行政院國家科學委員會補助專題研究計畫成果報告

反應為連續值的適合部署設計（2/3）

計畫類別：■個別型計畫　□ 整合型計畫

計畫編號：NSC90-2118-M-032-015

執行期間：90 年 08 月 01 日　至　91 年 10 月 31 日

計畫主持人：黃文濤

共同主持人：

計畫參與人員：吳昭賢、郭信霖

本成果報告包括以下應繳交之附件：

□赴國外出差或研習心得報告一份

□赴大陸地區出差或研習心得報告一份

■出席國際學術會議心得報告及發表之論文各一份

□國際合作研究計畫國外研究報告書一份

執行單位： 淡江大學

中 華 民 國　91 年　12 月　5 日

反應為連續值的適合部屬設計（2/3）

中文摘要：

本計畫共考慮兩個相關問題，即樣本數最佳解以及 k 母體中最佳母體之確認。前者在貝氏結構下，壽命為指數分佈，刻度參數之先驗分佈為伽碼分佈，考慮三種損失下之損失函數。則本計畫之成果提出了貝氏最佳解及貝氏決策。後者在考慮 k 個常態母體下，給定均值及變異數的門檻要求下，提出了貝氏最佳母體之確認程序，並證明了它的大樣本最佳性質。

英文摘要：

In this project, we consider two problems. The first problem is related to optimal allocation problem for choosing sample size. The second problem is related identification of the best normal population under some thresholds of related parameters. In the first problem we have proposed a Bayes solution which can be used to compute the optimal n. For the second problem, we consider $k(k>2)$ populations whose mean θ_i and variance σ_i^2 are all unknown. For given control values θ_0，σ_0^2 and δ_0, we are interested in identifying some population whose mean is most closed to θ_0. A Bayes approach is set up and an empirical Bayes procedure is proposed which has been shown to be asymptotically optimal.

Keywords and phrases：Best population identification, multiple criteria, empirical Bayes rule.

1. Introduction

In this second year project, we consider two kinds of problem.

The first problem relates to optimal allocation of number of items for life testing. In a batch of N items, we take a part of n items for experiment. Based on the experimental data, we need to conclude a decision if the batch is to be accepted or rejected. Three kinds of losses are considered. Loss due to rejection of batch if it meets requirement, cost of each item for experiment and cost of unit time for testing. Since life time is type I-censored, two quantities are to be determined, they are the number of n and the censoring time t. We propose a Bayes solution for this design problem. The solution consists of optimal values for n and t and a decision that the batch is to be accepted or rejected.

The second problem relates to identification of certain population which meets some requirements. Suppose k Normal populations π_i are under consideration and samples are taken from each population. π_i has mean θ_i and variance σ_i which are both unknown. Let θ_0 , δ_0 and $\sigma_0{}^2$ are control values. We consider the following requirements

Let $S_1 = \{\pi_i \mid \sigma_i^2 \le \delta_0^2\}$, π_i is considered the best if $\pi_i \in S_1$, $|\theta_i - \theta_0| \le \delta_0$ and $|\theta_i - \theta_0| = \min|\theta_j - \theta_0|$ where minimum is taken over π_j in S_1.

It is to be noted that neither (1) includes (2) nor (2) includes (1). They are also not exclusive to each other. In the formulation of problem, we assume that the mean θ has a normal prior with unknown parameters. A Bayes procedure is proposed to identify the best population for each requirement respectively.

2. Optimal allocation and Its Decision

Let $X_1,\dots,X_n$ denote the lifetimes of the n components put on a life test experiment. It is assumed that $X_1,\dots,X_n$ are mutually independent, follow an exponential distribution having expected lifetime $\theta = \lambda^{-1}$. We denote such an exponential distribution by $E(\lambda)$. Let $X_{(1)} \le \dots \le X_{(n)}$ be the order statistics of $X_1,\dots,X_n$. Since the sample is subject to type I censoring at time t, the true observations are :

$Y_i = \min(X_{(i)}, t)$, $i = 1,\dots,n$.

Then,

$M \equiv M(n,t) = \max\{i \mid X_{(i)} \le t \, , i = 1,\dots,n\}$

is the number of failures by time t, M and $Y(M,t) = (Y_1,\dots,Y_M)$ are the observable random variables and

$Y(n,t,M) = \sum_{i=1}^{M} Y_i + (n-M)t$ is the total lifetime of the n items up to time t.

Suppose that a batch of lifetime

components is presented for acceptance sampling. Let a denote an action on this problem of acceptance sampling. When $a=1$, it means that the batch is accepted, and when $a=0$, it means to reject the batch. For the given sample size n, censoring time t and parameter λ, the loss of taking action a is defined as :

$$L(a,\lambda,n,t)=ah(\lambda)+(1-a)C_3+nC_1+tC_2,$$

where C_1, C_2 and C_3 are positive constants,

C_1 : the cost per item inspected,

C_2 : the cost per unit time used for life test,

C_3 : the loss due to rejecting the batch,

and $h(\lambda)$: the loss of accepting the batch. Since $\theta=\lambda^{-1}$ is the expected lifetime, larger λ indicates smaller θ.

Then, for the fixed sample size n and the censoring time t, the Bayes decision function $\delta_B(\ |n,t)$, is given by :

$$\delta_B(m,y(m,t)|n,t) = \begin{cases} 1, & if\ \varphi_g(m,y(n,t,m)) \le C_3, \\ 0, & otherwise. \end{cases}$$

where

$$\varphi_g(m,y(n,t,m)) = \int_0^\infty h(\lambda)g(\lambda|m,y(m,t))\,d\lambda$$

Algorithm :

(1) start with $n=0$, compute $r(0,0,\delta_B)$.

(2) For each $n=1,\dots,n^*$, compute $r(n,t,\delta_B(\ |n,t))$ and minimize $r(n,t,\delta_B(\ |n,t))$ with respect to t. We denote the minimizer by $t_B(n)$.

(3) Computer the risks among $r(0,0,\delta_B)$ and $r(n,t_B(n),\delta_B(\ |n,t_B(n)))$.

Let $S=\{n\in I_n \,|\, r(n,t_B(n))<r(0,0,\delta_B)\}$.

Then, n_B is determined as :

$$n_B=\begin{cases} 0, & if\ S=\phi \\ \min\{n|n\in S\}, & if\ S\neq\phi \end{cases}$$

3. **Bayes solutions to Best Population Identification Consider the following loss**

$$L(a,\theta,\sigma)=\alpha\left[\sum_{i=0}^{k}a_i\Delta_i^2-\Delta_{[1]}^2\right] +(1-\alpha)\sum_{i=0}^{k}a_i(\sigma_i^2-\sigma_0^2)I_{\{\sigma_i>\sigma_0\}}$$

for some prefixed $\alpha(0<\alpha<1)$, where $\Delta_i=|\theta_i-\theta_0|$ and $\Delta_{[1]}=\min_{0\le i\le k}\Delta_i$.

For each $i=1,\dots,k$, let $X_{i1},\dots,X_{iM}$ be an independent random sample of size M

from a normal population π_i with mean θ_i and σ_i^2 0. It is assumed that θ_i is a realization of a random variable Θ_i with a normal prior distribution $N(\mu_i, \tau_i^2)$, where μ_i and τ_i^2 are both unknown. The random variables $\Theta_1, \ldots, \Theta_k$ are assumed to be mutually independent.

For each population π_i, $i = 1, \ldots, k$, we estimate the unknown parameters $N(\mu_i, \tau_i^2)$ and σ_i^2 based on the past data X_{ijt}, $j = 1, \ldots, M$, $t = 1, \ldots, n$. We denote

$$\begin{cases} X_{i.t} = \dfrac{1}{M}\sum_{j=1}^{M} X_{ijt}, \quad X_i(n) = \dfrac{1}{n}\sum_{t=1}^{n} X_{i.t}, \\ W_{i.t}^2 = \dfrac{1}{M-1}\sum_{j=1}^{M}\left(X_{ijt} - X_{i.t}\right)^2, \end{cases}$$

$$S_i^2(n) = \frac{1}{n-1}\sum_{t=1}^{n}\left(X_{i.t} - X_i(n)\right)^2$$

$$W_i^2(n) = \frac{1}{n}\sum_{t=1}^{n} W_{i.t}^2.$$

Define

$$\begin{cases} \hat{\mu}_{in} = X_i(n), \quad \hat{\sigma}_{in}^2 = W_i^2(n) \\ \hat{\upsilon}_{in}^2 = S_i^2(n), \quad \hat{\tau}_{in}^2 = \max\left(\hat{\upsilon}_{in}^2 - \dfrac{\hat{\sigma}_{in}^2}{M}, 0\right) \end{cases}$$

Also, for $i = 0, 1, \ldots, k$, we define

$$\hat{\phi}_{in}(x_i) = \begin{cases} \theta_0 & \text{if } i = 0; \\ \left(\hat{\tau}_{in}^2 x_i + \dfrac{\hat{\sigma}_{in}^2}{M}\hat{\mu}_{in}\right)\Big/\hat{\upsilon}_{in}^2 & \text{otherwise.} \end{cases}$$

$$\hat{\psi}_{in}^2(\tilde{x}_i) = \begin{cases} \sigma_0^2 & \text{if } i = 0; \\ 1\Big/\left(\dfrac{M}{\hat{\sigma}_{in}^2} + \dfrac{1}{\hat{\tau}_{in}^2}\right) & \text{otherwise} \end{cases}$$

and

$$\hat{U}_{in}(\tilde{x}_i) = \begin{cases} \delta_0^2 & \text{if } i \\ \hat{\psi}_{in}^2(\tilde{x}_i) + \left(\hat{\phi}_{in}(\tilde{x}_i) - \theta_0\right)^2 & \text{other} \end{cases}$$

and also define

$$\hat{U}_{in}'(\tilde{x}_i) = \begin{cases} \delta_0^2 & \text{if } i = 0; \\ \delta_1^2 & \text{if } \hat{\sigma}_{in}^2 > \sigma_0^2; \\ \hat{U}_{in}(\tilde{x}_i) & \text{otherwise} \end{cases}$$

For each $x \in \chi$, let

$$Q_n(\tilde{x}) = \left\{ i \,\middle|\, \hat{U}_{in}'(\tilde{x}_i) = \max_{0 \le j \le k} \hat{U}_{jn}'(\tilde{x}_j), i = 0, 1, \ldots, k \right\}$$

Again, define

$$i_n^* = i_n^*(\tilde{x}) = \begin{cases} 0 & \text{if } Q_n(\tilde{x}) = \{ \\ \min\{ i \mid i \in Q_n(\tilde{x}), i \neq 0 \} & \text{otherwise.} \end{cases}$$

Then, an empirical Bayes identification rul $\tilde{d}^{*n} = (d_0^{*n}, d_1^{*n}, \ldots, d_k^{*n})$ as follows

$$\begin{cases} d_{i_n^*}^{*n}(\tilde{x}) = 1, \\ d_j^{*n}(\tilde{x}) = 0, \quad \text{for } j \neq i_n^*. \end{cases}$$

Theorem : Assume $\sigma_i^2 \neq \sigma_0^2$, for al $i = 1, \ldots, k$. The proposed empirical Baye identification rule $\tilde{d}^{*n}(\tilde{x})$ is asymptoticall optimal.

4. Reference

【1】Gupta, S.S., et al. (1994). Empirical Bayes rules for selecting the best normal population compared with a control, *Statistics and Decisions* **12**, 125-147.

【2】Huang, W.T. and Lai, Y.T. (1999). Empirical Bayes procedure for selecting the best population with multiple criteria, *Ann. Inst. Statis. Math.* Vol. **51**, No. 2, 281-299.

第 53 屆國際統計學會(The 53rd Session of ISI)會報告後

淡江大學管科所　黃文濤

第 53 屆國際統計學會(The 53rd Session of International Satatistics Institute)於九十年八月二十二至二十九日在韓國漢城市的 COEX(Convention and Exhibition Center)舉行。一切順利，圓滿閉幕。

我的演講排在 8 月 25 日上午，小組的題目是〝Multiple comparisons,Ranking,Selection and Related Topics〞.我當該 session chairman. 我演講的題目是〝Generalized subset selection under Normality〞。它解決了 Ranking and Selection 這領域近二十年來的部份問題，同時很大地擴展了古典的擇優理論方法。主要問題是在樣本數不同，各母體變異數不同且未知的情況下選取最大均值母體一直是研究者，想解決又未解決的問題(非逐次方法)。此外我們也探討了對應最大噪音比(noise ratio)以及 quantile 之擇優問題，這些都牽涉到另一個未知參數(變異數)。

因為我是這個 session 主持人，所以排在最後一個講。來聽講的坐有講堂的八分滿，人數還真不少。其中有位是美國教授雖已退休，但他不只談到他的本行，又談到一大部分統計的教學的經驗，這佔去別人的時間有十多分鐘，但鑑於他的熱誠以及當時聽眾的喜歡，我還是准了他的要求，把會議時間往後延。另一個印象較深的是一位日籍教授和他的博士研究生（大陸籍）提出一篇文章探討日本經濟一直低迷不振的主要原因，那是因為政府花在公共建設費用過於龐大所致。他提出了數據及理論，宣稱他們是在這個問題上第一次真正指出問題所在的人而且理論客觀有據。惜時間有限未能讓他們大發揮。

這個年會共有 invited 論文領域達 76 個之多，contributed paper 領域 164 個。每二年舉行一次，下次 2003 年 8 月左右在德國柏林市召開。

韓國為召開這次會議，籌劃了許久，動用了許多人力，發動了幾乎所有在漢城的統計學界。在大會開幕上，韓國大統領金大中先生也來致詞，可見重視的程度。大會在 28 日晚間舉行了盛大晚宴(farewell party)，而學術研討一直到 29 日下午才全部結束。

在開會期間，我也應邀參加了韓國普渡大學統計系校友的歡宴，席中有普渡大學統計系教授 David Moore 及他的太太(韓籍)。

這次能參加 ISI 年會我要感謝國科會的補助，使我有機會與外國統計界交流，互相探討研究題目，並當了 session chair,在國際上,台灣統計學界也受重視。（附上論文一篇為演講時所用搞）。