 나타내는 것으로 한다. 이때 A_{hk}層에 屬하는 第二次抽出單位(個體)의 數를 N_{hk}, 第一次抽出單位(地點)의 數를 M_h라 한다. 또 各層의 가로, 세로의 累計는 끝欄에 준 記號로 나타내기로 한다.

A_{11} (N_{11})	A_{12} (N_{12})	……	A_{1t} (N_{1t})	$A_{1\cdot}$ ($N_{1\cdot}; M_1$)
A_{21} (N_{21})	A_{22} (N_{22})	……	A_{2t} (N_{2t})	$A_{2\cdot}$ ($N_{2\cdot}; M_2$)
	……	……	……	
A_{s1} (N_{s1})	A_{s2} (N_{s2})	……	A_{st} (N_{st})	$A_{s\cdot}$ ($N_{s\cdot}; M_s$)
$A_{\cdot 1}$ (N_1)	$A_{\cdot 2}$ (N_2)	……	$A_{\cdot t}$ (N_t)	$N; M$

母集團의 構造

위의 母集團構造에 의하여 各 $A_{h\cdot}$ 層에서 m_h 개의 第一次標本이, 또 A_{hk} 層에서 n_{hk} 개의 第二次標本이 얻어지는 것으로 생각하고 $\sum_{h=1}^{s} n_{hk}=n_k$, $\sum_{h=1}^{s} m_h=m$, $\sum_{k=1}^{t} n_k=n$ 로 나타내기로 한다.

注) ( )내는 第一次 및 第二次抽出單位의 數를 나타낸다. $A_{h\cdot}$ 는 $\frac{1}{p_h}=\frac{m_h}{M_h}$ 인 地點抽出(第一次抽出)確率이, 또 A_{hk} 는 $\frac{1}{q_{hk}}=\frac{n_{hk}}{N_{hk}}$ 인 第二次抽出確率이 주어져있다고 생각한다.

(ii) 推計方式

(i)의 標本에 의한 母集團의 階級別合計 x_k, 總合計 x 의 推定量 및 A_{hk} 에 屬하는 標本合計를 각각 X_k, X 및 X_{hk} 로 나타내면 單純推計方式에 의하여

(1) $$X_k=\sum_{h=1}^{s} p_h q_{hk} X_{hk}, \qquad k=1, 2, \cdots, t$$

(2) $$X=\sum_{h=1}^{s}\sum_{k=1}^{t} p_h q_{hk} X_{hk}$$

이다.

$A_{k\cdot}$ 가운데의 第 j 地點에서의 k 階級에 屬하는 個體의 分散을 σ_{hkj}^2, A_{hk} 層의 地點間分散을 $\sigma_{bh\cdot}^2$ 로 나타내면 X 의 分散은

(3) $$\sigma_X{}^2 \fallingdotseq \sum_{h=1}^{s}\sum_{k=1}^{t} p_h(q_{hk}-1)\alpha_{hk}+\sum_{h=1}^{s}(p_h-1)\beta_{h\cdot}$$

$$\sum_{j=1}^{M_h} N_{hkj}\sigma_{hkj}^2=\alpha_{hk}, \qquad M_h\sigma_{bh\cdot}^2=\beta_{h\cdot}$$

로 나타낼수 있다. 여기서 N_{hkj} 는 A_{hk} 層내의 第 j 地點에 屬하는 個體數를 나타낸다.

(iii) 標本設計상의 一般的制約條件

標本設計상의 制約條件으로는 標本地點의 制約

$$\sum_{h=1}^{s} m_h=m$$

와 個體標本數의 制約

$$E\left(\sum_{h=1}^{s}\sum_{k=1}^{t} n_{hk}\right)=\sum_{h=1}^{s}\sum_{k=1}^{t}\frac{N_{hk}}{p_h q_{hk}}=n$$

을 聯關시켜 주는 制約條件 중에서 가장 一般性을 잃지 않는 것으로 생각되는 單純費用函數

(4) $$\sum_{h=1}^{s} C_{h1}\frac{M_h}{p_h}+\sum_{h=1}^{s}\sum_{k=1}^{t} C_{h2}\frac{N_{hk}}{p_h q_{hk}}=C$$

단, C_{h1} 은 $A_{h\cdot}$層의 一標本地點當費用
C_{h2} 는 $A_{h\cdot}$層의 單位副次標本當費用
C 는 總豫算額

를 택하기로 한다.

(iv) 最適抽出比와 그의 決定方程式

(4)의 制約條件下에서 (3)을 最小로 하는 最適抽出比

$$(p_h^{(0)},\ q_{hk}^{(0)}\ ;\ h=1,2,\cdots,s,\ \ k=1,2,\cdots,t)$$

의 決定方程式은

(5) $$L=\sigma_x^2+\lambda\left(\sum_{h=1}^{s}C_{h1}\frac{M_h}{p_h}+\sum_{h=1}^{s}\sum_{k=1}^{t}C_{h2}\frac{N_{hk}}{p_hq_{hk}}-C\right)$$

를 써서 나타낼 수 있다. 즉

(6) $$\frac{\partial L}{\partial p_h}=\sum_{k=1}^{t}(q_{hk}-1)\alpha_{hk}+\beta_{h\cdot}-\frac{\lambda}{p_h^2}\left\{C_{h1}M_h+\sum_{k=1}^{t}C_{h2}\frac{N_{hk}}{q_{hk}}\right\}=0,\qquad h=1,2,\cdots\cdots s$$

(6′) $$\frac{\partial L}{\partial q_{hk}}=p_h\alpha_{hk}-\lambda\frac{C_{h2}N_{hk}}{p_hq_{hk}^2}=0,\qquad h=1,2,\cdots\cdots s,\ \ k=1,2,\cdots\cdots t,$$

따라서 (6′)에서

(7) $$p_h^2=\frac{\lambda C_{h2}N_{hk}}{q_{hk}^2\alpha_{hk}},\qquad h=1,2,\cdots\cdots,s$$

를 얻는다. 또한 (6) 및 (6′)에서

(7′) $$p_h^2=\frac{\lambda C_{h1}M_h}{\beta_{h\cdot}-\sum_{k=1}^{t}\alpha_{hk}},\qquad h=1,2,\cdots\cdots,s$$

를 얻는다.

따라서 (7), (7′)에 의하여

(8) $$q_{hk}^{(0)}=\sqrt{\frac{C_{h2}N_{hk}}{C_{h1}M_h}}\sqrt{\frac{\beta_{h\cdot}-\sum_{k=1}^{t}\alpha_{hk}}{\alpha_{hk}}},\qquad h=1,2,\cdots\cdots,s,\ \ k=1,2,\cdots\cdots,t$$

가 성립한다. 또 (7), (7′)를 (4)에 代入하면

(9) $$\sqrt{\lambda}=\frac{1}{C}\left\{\sum_{h=1}^{s}\sqrt{C_{h1}M_h\left(\beta_{h\cdot}-\sum_{k=1}^{t}\alpha_{hk}\right)}+\sum_{h=1}^{s}\sum_{k=1}^{t}\sqrt{C_{h2}N_{hk}\alpha_{hk}}\right\}.$$

따라서

(10) $$p_h^{(0)}=\frac{1}{C}\sqrt{\frac{C_{h1}M_h}{\beta_{h\cdot}-\sum_{k=1}^{t}\alpha_{hk}}}\left\{\sum_{h=1}^{s}\sqrt{C_{h1}M_h\left(\beta_{h\cdot}-\sum_{k=1}^{t}\alpha_{hk}\right)}+\sum_{h=1}^{s}\sum_{k=1}^{t}\sqrt{C_{h2}N_{hk}\alpha_{hk}}\right\},\qquad h=1,2,\cdots\cdots,s$$

를 얻는다.

3. 標本企劃에 대한 實踐上의 問題

標本企劃를 실시한 경우 그 각 段階에서 일어나는 구체적인 여러 制約을 검토하여 본다.

(i) 抽出作業상의 制約

抽出作業에 있어서 각層 A_{hk} 마다 다른 抽出比 p_h, q_{hk} 를 사용하는 것이 쉬울것 같지만 list 의 質과 量에 따라서 뜻밖의 混亂을 초래하기 쉬우므로 p_h 와

q_{hk} 가 量的으로 獨立, 즉

(11) $$q_{1k}=q_{2k}=\cdots\cdots=q_{sk}=q_k, \qquad k=1,2,\cdots\cdots,t$$

라 하면 (4), (6), (6′)는 각각

(12) $$\sum_{h=1}^{s}C_{h1}\frac{M_h}{p_h}+\sum_{h=1}^{s}\sum_{k=1}^{t}C_{h2}\frac{N_{hk}}{p_hq_k}=C,$$

(13) $$\sum_{k=1}^{t}(q_k-1)\alpha_{hk}+\beta_{h\cdot}-\frac{\lambda}{{p_h}^2}\left\{C_{h1}M_h+\sum_{k=1}^{t}C_{h2}\frac{N_{hk}}{q_k}\right\}=0,$$
$$h=1,2,\cdots\cdots,s,$$

(13′) $$\sum_{h=1}^{s}\alpha_{hk}p_h-\frac{\lambda}{{q_k}^2}\left\{\sum_{h=1}^{s}C_{h2}\frac{N_{hk}}{p_h}\right\}=0,$$
$$k=1,2,\cdots\cdots,t$$

로 되어 $p_h^{(0)}$, $q_k^{(0)}$, $h=1,2,\cdots\cdots,s$, $k=1,2,\cdots\cdots,t$,를 구한다는 것은 일반적으로 困難하다.

$p_h^{(0)}$, $q_k^{(0)}$의 算出을 용이하게 하여주는 條件을 열거하여 본다.

(a) 母集團 size matrix $N=(N_{hk})$의 行數 또는 列數의 制限.

$t=1$로 하면 $q_{hk}=q_h$로 되기 때문에 그 解法은 困難하다. 한편, $s=1$로 하고 地域符號 k를 생략하면 (12), (13), (13′)의 解는 아래와 같다.

(14) $$p^{(0)}=\frac{C_1M}{C}+\frac{\sqrt{C_1M}}{C}\,\frac{\sum_{k=1}^{t}\sqrt{C_2N_k\alpha_k}}{\beta_{\cdot}-\sum_{k=1}^{t}\alpha_k},$$

(15) $$q_k^{(0)}=\sqrt{\frac{C_2N_k(\beta_{\cdot}-\sum_{k=1}^{t}\alpha_k)}{C_1M\alpha_k}},$$

(16) $$\lambda=\frac{\{\sqrt{C_1M(\beta_{\cdot}-\sum_{k=1}^{t}\alpha_k)}+\sum_{k=1}^{t}\sqrt{C_2N_k\alpha_k}\}^2}{C^2}$$

(b) p, q의 制限.

(17) $$p_1=p_2=\cdots\cdots\cdots\cdots=p_s=p$$

라 하면 (13′)에 의하여

(18) $$p^2=\frac{\lambda\sum_{h=1}^{s}\sum_{k=1}^{t}C_{h2}N_{hk}}{{q_k}^2\sum_{h=1}^{s}\sum_{k=1}^{t}\alpha_{hk}}$$

를 얻는다. 한편 $\sum_{h=1}^{s}(13)\times p-\sum_{k=1}^{t}(13')\times q_k$에서

(19) $$p=\sqrt{\frac{\lambda\sum_{h=1}^{s}C_{h1}M_h}{\sum_{h=1}^{s}(\beta_{h\cdot}-\sum_{k=1}^{t}\alpha_{hk})}}$$

가 되며, (18)과 比較하면

(20) $$q_k=\sqrt{\frac{\sum_{h=1}^{s}\sum_{k=1}^{t}C_{h2}N_{hk}}{\sum_{h=1}^{s}C_{h1}M_h}}\;\sqrt{\frac{\sum_{h=1}^{s}(\beta_{h\cdot}-\sum_{k=1}^{t}\alpha_{hk})}{\sum_{h=1}^{s}\sum_{k=1}^{t}\alpha_{hk}}}$$

가 성립한다. 또 (12)에 (18), (19)를 代入하여

(21) $$\lambda=\frac{1}{C^2}\left[\sqrt{\{\sum_{h=1}^{s}C_{h1}M_h\}\{\sum_{h=1}^{s}(\beta_{h\cdot}-\sum_{k=1}^{t}\alpha_{hk})\}}\right.$$

$$+\sqrt{\{\sum_{h=1}^{s}\sum_{k=1}^{t}C_{h_2}N_{hk}\}\{\sum_{h=1}^{s}\sum_{k=1}^{t}\alpha_{hk}\}}\,]^2$$

를 얻는다. (19), (20), (21)은 (14), (15), (16)의 擴張이다.

(c) 費用函數의 改良.

(12) 보다 복잡한 費用函數를 써도 (b)를 보다 一般化할 수는 없다. 逆으로 (12) 보다 單純化하여

$$C_{11}=C_{21}=\cdots\cdots\cdots\cdots=C_{s1}=C_1$$

또는

$$C_{12}=C_{22}=\cdots\cdots\cdots\cdots=C_{s2}=C_2$$

로 하여도 같은 결과를 얻는다.

(d) 母集團 size matrix N의 特異化.

N을 對角行列이라 생각할 때에는 제 2절에서 취급한 일반적인 層化二段抽出法에 준하여 最適値를 算出할 수 있다.

이 때 (12), (13), (13′)는

(22) $$\sum_{h=1}^{s}C_{h1}\frac{M_h}{p_h}+\sum_{h=1}^{s}\frac{C_{h2}}{p_h}\ \frac{N_{hh}}{q_h}=C,$$

(23) $$(q_h-1)\alpha_{hh}+\beta_{h\cdot}-\frac{\lambda}{p_h^2}\left\{C_{h_1}M_h+C_{h_2}\frac{N_{hh}}{q_h}\right\}=0,\qquad h=1,2,\cdots\cdots,s$$

(23′) $$\alpha_{hh}p_h-\lambda\frac{C_{h_2}}{q_h^2}\ \frac{N_{hh}}{p_h}=0,\qquad h=1,2,\cdots\cdots,s$$

으로 變形되며, 解는

(24) $$p_h^{(0)}=\sqrt{\frac{\lambda C_{h_1}M_h}{\beta_{h\cdot}-\alpha_{hh}}},$$

(25) $$q_h^{(0)}=\sqrt{\frac{C_{h2}N_{hh}(\beta_{h\cdot}-\alpha_{hh})}{C_{h_1}M_h\alpha_{hh}}},$$

(26) $$\lambda=\frac{1}{C}\sum_{h=1}^{s}\{\sqrt{C_{h_1}M_h(\beta_{h\cdot}-\alpha_{hh})}+\sqrt{C_{h_2}N_{hh}\alpha_{hh}}\}$$

이다.

(ii) 集計面에서의 制約條件

이제 (17) 대신에

(27) $$p_1q_{1k}=p_2q_{2k}=\cdots\cdots=p_sq_{sk}=r_k,\qquad k=1,2,\cdots\cdots,t$$

로 하면 階級別로 單純集計가 가능하며 集計面에서 다소 簡約化할 수 있다.

이 때 (3), (4)는 각각

(28) $$\sigma_x^2\fallingdotseq\sum_{k=1}^{t}(\sum_{h=1}^{s}\alpha_{hk})r_k+\sum_{h=1}^{s}\{(\beta_{h\cdot}-\sum_{k=1}^{t}\alpha_{hk})\}p_h-\sum_{h=1}^{s}\beta_{h\cdot},$$

(29) $$\sum_{h=1}^{s}\frac{C_{h_1}M_h}{p_h}+\sum_{k=1}^{t}\frac{\sum_{h=1}^{s}C_{h_2}N_{hk}}{r_k}=C$$

로 되며

(30) $$p_h{}^{(0)}=\sqrt{\frac{\lambda C_{h_1}M_h}{\beta_{h\cdot}-\sum_{k=1}^{t}\alpha_{hk}}}$$

(31) $$r_k{}^{(0)}=\sqrt{\frac{\lambda\sum_{h=1}^{s}C_{h_2}N_{hk}}{\sum_{h=1}^{s}\alpha_{hk}}}$$

(32) $$q_{hk}{}^{(0)}=\sqrt{\frac{\sum_{k=1}^{s}C_{h_2}N_{hk}}{C_{h_1}M_h}}\sqrt{\frac{\beta_{h\cdot}-\sum_{k=1}^{t}\alpha_{hk}}{\sum_{h=1}^{s}\alpha_{hk}}}$$

(33) $$\sqrt{\lambda}=\frac{1}{C}\left[\sum_{h=1}^{s}\sqrt{(\beta_{h\cdot}-\sum_{k=1}^{t}\alpha_{hk})C_{h_1}M_h}+\sum_{k=1}^{t}\sqrt{(\sum_{h=1}^{s}\alpha_{hk})(\sum_{h=1}^{s}C_{h_2}N_{hk})}\right]$$

를 얻는다.

즉 이 방법으로는 항상 最適抽出比의 算出이 가능하며 (8), (9), (10)과 비교할 때 $\sqrt{\lambda}$의 算出 및 $p_h^{(0)}$의 算出이 훨신 有利함을 알 수 있다. 抽出作業面에 대해서도 有利하게 作用하고 있다.

(iii) 實査에 관한 制約條件

이 制約條件은 豫算條件 (4)와 함께 중요하다고 생각되며, 특히 <u>單位第一次標本內의 第二次標本數가 대체로 均一하게 유지됨</u>이 필요하다.

이 條件을 (17) 또는 (27)와 같이 單純하게 formulate 하기에는 너무나 복잡하기 때문에 다음과 같이 사용하는데 지장이 없을 정도로 各層의 基準을 구체화하기로 한다.

우선 條件으로서

(a) 階級層化는 x와 相關度가 높은 數量 y의 大小에 따르는 數量階級이라 하고, 그 階級分點을 $y_1>y_2>\cdots\cdots>y_{t-1}$이라 하면 $A_{\cdot1}$, $A_{\cdot2}$, $\cdots\cdots$, $A_{\cdot t}$는 각각

(34) $$y\geqq y_1,\quad y_1>y\geqq y_2,\cdots\cdots,\quad y_{t-1}>y$$

에 의하여 결정된다.

(b) 各單位地點에 屬하는 個體數는 대체로 一定하다고 본다. 이에 對應하는 標本設計方針은 다음 두가지 경우 (c), (d)로 나누어 생각할 수 있다.

(c) 地域階層 $A_{i\cdot}$는 $A_{\cdot j}$; $1\leqq j\leqq i-1$에 屬하는 個體를 포함하지 않고 $A_{\cdot i}$에 屬하는 個體를 적어도 하나 包含하도록 構成하는 것이 좋다. 따라서 이때 $s=t$가 성립하며, 이 層化基準에 의하면 母集團 size matrix N은

(35) $$N=\begin{pmatrix} N_{11} & N_{12} & \cdots\cdots & N_{1t} \\ 0 & N_{22} & \cdots\cdots & N_{2t} \\ & \cdots\cdots\cdots & & \\ 0 & \cdots\cdots & 0 & N_{tt} \end{pmatrix}$$

로 표시되는 上三角行列이 된다.

(d) 以上의 各層에 대한 標本의 割當에 (27)를 만족시키는 條件下에서의 最適割當法을 적용하면, 결국 上位地域層 (h의 값이 적은 경우)에 屬하는 上位

階級 (k의 값이 적은 경우)標本의 增大는, 同地域階層의 下位階級標本의 減少에 의하여 相殺되어 各標本地點內의 標本個體數가 均一化된다.

以上은 구체적인 調査區의 data의 解析에 의한, 충분한 檢討를 필요로 한다. 또 (b)의 條件은 일반으로 보장되지 않으므로 地域을 個體數에 따라서 미리 層化할 必要性이 예상된다.

그러나 위의 (a), (b), (c), (d)에 보인 層化 및 抽出方針이 (i), (ii), (iii)의 條件에 적합하며 또한 구체적인 設計問題에 있어서 가장 情報가 풍부하며 적절하다고 結論내릴 수 있다고 생각된다.

參 考 文 獻

〔1〕 Hansen, Hurwitz and Madow; *Sample survey methods and theory*, *I*, *II*, Wiley & Sons, Inc, New York, 1953.

〔2〕 林知己夫, サンプリング調査はどう行うか, 東京大學出版部, 1951.

서울大學校