

# 水稻褐飛蟲田間密度調查經濟 取樣大小之研究

彭雲明 林燦隆 陳秋男

ECONOMIC SAMPLE SIZE ON THE POPULATION  
DENSITY ESTIMATION OF RICE  
BROWN PLANTHOPPER

YUN-MING PONG, TSAN-LONG LIN and CHIOU-NAN CHEN

抽印自國立臺灣大學農學院研究報告第二十一卷第一號

中華民國七十年四月

Reprinted from Memoirs of the College of Agriculture

National Taiwan University

Taipei, Taiwan, China.

Vol. 21 No. 1, April 1981

# 水稻褐飛蝨田間密度調查經濟 取樣大小之研究<sup>1</sup>

彭雲明 林燦隆 陳秋男<sup>2</sup>

## 摘 要

本省水稻褐飛蝨之分佈雖大多為集中型，但每叢水稻蟲數之分佈極為複雜，負二項分佈難以有效配合，為符合 Stein 二階取樣所需之常態分佈前提，對褐飛蝨觀測資料做  $\log(x+0.1)$ ,  $\sinh^{-1}\sqrt{x}$ ,  $r+\delta \log\left(\frac{x-\xi}{\lambda}\right)$ ,  $r+\delta \log\left(\frac{x-\xi}{\lambda+\xi-x}\right)$ ,  $r+\delta \sinh^{-1}\left(\frac{x-\xi}{\lambda}\right)$  五種轉換，轉換後的資料其峯度歪度值雖有降低之趨勢，但仍偏離常態。對觀測值直接做轉換，因  $\overline{h(\bar{x})}$  估計  $h(\mu)$  ( $h$  為轉換函數) 有偏差存在亦即  $Eh(\bar{x})=h(\mu)+\text{Bias}$ ，不適合應用於二階取樣。

以模擬試驗的方法研究樣品平均值  $\bar{x}$  經  $\log(\bar{x}+0.1)$ ,  $\sinh^{-1}\bar{x}$ ,  $\sinh^{-1}\sqrt{\bar{x}}$  三種函數轉換後其分佈的性質。假設一與褐飛蝨調查結果相同之無限族羣，由此族羣中抽出 100 樣品 (樣品大小分別為  $n=5, 10, 15, 20, 25, 30$  六種情形)，計算其樣品平均值  $\bar{x}_{ni}$ ,  $i=1, 2, \dots, 100$  及轉換後的樣品平均值  $h(\bar{x}_{ni})$ ,  $i=1, 2, \dots, 100$ ，分別模擬  $\bar{x}_n$  及  $h(\bar{x}_n)$  之分佈，發現  $h(\bar{x}_n)$  之變異係數較  $\bar{x}_n$  為小， $h(\bar{x}_n)$  之學生氏化序列  $\sqrt{n}[\overline{h(\bar{x}_n)}-h(\mu)]/[h'(\bar{x}_n)s]$  比  $\sqrt{n}(\bar{x}_n-\mu)/s$  更易趨近  $t$  分佈。

Stein 二階取樣法所決定的樣品大小  $n$  為隨機變值，其期望值  $E(n)$  為族羣變方  $\sigma^2$  及初期樣品大小  $n_0$  之函數。 $E(n)$  之值隨  $n_0$  值增加而遞減 (在固定的精密度下)，此趨勢在  $n_0=E(n)$  時停止。 $n_0=21$  時在指定各種不同的理想平均差異 (精密度) 下幾乎均可得到最適當  $E(n)$  值，因此本文建議初期先調查 21 叢水稻，而後再視所需之精密度決定最後樣品大小。

## 一、緒 言

1961 年以前水稻褐飛蝨 (*Nilaparvata lugens*) 在本省原為偶發性害蟲，但自 1962 年以後其為害面積增加，為害程度亦愈形嚴重<sup>(1)</sup>，為了有效的控制其在田間之密度，必須尋求經濟有效的估計法。

1945 年 Wald<sup>(18)</sup> 所發展的逐次檢定法雖主要的目的為檢定亦可用於估計，此法堪稱經濟有效。因為此法在一特定的檢定強度下 (strength of test)，其平均樣品大小 (sample size) 較一般固定樣品大小之檢定為小。此法被提出後不久即被生物學家應用於檢定生物族羣之密度，例如：Oakland<sup>(14)</sup> 將此法應用在檢定鮭魚感染條蟲囊胞之比率，Waters<sup>(19)</sup> 將其應用於檢定森林害蟲 Spruce budworm 為害的程度，Wilson (1973) 根據逐次檢定的方式發展出逐次估算的方法用來估計松樹鋸蠅的族羣密度。1977 年日本的昆蟲學家 Kuno<sup>(11)</sup>，將逐次檢定及估算的方法應用於水稻褐飛蝨田間密度之調

1. 本研究承蒙國家科學委員會補助經費，謹誌申謝。  
2. 國立臺灣大學農藝學系助教，教授，臺灣植物保護中心研究員。  
本文於民國 70 年 2 月 19 日收到。

查。上述幾位學者在利用逐次法時均假設其所研究之對象之空間分佈型 (spatial pattern) 須呈負二項分佈  $f(x) = \left(\frac{k+x-1}{x}\right) p^k q^x$ ，此外，還需要此分佈的介量 (parameter)  $k$  值必須是穩定的，因為在建立逐次檢法之前須先估計  $k$  值<sup>(7)</sup>。

雖然 Anscombe<sup>(8)</sup>認為同一種 (species) 昆蟲，不論其族羣密度如何變異，其  $k$  值經常是穩定不變的，但事實上並非絕對如此<sup>(16)</sup>。且 Lin & Chen (1979) 研究本省水稻褐飛蝨之空間分佈型發現在所分析之 800 組資料中雖大多數呈集中型分佈，但只有 350 組 (佔 44%)，呈負二項分佈，同時其  $k$  值變異很大。是故很顯然，在本省無法藉負二項分佈建立逐次估算及檢定的方法。

本文試圖以變數變換的方法使資料之分佈趨近常態，以模擬試驗的方式觀察經轉換後的樣品  $t$  值與理論  $t$  值符合之程度，探討 Stein 二階取樣檢定的可行性，並找出較適當的初次取樣大小 (initial sample size)。

## 二、資 料

本文所研究的 128 組褐飛蝨資料係由臺灣植物保護中心所調查。調查期間為 1975, 1976 兩年，自 9 月中旬至 11 月中旬第二期水稻生育期間，每隔 10 天左右調查一次，共計調查 5~7 次。每次調查 1000 株左右的水稻，將其上之蟲數按性別 (雌、雄)、各發育期 (若蟲、成蟲)、形態 (長翅、短翅) 分別予以記錄，並整理成下列之形式：

蟲 數 (x)	0	1	2	3	...	n
水 稻 株 數 $f_x$	$f_0$	$f_1$	$f_2$	$f_3$	...	$f_n$

式中  $\sum_{x=0}^n f_x = N (N=1000)$ 。由於調查的樣品頗大，具樣品由來族羣之相當代表性，因此將其作為模擬試驗設立假設族羣之依據。本文曾依據 1975 年所調查的 36 組資料設立 36 個假設族羣，每個假設族羣蟲數之密集度及變方均不相同。

## 三、昆蟲資料之轉換與 Stein 二階取樣

為便於往後之討論先概略說明 Stein 之二階取樣。

### (一) Stein 二階取樣

此取樣法由在 Stein 年 1954 提出，是逐次估算法 (sequential estimation) 之特別情形，其主要目的在於估算而非檢定。其取樣過程分為兩個步驟。設  $x_1, x_2, \dots$  是相互獨立並呈常態分佈  $N(\mu, \sigma^2)$  之隨機變數。先從此隨機變數構成之族羣中抽取  $n_0$  個觀測值  $x_1, x_2, x_3, \dots, x_{n_0}$  計算其樣品變方  $S^2 = \left\{ \sum_{i=1}^{n_0} x_i^2 - \frac{1}{n_0} \left( \sum x_i \right)^2 \right\} / (n_0 - 1)$ ，再以下式求得所需之取樣數：

$$n = \max \left\{ \left\lceil -\frac{S^2}{Z} \right\rceil + 1, n_0 + 1 \right\} \quad (1)$$

式 (1) 中之  $Z = d^2 / t_1^2$ ， $2d$  為信賴區間之長度， $t_1$  為自由度之  $n_0 - 1$  理論分佈  $t$  值並賦與信賴區間之信賴係數。為方便計將式 (1) 改寫成式 (2)：

$$n = \max \left\{ \left[ \frac{t_0^2 S^2}{d^2} \right] + 1, n_0 + 1 \right\} \quad (2)$$

用上式求得  $n$  後，再從族羣中作第二次抽樣，抽取  $(n - n_0)$  個變值，然後用先後兩次取得之  $n$  個變值求平均，作為族羣平均  $\mu$  之估值。

### (二) 昆蟲資料之常態性轉換函數

在(一)中曾述及二階取樣之前提須為常態分佈，但昆蟲分佈多為集中型分佈，變方大於平均，且變方隨平均值之變異而變異，使用二階取樣法之前，須先轉換使其接近常態。

Lin 與 Chen (1979) 曾研究下述三種轉換之可行性：

$$1. x' = k^{1/2} \sinh^{-1}(kx)^{1/2(4,5)} \quad 2. x' = \log(x+0.1) \quad 3. x' = (x+a)^b$$

發現資料經此三種轉換後其歪度值  $g_1$  及峯度值  $g_2$  大都有降低之趨勢，但趨近常態的程度仍難令人滿意，因其  $|g_1|$  及  $|g_2|$  值仍分別大於 2 倍標準差；亦即歪度、峯度仍然存在，而且更不好的是有些資料經轉換後其歪度與峯度反而變大。

本文將嘗試探討 Johnson (1947) 轉換的可行性，此轉換之提出較 Bartlett (1947)<sup>(4)</sup> 為晚，其轉換形式有三種：1.  $x' = r + \delta \log \left( \frac{x - \xi}{\lambda} \right)$ ；2.  $x' = r + \delta \log \left( \frac{x - \xi}{\lambda + \xi - x} \right)$ ；3.  $x' = r + \delta \sinh^{-1} \left( \frac{x - \xi}{\lambda} \right)$ ，其中  $r, \delta, \xi, \lambda$  為轉換介量 (transformation parameter)，其值是根據  $x'$  為常態分佈逆推  $x$  之分佈，按  $x$  之分佈求出  $x$  之各級動差，而後以動差法估算而得。今以前述之 128 組資料分別做此三種轉換，為與  $x' = \log(x+0.1)$ ， $x' = \sinh^{-1} \sqrt{x}$  轉換做一比較亦同時做此二轉換，其結果列於表(1)。由表(1)知資料經此等轉換而呈常態者非常少，大部份資料經轉換後其歪度、峯度仍存在，為數甚多之資料經轉換後歪度、峯度更加顯著。可見欲借助該等函數將褐飛蝨資料轉換成常態，其可行性似乎不高，縱令轉換後之資料呈常態，在二階取樣之應用上亦有其他困難存在。

表 1 5 種轉換函數在 128 組資料中之效果

Table 1. Effectiveness of the five transformation functions in the 128 sets of data

transformation function	and $\begin{cases}  g_1/SEg_1  < 1.96 \\  g_2/SEg_2  > 1.96 \end{cases}$	or $\begin{cases}  g_1/SEg_1  > 1.96 \\  g_2/SEg_2  > 1.96 \end{cases}$
$\sinh^{-1}(x)$	3 <sup>a</sup>	125 <sup>b</sup> (29 <sup>c</sup> , 37 <sup>d</sup> )
$\log(x+0.1)$	3	125 (32, 40)
$r + \delta \log \left( \frac{x - \xi}{\lambda} \right)$	13	115 (1, 15)
$r + \delta \log \left( \frac{x - \xi}{\lambda + \xi - x} \right)$	1	127 (22, 39)
$r + \delta \sinh^{-1} \left( \frac{x - \xi}{\lambda} \right)$	5	123 (18, 25)

a. Number of sets of data (out of 128 sets of transformed data) whose distribution approach normal distribution.

b. Number of sets of data (out of 128 sets of transformed data) whose distribution still fail to approach normal distribution.

c. Number of sets of data (in 'b') whose  $g_1$  value is larger than that of the untransformed.

d. Number of sets of data (in 'b') whose  $g_2$  value is larger than that of the untransformed.

(⇒) 轉換後分佈之性質及二階取樣實施之困難

設轉換後之資料 (以  $h(x)$  表示,  $h$  表轉換函數) 呈常態分佈, 其平均與變方分別以  $\mu_h, \sigma_h^2$  表示。今將  $h(x)$  在以原族羣平均值  $\mu$  為起點作泰勒級數 (Taylor series expansion) 展開如式(3), 則  $\mu_{h_1}$  及  $\sigma_h^2$  如式(4)式(5):

$$h(x) = h(\mu) + h^{(1)}(\mu)(x - \mu) + \frac{h^{(2)}(\mu)}{2!}(x - \mu)^2 + \frac{h^{(3)}(\mu)}{3!}(x - \mu)^3 + \dots + R \quad (3)$$

$$\mu_h = E[\bar{h}(x)] = E[h(x)] = h(\mu) + \frac{h^{(2)}(\mu)}{2!}\mu_2 + \frac{h^{(3)}(\mu)}{3!}\mu_3 + \dots + R' \quad (4)$$

式中  $\mu_r$  表示原族羣在平均值近旁的  $r$  次動差。

$$\begin{aligned} \sigma_h^2 = E[h(x) - \mu_h]^2 &= h'(\mu)\sigma^2 + \frac{h^{(1)}(\mu)h^{(2)}(\mu)}{2!}\mu_3 + \frac{[h^{(2)}(\mu)]^2}{(2!)^2}[\mu_4 + \mu_2^2] \\ &+ \frac{h^{(1)}(\mu)h^{(3)}(\mu)}{3!}\mu_4 + \dots + R'' \end{aligned} \quad (5)$$

由(4)知  $\bar{h}(x)$  並非  $h(\mu)$  之無偏估值, 其偏差 (如式(6)) 不但與轉換函數之各階導數有關並且與原族

$$\text{Bias} = \frac{h^{(2)}(\mu)}{2!}\sigma^2 + \frac{h^{(3)}(\mu)}{3!}\mu_3 + \frac{h^{(4)}(\mu)}{4!}\mu_4 + \dots + R' \quad (6)$$

羣之各級中心動差  $\mu_j, j=1, 2, \dots$  有關, 不易估算。是故式(7)所成之分佈為非中心  $t$  分佈 (noncentral  $t$ -distribution) 其偏差介量為  $\delta = \text{Bias}/\sigma_h$ , 自由度為  $n_0 - 1$  (6)。

$$\frac{\bar{h}(x) - h(\mu)}{S_h/\sqrt{n_0}} = \frac{\frac{\sqrt{n_0}[h(x) - \mu_h]}{\sigma_h} + \delta}{\sqrt{\bar{x}_{n_0-1}^2/(n_0 - 1)}} \sim t'(n_0 - 1, \delta) \quad (7)$$

與式(2)中之  $t_1$  分佈不符 ( $t_1$  為自由度為  $n_0 - 1$  之  $t$  分佈), 而且此取樣的目的在估算信賴區間  $\bar{x} - d < \mu < \bar{x} + d$  亦即轉換後之  $\bar{h}(x) - d_h < h(\mu) < \bar{h}(x) + d_h$ , 並非  $\bar{h}(x) - d_h < \mu_h < \bar{h}(x) + d_h$ 。因此一轉換函數除具有將資料轉換呈常態分佈之能力外還須具備  $h(\mu) = \mu_h$  之性質, 但此轉換函數不易覓得, 今嘗試探討樣品平均值之轉換。

#### 四、樣品平均值轉換 $h(\bar{x})$ 在二階取樣之可行性

(一) 隨機變數序列  $\sqrt{n}[h(\bar{x}_n) - h(\mu)]/h'(\bar{x}_n)S_n$  之極限分佈

隨機變數  $x$  若其平均值  $\mu$  與變方  $\sigma^2$  均為有限值, 由 Lindeberg-Lévy 之中央極限定理知  $Y_n = \sqrt{n}(\bar{x}_n - \mu)/\sigma$  之極限分佈為  $\lim_{n \rightarrow \infty} F_n(y) = \lim_{n \rightarrow \infty} p[\sqrt{n}(\bar{x}_n - \mu)/\sigma \leq y] = \Phi(y)$ , 在此  $\bar{x}_n$  表示樣品數為  $n$  之樣品平均,  $\Phi(y)$  為標準常態分佈之分佈函數, 亦即下式成立:

$$\frac{\sqrt{u}}{\sigma}(\bar{x}_n - \mu) \xrightarrow{\mathcal{L}} Z, Z \sim N(0, 1) \quad (8)$$

式中  $\mathcal{L}$  表「分佈收斂於」。根據統計理論 [例如參照 Bickel & Docksum (1977) (Appendix A. 14. 17)] 知若式(8)成立則式(9)亦成立, 或寫成式(10):

$$\frac{\sqrt{n}}{\sigma} [h(\bar{x}_n) - h(\mu)] \xrightarrow{D} h'(\mu)Z \quad (9)$$

$$\frac{\sqrt{n} [h(\bar{x}_n) - h(\mu)]}{h'(\mu)\sigma} \xrightarrow{D} Z \quad (10)$$

按 Khintchine 大數法則知  $\bar{x}_n \xrightarrow{P} \mu$  ( $P$  表機率收斂於)，因此  $h'(\bar{x}_n) \xrightarrow{P} h'(\mu)$  又  $S_n^2 \xrightarrow{P} \sigma^2 \Rightarrow S_n \xrightarrow{P} \sigma$ ，所以  $h'(\bar{x}_n)S_n \xrightarrow{P} h'(\mu)\sigma$  亦即式 (11) 成立：

$$\frac{h'(\bar{x}_n)S_n}{h'(\mu)\sigma} \xrightarrow{P} 1 \quad (11)$$

由式 (10) 及式 (11) 得知式 (12) 成立，亦即  $\sqrt{n} [h(\bar{x}_n) - h(\mu)] / h'(\bar{x}_n)S_n$  收斂於標準常態分佈。雖然如

$$\frac{\sqrt{n} [h(\bar{x}_n) - h(\mu)]}{\frac{h'(\mu)\sigma}{h'(\bar{x}_n)S_n}} = \frac{\sqrt{n} [h(\bar{x}_n) - h(\mu)]}{h'(\bar{x}_n)S_n} \xrightarrow{D} Z \quad (12)$$

此，但在  $n$  值較小時其分佈仍不得而知。對任意常態逢機變數  $x$  而言， $\sqrt{n}(\bar{x}_n - \mu) / S_n \xrightarrow{D} Z$  當  $n \rightarrow \infty$ ，在  $n$  值為有限值時  $\sqrt{n}(\bar{x}_n - \mu) / S_n$  呈  $t$  分佈（自由度為  $n-1$ ）。由此推測雖  $x$  不呈常態分佈，但  $\sqrt{n} [h(\bar{x}_n) - h(\mu)] / h'(\bar{x}_n)S_n$  在  $n$  值為有限值時亦有可能趨近  $t$  分佈。今以模擬試驗之方式探討此種臆測之真確性。

(二) 以模擬試驗之方式探究小樣品時 ( $n \leq 30$ ) 逢機變數  $h(\bar{x}_n)$  之分佈特性及  $\sqrt{n} [h(\bar{x}_n) - h(\mu)] / h'(\bar{x}_n)S_n$  之近似分佈

### 1. 模擬試驗：假設族羣之設立與逢機樣品之抽出。

在一、中曾述及本文所研究的資料對其所來自之族羣具有相當之代表性，因此假設一無限族羣具有與此樣品相同之機率分佈，（即  $p(X=x) = f_x / \sum f_x$ ），為此樣品所來自之族羣。今由此族羣中逢機抽出 100 組樣品，其樣品大小  $n$ 。一共有  $n=5, 10, 15, 20, 25, 30$ ，6 種情形。逢機樣品抽出之方法為：由 Monte Carlo 法產生齊一 (uniform) 之逢機數值  $Z$ ， $0 \leq Z \leq 1$ ，此處所用者為 Carnahan *et al.* (1969) 的方法，其產生式為：

$$y_{i+1} = ay_i \pmod{m}$$

$$Z_i = y_{i+1} / m$$

式中  $a=2^{10}+3$ ， $m=2^{20}$ ， $y_i=566387$ 。把  $Z_i$  值作為  $x$  分佈自小到大的累計機率看待， $i$  代表運算之次數號。 $Z_i$  值產生後，將其放入假設族羣之機率  $p_j = p_{r_o}[X=x_j]$ ， $x_j=0, 1, 2, \dots$  分佈中，若  $Z_i$  值落下列區間則令所得之逢機變值  $x_i = x_j$ 。

$$p_0 + p_1 + \dots + p_{j-1} < Z_i \leq p_0 + p_1 + \dots + p_j$$

此一逢機數  $x_i$  值相當於從假設族羣逢機取出之一觀測值，重複此過程 500 次相當於逢機取出 500 觀測值  $x_1, x_2, \dots, x_{500}$ 。將  $x_1, x_2, x_3, x_4, x_5$  五個觀測值視為第 1 個樣品，將  $x_6, x_7, x_8, x_9, x_{10}$  五個觀測值視為第 2 個樣品，……，依此方式共可產生 100 個樣品其樣品數為  $n=5$ 。樣品數  $n=10$  之 100

樣品其產生的方式與樣品數  $n=5$  時之情形相似，即逢機抽取 1000 個觀測值， $x_1, x_2, \dots, x_{1000}$ ，而後將  $x_1, x_2, \dots, x_{10}$  視為第 1 個樣品， $x_{11}, x_{12}, \dots, x_{20}$  視為第二個樣品，……，依此方式產生出 100 個樣品，其樣品數為  $n=10$ 。樣品數為  $n=15, 20, 25, 30$  時，100 個樣品產生的方式與樣品數  $n=5, 10$  時之情況類似。抽取 100 樣品的用意為：在樣品數為  $n$  時，以 100 個樣品之介值 (statistic)  $\bar{x}_n, h(\bar{x}_n), \sqrt{n}(\bar{x}_{nk}-\mu)/S, \sqrt{n}[h(\bar{x}_{nk})-h(\mu)]/[h'(\bar{x}_{nk})S] k=1, 2, \dots, 100$ ，分別模擬  $\bar{x}_n, h(\bar{x}_n), \sqrt{n}(\bar{x}_n-\mu)/S, \sqrt{n}[h(\bar{x}_n)-h(\mu)]/h'(\bar{x}_n)S$  之分佈，並研究其分佈之性質。整個模擬試驗之過程如圖 1 所表示。

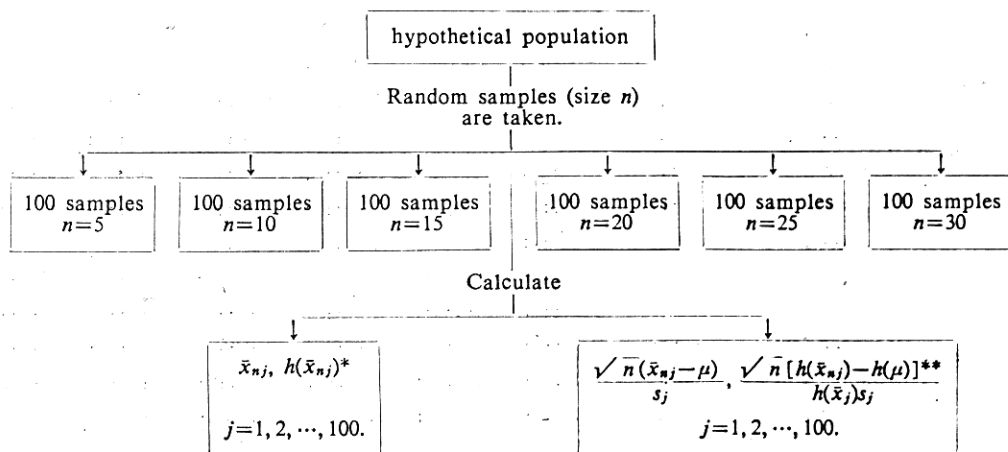


圖 1 以模擬試驗之方式研究  $\bar{x}_n, h(\bar{x}_n), \sqrt{n}(\bar{x}_n-\mu)/s, \sqrt{n}[h(\bar{x}_n)-h(\mu)]/[h'(\bar{x}_n)s]$  分佈之性質

Fig. 1. The simulation procedure by which the distribution properties of statistics  $\bar{x}_n, h(\bar{x}_n), \sqrt{n}(\bar{x}_n-\mu)/s, \sqrt{n}[h(\bar{x}_n)-h(\mu)]/[h'(\bar{x}_n)s]$  are studied.

\* The values of  $\bar{x}_{nj}$  and  $h(\bar{x}_{nj})$  are used to simulate the distributions of  $\bar{x}_n$  and  $h(\bar{x}_n)$  respectively.

\*\* The values of  $\sqrt{n}(\bar{x}_{nj}-\mu)/s_j$  and  $\sqrt{n}[h(\bar{x}_{nj})-h(\mu)]/[h'(\bar{x}_{nj})s_j]$  are used to simulate the distributions of  $\sqrt{n}(\bar{x}_n-\mu)/s$  and  $\sqrt{n}[h(\bar{x}_n)-h(\mu)]/[h'(\bar{x}_n)s]$  respectively.

為使所研究之結果較具一般性，根據 1975 年所調查之 36 組資料設立 36 個假設族羣。每一假設族羣均重複如圖 1 之模擬試驗過程。

2.  $E[h(\bar{x}_n)]$  趨近  $h(\mu)$  之程度及  $h(\bar{x}_n)$  之變異係數：

理論上  $h(\bar{x}_n)$  為  $h(\mu)$  之漸近無偏估算式 (見式 (14))

$$Eh(\bar{x}_n) = h(\mu) + \frac{h''(\mu)}{2} \cdot \frac{\sigma^2}{n} + R_n \quad (14)$$

當  $n \rightarrow \infty$  時， $E[h(\bar{x}_n)] = h(\mu)$ ，在小樣品時  $Eh(\bar{x}_n)$  趨近  $h(\mu)$  之程度除受樣品數  $n$  之影響外，還受族羣平均值  $\mu$  與  $\sigma^2$  之影響。今以 36 個假設族羣中之 5 個族羣研究不同平均  $\mu$ 、變方  $\sigma^2$  及樣品大小  $n$  對  $Eh(\bar{x}_n)$  趨近  $h(\mu)$  之程度，其結果列於表 2a, 2b。表 2a, 2b 中之五個族羣平均及變方值係由 36 個假設族羣中選出，目的在比較在不同族羣平均下之趨勢。表 2a 中之數值係以  $h(\bar{x}_{nk}) k=1, 2, \dots, 100$ ，計算其平均值  $\frac{1}{100} \sum_{k=1}^{100} h(\bar{x}_{nk})$  做為在樣品大小為  $n$  時  $Eh(\bar{x}_n)$  之近似值，以與  $h(\mu)$  值比較。表 2b 則為相對誤差  $|Eh(\bar{x}_n) - h(\mu)|/h(\mu)$  之值。係根據表 2a 之  $Eh(\bar{x}_n)$  與  $h(\mu)$  值計算而得。由表 2b 顯示  $n=5$  時，在 5 個假設族羣中，其相對誤差在 2~6% 之間，在  $n=10$  時相對誤差之範圍在 0.9~

表 2a  $Eh(\bar{x}_n)$  趨近  $h(\mu)$  之程度  
 Table 2a. The extent of  $Eh(\bar{x})^a$  to approach toward  $h(\mu)^b$

population mean, variance	transformation function	sample size $n$						$h(\mu)$
		5	10	15	20	25	30	
1 0.6292 0.985	$\log(\bar{x}+0.1)$	—	-0.3805	-0.2924	-0.3093	-0.3069	-0.2969	-0.3157
	$\sinh^{-1}\bar{x}$	—	0.5971	0.6289	0.6142	0.6121	0.6163	0.5937
	$\sinh^{-1}\sqrt{\bar{x}}$	—	0.7115	0.7394	0.7326	0.7328	0.7361	0.7274
2 5.8852 48.132	$\log(\bar{x}+0.1)$	1.7157	1.7153	1.7461	1.7363	1.7409	1.7371	1.7893
	$\sinh^{-1}\bar{x}$	2.4014	2.3998	2.4301	2.4201	2.4147	2.4208	2.4727
	$\sinh^{-1}\sqrt{\bar{x}}$	1.5885	1.5866	1.6003	1.5954	1.5975	1.5956	1.6194
3 10.4396 76.907	$\log(\bar{x}+0.1)$	2.2691	2.3347	2.3454	2.3539	2.3487	2.3468	2.3551
	$\sinh^{-1}\bar{x}$	2.9553	3.0206	3.0313	3.0398	3.0346	3.0327	3.0410
	$\sinh^{-1}\sqrt{\bar{x}}$	1.8500	1.8802	1.8850	1.8889	1.8863	1.8853	1.8891
4 14.1856 412.531	$\log(\bar{x}+0.1)$	2.4756	2.5748	2.6227	2.6600	2.6536	2.6601	2.6592
	$\sinh^{-1}\bar{x}$	3.1673	3.2617	3.3097	3.3474	3.3409	3.3474	3.3466
	$\sinh^{-1}\sqrt{\bar{x}}$	1.9525	1.9965	2.0193	2.0374	2.0341	2.0371	2.0364
5 19.3816 96.648	$\log(\bar{x}+0.1)$	2.8796	2.9248	2.9484	2.9451	2.9494	2.9479	2.9695
	$\sinh^{-1}\bar{x}$	3.5678	3.6132	3.6369	3.6336	3.6380	3.6364	3.6581
	$\sinh^{-1}\sqrt{\bar{x}}$	2.1442	2.1662	2.1778	2.1761	2.1782	2.1774	2.1880

- a. The expectations of the transformed sample means at different sample sizes.  
 b. The transformed population means of the five hypothetical populations.

表 2b 以  $h(\bar{x})$  估算  $h(\mu)$  時之相對偏差  
 Table 2b. Values of the relative bias of  $Eh(\bar{x})$

population mean, variance	transformation function	sample size $n$					
		5	10	15	20	25	30
1 0.6292 0.985	$\log(\bar{x}+0.1)$	—	0.2052	0.0738	0.0202	0.0278	0.0595
	$\sinh^{-1}\bar{x}$	—	0.0057	0.0592	0.0345	0.0309	0.0380
	$\sinh^{-1}\sqrt{\bar{x}}$	—	0.0218	0.0164	0.0071	0.0074	0.0119
2 5.8852 48.132	$\log(\bar{x}+0.1)$	0.0411	0.0413	0.0241	0.0296	0.0270	0.0291
	$\sinh^{-1}\bar{x}$	0.0288	0.0294	0.0172	0.0212	0.0194	0.0209
	$\sinh^{-1}\sqrt{\bar{x}}$	0.0190	0.0202	0.0117	0.0148	0.0135	0.0146
3 10.4396 76.907	$\log(\bar{x}+0.1)$	0.0365	0.0086	0.0041	0.0005	0.0027	0.0035
	$\sinh^{-1}\bar{x}$	0.0281	0.0067	0.0031	0.0003	0.0021	0.0027
	$\sinh^{-1}\sqrt{\bar{x}}$	0.0206	0.0047	0.0021	0.0001	0.0014	0.0020
4 14.1856 412.531	$\log(\bar{x}+0.1)$	0.0690	0.0317	0.0137	0.0003	0.0021	0.0003
	$\sinh^{-1}\bar{x}$	0.0535	0.0253	0.0110	0.0002	0.0017	0.0002
	$\sinh^{-1}\sqrt{\bar{x}}$	0.0412	0.0195	0.0083	0.0004	0.0011	0.0003
5 19.3816 96.648	$\log(\bar{x}+0.1)$	0.0302	0.0150	0.0071	0.0082	0.0067	0.0072
	$\sinh^{-1}\bar{x}$	0.0246	0.0122	0.0057	0.0066	0.0054	0.0059
	$\sinh^{-1}\sqrt{\bar{x}}$	0.0200	0.0099	0.0046	0.0054	0.0044	0.0048

4%之間(在第1假設族羣經  $\log(\bar{x}_n+0.1)$  轉換之誤差為 20.5%為唯一之例外, 由此可見  $\log(\bar{x}_n+0.1)$  之轉換不太理想),  $n=15$  時, 相對誤差範圍在 0.4~7%之間, 誤差均甚微小,  $n \geq 15$  時對每一族羣而言, 相對誤差已穩定不再有太大的變異。這些變異有一小部分可能來自以  $\frac{1}{100} \sum_{k=1}^{100} h(\bar{x}_{nk})$  做為  $Eh(\bar{x}_n)$  之近似值之逢機誤差。若剔除此種誤差, 表 2b 中之誤差值可能還要小些。表 2b 亦顯示  $\mu$  值之大小對相對誤差值有影響, 此種影響在  $n \leq 15$  較為明顯,  $n > 15$  時其影響在  $\mu=5 \sim \mu=10$  較顯著, 大致而言在小樣品時  $\mu$  (及  $\sigma^2$ ) 之影響亦不如  $n$  值之影響大。在所研究的三種轉換中, 不論  $\mu$  值之大小, 及  $\mu, \sigma^2$  值之變異(即在各種不同之族羣中) 似乎均以  $\sinh^{-1}\sqrt{\bar{x}}$  轉換為最佳, 其相對誤差最小。 $\sinh^{-1}\bar{x}_n$  轉換之誤差介於  $\log(\bar{x}_n+0.1)$  轉換與  $\sinh^{-1}\sqrt{\bar{x}}$  轉換之間。

由上述之結果知  $Eh(\bar{x}_n)$  在小樣品時趨近  $h(\mu)$  之程度甚佳, 今討論其轉換後  $h(\bar{x}_n)$  之變異係數與未轉換者  $\bar{x}_n$  之關係。 $h(\bar{x}_n)$  之變異係數  $c.v. = h'(\mu)\sigma_x/h(\mu)$  較未轉換者之  $c.v. = \sigma_x/\mu$  為小, 為比較其大小之差異將  $\bar{x}_n$  之變異係數視為 1, 此時  $h(\bar{x}_n)$  之相對變異係數為  $h'(\mu)\mu/h(\mu)$ , 是族羣平均值  $\mu$  值之函數。 $\mu$  值越大則相對變異係數即愈小, 如表 3 所示, 其變小的幅度隨  $\mu$  值之增加而漸減, 例如,  $\mu=2, \mu=3$  時  $\sinh^{-1}\sqrt{\bar{x}_n}$  之相對變異係數分別為 0.35617、0.32880, 其差為 0.04497,  $\mu=3, \mu=4$  時  $\sinh^{-1}\sqrt{\bar{x}_n}$  之相對變異係數之差為 0.02737、 $<0.04497$ 。在所研究的三種轉換函數中以  $\sinh^{-1}\sqrt{\bar{x}_n}$  之變異係數為最小  $\log(\bar{x}_n+0.1)$  轉換在  $\mu=1$  附近  $\log(\mu+0.1) \approx 0$  時使  $h'(\mu)\sigma_x/h(\mu)$  得值變得很大, 此為  $\log(\bar{x}_n+0.1)$  轉換之缺點。 $\sinh^{-1}\bar{x}_n$  之變異係數則介於  $\log(\bar{x}_n+0.1)$  與  $\sinh^{-1}\sqrt{\bar{x}_n}$  之間。綜合表 2、表 3 可知  $\bar{x}_n$  經轉換後, 其分佈會集中於  $h(\mu)$  值附近且  $h(\bar{x}_n)$  為  $h(\mu)$  之無偏估值。

表 3 在不同的族羣  $h(\bar{x})$  和  $\bar{x}$  的相對變異係數

Table 3. Relative c. v. of  $h(\bar{x})$  and  $\bar{x}$  from different populations

population mean $\mu$	relative c. v. of $\bar{x}$	relative c. v. of $h(\bar{x})$		
		$\log(\bar{x}+0.1)$	$\sinh^{-1}\bar{x}$	$\sinh^{-1}\sqrt{\bar{x}}$
0.1	1	0.31066	0.95346	0.48446
1	1	9.53823	0.80228	0.40114
2	1	1.28346	0.61956	0.35617
3	1	0.85535	0.52170	0.32880
4	1	0.69114	0.46319	0.30978
5	1	0.60175	0.42405	0.29552
10	1	0.42814	0.33188	0.25513
15	1	0.36592	0.29326	0.23462

### 3. $\bar{x}_n$ 與 $h(\bar{x}_n)$ 之分佈趨近常態分佈之比較:

先以模擬試驗之方式探討  $\bar{x}_n$  隨  $n$  值之增加而趨近常態之情形, 而後再討論其轉換後的情形。對一假設族羣而言, 按圖 1 之方式逢機取得 100 個  $\bar{x}_{nk}$  ( $k=1, 2, \dots, 100$ ) 值, 並以此模擬  $\bar{x}_n$  之分佈。根據此 100 個  $\bar{x}_{nk}$  ( $k=1, 2, \dots, 100$ ) 值計算  $g_1, g_2$  ( $g_1$ -statistic,  $g_2$ -statistic),  $SEg_1$  ( $g_1$  值之標準偏差),  $SEg_2$  ( $g_2$  值之標準偏差) 並測驗  $\bar{x}_n$  之峯度、歪度是否存在, 若  $|g_1/SEg_1|, |g_2/SEg_2|$  均小於 1.96, 則稱  $\bar{x}_n$  之分佈趨近常態, 否則認為其仍偏離常態。在  $n$  值固定時 (在本文中  $n$  值共分  $n=5, 10,$

15, 20, 25, 30 六種情形), 來自 36 個假設族羣之 36 個  $\bar{x}_n$  分佈, 按上述方式檢定其分佈是否趨近常態, 其結果記錄於表 4。由表 4 知  $n=5$  時, 來自 36 個假設族羣之 36 個  $\bar{x}$  分佈只有 3 個趨近常態分佈,  $n=10$  時有 8 個  $\bar{x}$  之分佈趨近常態,  $n$  值增加時趨近常態的個數隨之增加, 在  $n=20$  時已漸趨穩定, 此時來自 36 個假設族羣之  $\bar{x}$  分佈有 22 個趨近常態分佈, 約佔 55.5~61.1% (22/36)。

表 4 樣品平均值  $\bar{x}_n$  與轉換後的樣品平均值之常態性測驗

Table 4. Normality test for the distributions of sample means and transformed sample means\*

transformation function	sample size $n$					
	5	10	15	20	25	30
$\log(\bar{x}+0.1)$	8(22.2)**	11(30.5)	21(58.3)	21(58.3)	23(63.8)	25(69.4)
$\sinh^{-1}\bar{x}$	10(27.7)	15(41.6)	26(72.2)	28(77.7)	28(77.7)	28(77.7)
$\sinh^{-1}\sqrt{\bar{x}}$	10(27.7)	14(38.8)	24(66.6)	26(72.2)	28(77.7)	28(77.7)
$\bar{x}$	3( 8.3)	8(22.2)	11(30.5)	22(61.1)	20(55.5)	20(55.5)

\* With a given sample size,  $n$ , 36  $\bar{x}$  or  $h(\bar{x})$ -distributions are generated by simulation method, each from one hypothetical population.

\*\* Number (the integer) and percentage of the  $h(\bar{x})$ -distributions, with both  $|g_1/SEg_1|$  and  $|g_2/SEg_2|$  less than 1.96, in the 36 generated  $h(\bar{x})$ -distributions.

檢定  $h(\bar{x}_n)$  是否趨近常態的方式與檢定  $\bar{x}_n$  者相似, 對任一假設族羣而言, 將逢機取得之 100 個  $\bar{x}_n$  ( $n=1, 2, \dots, 100$ ) 值轉換成  $h(\bar{x}_{nk})$  ( $k=1, 2, \dots, 100$ ) 模擬  $h(\bar{x}_n)$  之分佈, 根據  $h(\bar{x}_{nk})$ ,  $k=1, 2, \dots, 100$  計算  $g_1, g_2, SEg_1, SEg_2$  值以檢定  $h(\bar{x}_n)$  之常態性, 所得之結果記錄於表 4。在  $n=5$  時, 來自 36 個假設族羣的 36 個  $h(\bar{x}_n)$  分佈有 8~10 個趨近常態,  $n=10$  時, 11~15 個趨近常態,  $n$  值增加則趨近常態的個數亦增加, 大約在  $n>20$  時漸趨穩定。在相同的  $n$  值下轉換後的  $h(\bar{x}_n)$  比未轉換者趨近常態的程度較佳,  $n=30$  時  $h(\bar{x}_n)$  趨近常態的比率為 77.7% 較未轉換的 55.5% 高出很多。

#### 4. $\sqrt{n}(\bar{x}_n - \mu)/S$ , 與 $\sqrt{n}[h(\bar{x}_n) - h(\mu)]/[h'(\bar{x}_n)S]$ 趨近 $t$ 分佈之比較:

由上兩部分, 分別知  $Eh(\bar{x}_n) = h(\mu)$  及  $h(\bar{x}_n)$  之分佈趨近常態的可能性很高, 由此推測  $\sqrt{n}[h(\bar{x}_n) - h(\mu)]/[h'(\bar{x}_n)S]$  趨近  $t$  分佈之可能性很大。今亦以模擬試驗之方式探討  $\sqrt{n}(\bar{x}_n - \mu)/S$ , 與  $\sqrt{n}[h(\bar{x}_n) - h(\mu)]/[h'(\bar{x}_n)S]$  趨近  $t$  分佈之差異。先討論未轉換的情形, 對一假設族羣而言, 在樣品數固定為  $n$  時, 由圖 1 知, 可得逢機數值  $\bar{x}_{nk}, S_k, k=1, 2, \dots, 100$ 。依據此 100 對  $\bar{x}_{nk}, S_k$ , 可計算出  $\sqrt{n}(\bar{x}_{nk} - \mu)/S_k, k=1, 2, \dots, 100$ , 模擬  $\sqrt{n}(\bar{x}_n - \mu)/S$  之分佈, 此 100 個樣品  $t$  值與理論  $t$  值比較, 其比較方法為: 將理論  $t$  分佈值劃分成 12 個區間, 即  $(-\infty, -t_{0.05}), (-t_{0.05}, -t_{0.14}), (-t_{0.14}, -t_{0.23}), \dots, (t_{0.14}, t_{0.05}), (t_{0.05}, \infty)$ , 計數 100 個樣品  $t$  值落於各區間的度類並與理論頻度比較, 以  $\chi^2$  測驗其適合度。若適合度測驗不顯著則稱  $\sqrt{n}(\bar{x}_n - \mu)/S$  之分佈趨近  $t$  分佈, 否則認為其偏離  $t$  分佈。當  $n$  值固定時, 來自 36 個假設族羣的 36 個  $\sqrt{n}(\bar{x}_n - \mu)/S$  分佈經  $\chi^2$  檢定其適合度之結果記錄於表 5。由表 5 知,  $n=5$  時 36 個  $\sqrt{n}(\bar{x}_n - \mu)/S$  分佈沒有一個趨近  $t$  分佈,  $n=10$  時, 趨近  $t$  分佈者僅有 1 個, 其趨近  $t$  分佈的程度隨樣品數  $n$  假之增加而增加, 但在  $n=30$  時, 趨近  $t$  分佈的比率只有 44.4% 還不到一半,  $\bar{x}_n$  經轉換後的情形, 係以  $\sqrt{n}[h(\bar{x}_{nj}) - h(\mu)]/[h'(\bar{x}_{nj})S_j], j=1, 2, \dots, 100$

表 5 適合度測驗的結果  
Table 5. The result of test of goodness-of-fit\*

transformation function	sample size $n$ :					
	5	10	15	20	25	30
$\log(\bar{x}+0.1)$	5(13.9)**	16(44.4)	23(63.9)	23(63.9)	23(63.9)	26(72.2)
$\sinh^{-1}\bar{x}$	6(16.7)	15(41.7)	20(55.6)	20(55.6)	21(58.3)	23(63.9)
$\sinh^{-1}\sqrt{\bar{x}}$	5(31.9)	15(41.7)	23(63.9)	22(61.1)	23(63.9)	25(69.4)
$\bar{x}$	0(0.0)	1(2.8)	11(30.6)	12(33.6)	11(30.6)	16(44.4)

\* With a given sample size,  $n$ ,  $36(\bar{x}-\mu)/s$  and  $\sqrt{n}[h(\bar{x})-h(\mu)]/[h'(\bar{x})s]$  distributions are tested.

\*\* Number and percentage of the distributions fitted well by  $t$ -distribution.

模擬  $\sqrt{n}[h(\bar{x}_n)-h(\bar{x}_n)S]/[h'(\bar{x}_n)S]$  之分佈， $\chi^2$  檢定其符合  $t$  分佈之結果亦記錄於表 5。由表 5 知轉換後之結果顯然比未轉換者要好，其趨近  $t$  值之比率亦隨  $n$  值之增加而增加，當  $n=15$  時已漸呈穩定，在  $n=30$  時已達 63.9~72.2%，換言之，約 70% 之假設族群其  $\sqrt{n}[h(\bar{x}_n)-h(\mu)]/[h'(\bar{x}_n)S]$  之分佈趨近理論  $t$  分佈。

經上述研究得知在小樣品時介值  $\sqrt{n}[h(\bar{x}_n)-h(\mu)]/[h'(\bar{x}_n)S]$  趨近理論  $t$  分佈之程度甚佳，且以  $h(\bar{x})$  估計  $h(\mu)$  具有無偏性之性質，因此解決了第三節所述一般觀測值直接轉換，即  $h(x)$  無法應用在二階取樣的困難。在所研究的三種轉換中，以  $\sinh^{-1}\sqrt{\bar{x}}$  轉換之變異最小，且對  $h(\mu)$  之估計誤差亦最小最有利用價值，往後所提之轉換  $h(\bar{x}_n)$  係指  $\sinh^{-1}\sqrt{\bar{x}}$  轉換而言。 $h(\bar{x})$  轉換應用於二階取樣，求樣品數之過程說明如下：先取一初期樣品，樣品大小為  $n_0$ ，求出  $\bar{x}$  及  $S^2$ ，再按式(15)決定最後樣品大小  $n$ 。

$$n = \max \left\{ \left[ \frac{th'(\bar{x})S}{d_h} \right]^2 + 1, n_0 + 1 \right\} \quad (15)$$

式中  $t$  值為自由度  $n_0-1$  及信賴度為  $1-\alpha$  的理論  $t$  值， $d_h$  則與估算所得  $h(\bar{x})$  之精密度有關。

### 五、選擇適當的初期樣品大小 $n_0$

Stein 最初提出二階取樣時，對  $n_0$  值大小的選擇並未有詳細的探討，只提到可任意選取  $n_0$  值。但任意選取  $n_0$  值並不能符合經濟的原則，因為若選擇較大的  $n_0$  值，則必使得  $h'(\bar{x})S_{\bar{x}}$  很小，對於預先固定的顯著差數  $th'(\bar{x})S_{\bar{x}} \leq d_h$  固然是滿足，但往往超過所需要，而造成不必要的浪費，若選取較小之  $n_0$  值，則  $[h'(\bar{x})S_{\bar{x}}]^2$  對  $[h'(\mu)\sigma_x]^2$  所提供之認識不夠，故往往需要再多取為數甚多之觀測值做為第二部分樣品才能滿足預先固定之精密度（見表 6）因此亦不符經濟原則。今即按  $n_0$  與  $E(n)$  之關係及  $n$  之分佈探討較適當的  $n_0$  值。 $n_0$  與  $E(n)$  之關係式如下：

$$E(n) = (n_0 + 1)p\{x_{n_0+1}^2 < y\} + \frac{\sigma^2}{z} p\{x_{n_0+1}^2 > y\} + \theta \quad (16)$$

式中  $0 < \theta < 1$ ,  $y = (n_0^2 - 1)z/\sigma^2$ ,  $z = d^2/t^2$ 。 $E(n)$  為  $n_0$  及  $\sigma^2$  之函數，然  $\sigma^2$  為一族羣的特徵值無法改變，可改變者僅為  $n_0$  值。最適當的  $n_0$  值是使得  $E(n)$  值為最小的，1953 年 Seelbinder 按式(17)

表 6 在第 I 種錯誤機率  $\alpha=0.05$  時不同的初期樣品大小  $n_0$ ，及預定之  $c$  下  $E(n)$  與  $n_{0.95}$  值表  
 Table 6. Values of  $E(n)$  and  $n_{0.95}$ \* under different initial sample sizes and different  $c$  values when error size  $\alpha=0.05$

$n_0-1$	$c=d_h/h'(\mu)\sigma$						
	0.1	0.2	0.3	0.4	0.5	0.6	0.7
240	388( 455)	241(241)					
120	392( 496)	121(121)					
80	396( 526)	101(132)	81( 81)				
60	400( 528)	100(132)	61( 61)	61(61)			
50	403( 546)	101(137)	53( 61)	51(51)			
40	408( 570)	102(143)	48( 64)	42(40)	41(41)		
30	417( 609)	104(152)	47( 68)	32(38)	31(31)		
20	435( 684)	109(171)	49( 76)	29(43)	22(28)	21(21)	
10	496( 909)	124(228)	56(101)	32(57)	21(37)	16(26)	13(19)
5	661(1468)	165(367)	74(164)	42(49)	47(59)	19(41)	14(30)

\* Values of  $E(n)$  and  $n_{0.95}$  are the numbers on the left of ( ) and in the ( ), respectively.

(式(16)之變型)，計算在不同  $n_0$  值下  $E(n)$  的值 ( $n_0$  值的範圍從 5 到 240)，列成表 (表 6)。

$$E(n) = n_0 \int_0^{x_{n_0}^2} f(x^2, n_0-1) dx^2 + \frac{t^2}{(n_0-1)c^2} \int_{x_{n_0}^2}^{\infty} x^2 f(x^2, n_0-1) dx^2 \quad (17)$$

式中  $c = \frac{dh}{h'(\mu)\sigma}$ 。藉表 6 可查出  $c$  值固定時，使得  $E(n)$  值最小的  $n_0$  值。

因  $n$  值為逢機值，雖從表 6 可查出導致最小  $E(n)$  值之  $n_0$  值，若  $n$  值之變異很大時，即使選取該  $n_0$  值，式(15)所求出之  $n$  值偏離  $E(n)$  值之可能性很大。所以瞭解  $n$  之分佈情形對  $n_0$  值之選擇亦不無幫助。 $n$  之變異雖然一般通常用  $\text{Var}(n)$  表示，但由於  $\text{Var}(n)$  與  $n$  分佈之百分位 (percentile) 無密切之關係，Moshman (1959) 認為以  $n_{0.95}$  (95th percentage point) 表示其變異較以  $\text{Var}(n)$  清楚。 $n_{0.95}$  之計算式如下：

$$n_{0.95} = \max \left\{ n_0, \frac{1}{\lambda} \left[ 95\text{th percentage point of } \frac{\chi^2(n)}{n} \right] + 1 \right\} \quad (18)$$

依據此式計算  $n_{0.95}$  值附於表 6  $E(n)$  值傍，可輔助選擇  $n_0$  值。今舉例說明之，在表 6 中若固定精密度  $c=0.3$  時，依據最小的  $E(n)$  值判斷應選擇  $n_0=31$ ，若為避免極端大值出現，可選擇  $n_0=41$ 。因為  $n_0=41$  時  $n_{0.95}=64$ ，較  $n_{0.95}=68$  ( $n_0=30$  時) 為小。 $n_{0.95}$  值之應用僅為輔助而已，主要的還是需要以最小的  $E(n)$  值來決定  $n_0$  值。

表 6 對  $n_0$  值的選擇提供了重要的理論根據，在  $c$  值 ( $c=d_h/h'(\mu)\sigma$ ) 為 0.2、0.3、0.4、0.5、0.6 時，其最適當的初期樣品大小  $n_0$  分別為：61、31、21、11、11，因為在此等初期樣品大小下可得最適當之  $E(n)$  值：100、47、29、21、16。今若固定初期樣品大小為 21，則在此條件下  $E(n)$  值為：109、49、29、22、21，與前述最適當之  $E(n)$  值相比較，差距不多，亦即，若不論  $c$  值為何，取初期樣品大小  $n_0=21$ ，幾乎可得最適當之  $E(n)$  值，故本文建議初期樣品大小宜取為  $n_0=21$ 。

## 六、應 用 例

今擬估計褐飛蝨若蟲之田間族羣密度  $\mu$ ，並檢定  $H_0: \mu=20$ ,  $H_A: \mu \neq 20$ 。按上節所述最適當的初期樣品大小為  $n_0=21$ ，故先逢機抽取 21 個觀測值，即先逢機調查 21 株水稻並記錄其上之若蟲數如下：

15 12 21 37 28 13 27 11 11 11 23 18 13 9 35 10 8 15 12 12 12

根據此 21 個觀測值計算  $S^2=71.96$ ，令  $\frac{dh}{h'(\bar{x})}=3$ ，則由式 (15) 可算出最後樣品總數  $n$ 。

$$n = \left[ t \cdot \frac{h'(\bar{x})}{d_h} \cdot S \right]^2 + 1 = 35 \quad (t = t_{20, 0.05} = 2.086, h(\cdot) = \sinh^{-1} \sqrt{\cdot})$$

因 35 較 21 為大，需再抽取 14 個觀測值，亦即再逢機調查 14 株水稻並記錄其蟲數：

3 10 13 20 42 8 28 25 31 13 0 24 12 16

由此 35 個觀測值可計算出  $\bar{x}=17.08$ ,  $h(\bar{x})=2.1122685$ ,  $h'(\bar{x})=0.028443668$ ,  $d_h=3 \cdot h'(\bar{x})=0.08533$ 。由此等數值知：

$$p\{h(\bar{x}) - d_h < h(\mu) < h(\bar{x}) + d_h\} = p\{2.0269 < h(\mu) < 2.1976\} = 0.95 \quad (19)$$

式 (19) 相當於  $p\{h^{-1}(2.0269) < \mu < h^{-1}(2.1976)\} = p\{13.91 < \mu < 19.76\} = 0.95$ ，因  $h(20) = 2.2033$  不在式 (19) 之區間內，故田間族羣密度  $\mu \approx 20$ ，較 20 稍小，但其上限已非常接近 20。

上述例子中  $h(\bar{x}) = \sinh^{-1} \sqrt{\bar{x}} = \log_e(\sqrt{\bar{x}} + \sqrt{\bar{x}^2 + 1})$ ,  $h'(\bar{x}) = \frac{1}{2\sqrt{\bar{x}(\bar{x} + 1)}}$ 。令  $d_h/h'(\bar{x}) = 3$  的目的是使  $|\bar{x} - \mu| \leq d$ ,  $d \approx 3$ 。例子中之觀測值是根據 1975 年 10 月 7 日調查之若蟲資料做為假設族羣，而後從此族羣中逢機抽出之觀測值，此假設族羣之族羣平均  $\mu = 19.382$ ,  $\sigma^2 = 96.648$ 。

## 引 用 文 獻

- (1) 程建中, 1976。褐飛蝨棲羣與水稻之相互關係。中興大學昆蟲學研究所碩士論文。
- (2) ABRAMOWITZ, M. and I. A. STEGUM. 1965. Handbook of mathematical functions with formulars, graphs and mathematical tables. p. 949-952. New York: Dover Publications, Inc.
- (3) ANSCOMBE, F. J. 1949. The statistical analysis of insect counts based on the negative binomial distribution. Biometrics 5: 165-173.
- (4) BARTLETT, M. S. 1947. The use of transformation. Biometrics 3: 39-52.
- (5) BEALL, G. 1942. The transformation from entomological field experiments so that the analysis of variance become applicable. Biometrika 32: 243-262.
- (6) BICKEL, P. J. and K. A. DOCKSUM. 1977. Mathematical statistics Basic idea and selected topics. 28-30, Taipei: 華泰書局翻印本。
- (7) BLISS, C. I. and A. R. G. OWEN, 1958. Negative binomial distribution with a common K. Biometrika 45: 37-58.
- (8) CARNAHAN, B., H. A. LUTTER and J. O. WILKES. 1969. Applied numerical methods. New York: John Wiley.
- (9) GRAYBILL, A. F. 1978. Theory and application of the linear model. p. 131. North Scituate: Duxbury.
- (10) JOHNSON, N. L. 1947. Systems of frequency curves generated by method of translation. Biometrika 36: 149-176.
- (11) KANO, E. 1979. Distribution pattern of the rice brown planthopper and field sampling technique. The Rice Brown Planthopper. 135-147. Compiled by Food and Fertilizer Technology Center for The Asian and Pacific Region.

- (12) LIN, T. L. and C. N. CHEN. 1979. The spatial distribution of the rice brown planthopper in Taiwan. Proceedings of the ROC-Japan Symposium on Rice Productivity. 123-134. Institute of Botany, Academia Sinica.
- (13) MOSHMAN, J. 1958. A method for selecting the size of the initial sample in Stein two sample procedure. *Ann. Math. Stat.* **29**: 1271-1275.
- (14) OAKLAND, G. B. 1950. An application of sequential analysis to whitefish sampling. *Biometrics* **6**: 59-67.
- (15) SEELBINDER, B. M. 1953. On Stein's two-stage sampling scheme. *Ann. Math. Stat.* **24**: 640-649.
- (16) SILVESTER, E. S. and E. L. COX. 1961. Sequential plans for sampling aphids on sugar beets in Kern County, California. *Jour. Econ. Ent.* **54**: 1080-1085.
- (17) STEIN, C. 1945. A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Stat.* **16**: 203-258.
- (18) WALD, A. 1947. *Sequential analysis*. New York: Wiley.
- (19) WATERS, W. E. 1955. Sequential sampling in forest insect surveys. *Forest Science* **1**: 68-77
- (20) WILSON, L. F. 1973. A new procedure for rapidly estimating European pine sawfly (*Hymenoptera: Diprionidae*) population levels in young pine plantations. *The Canadian Entomol. Entomol.* **103**: 1315-1322.

## ECONOMIC SAMPLE SIZE ON THE POPULATION DENSITY ESTIMATION OF RICE BROWN PLANTHOPPER<sup>1</sup>

YUN-MING PONG, TSAN-LONG LIN<sup>2</sup> and CHIOU-NAN CHEN<sup>3</sup>

### SUMMARY

Stein's two-stage sampling method is applicable when the observations taken from a normal distribution. Since the spatial distribution of rice brown planthopper is not normal, data transformations of the kinds  $\log(x+0.1)$ ,  $\sinh^{-1}\sqrt{x}$ ,  $r+\delta \log\left(\frac{x-\xi}{\lambda}\right)$ ,  $r+\delta \log\left(\frac{x-\xi}{x-\xi+\lambda}\right)$ ,  $r+\delta \sinh^{-1}\left(\frac{x-\xi}{\lambda}\right)$  are tried. The results show that the transformations are all ineffective.

Monte Carlo simulation is used to study the results when the sample mean  $\bar{x}$  is transformed by the transformations  $\log(\bar{x}+0.1)$ ,  $\sinh^{-1}\bar{x}$ ,  $\sinh^{-1}\sqrt{\bar{x}}$ . Infinite populations which have the same probability distribution as the data observed are generated, and from each population simulated 100 sets of samples are drawn. The sample size  $n$  varies from 5, 10, 15, 20, 25, 30. Results show that the variance of the transformed  $h(\bar{x})$  are much smaller than the untransformed  $\bar{x}$ . Also,  $h(\bar{x})$  is asymptotically unbiased and bias is negligible. The studentized variables  $\sqrt{n}[h(\bar{x})-h(\mu)]/[h'(\bar{x})s]$  approximate  $t$ -distribution better than the untransformed.

The sample size estimated by Stein's two-stage sampling method is a random variable. Its expectation  $E(n)$  is a function of the population variance, initial sample size  $n_0$  and the width of the confidence interval. Under various desired differences between means in different accuracy levels, the value of  $E(n)$  almost reaches its optimal value when initial sample size  $n_0$  equals 21. For economic reason, it seems reasonable to observed 21 hills of rice plant at the first stage. The final sample size is determined according to the precision level required.

1. This research was supported by the National Science Council, R. O. C.
2. Graduate student and Professor respectively, Department of Agronomy, National Taiwan University.
3. Senior Specialist, Taiwan Plant Protection Center