行政院國家科學委員會專題研究計畫成果報告

最佳化測量試驗之研究

Optimal Allocation of Measurements in a Linear Calibration Process

計畫類別: ☑個別型計畫 ☐整合型計畫

計畫編號: NSC-89-2118-M-002-010

執行期間: 89 年 8 月 1 日 至 90 年 7 月 31 日

計畫主持人: 廖 振 鐸

處理方式: ☑可立即對外提供參考
☐一年後可對外提供參考
☐二年後可對外提供參考

執行單位: 國立台灣大學農藝學研究所試驗統計組

中華民國 90 年 10 月 20 日

中 文 摘 要

本研究探討測量(calibration)過程中，用於校正的標準物(standard)及未知測量物(unknown)的分配問題。基於下列直線測量(linear calibration)模式

$$y_i = \alpha + \beta\tau + \varepsilon_i$$

其中 y_i 表示實際測量值，α 及 β 為未知的參數，代表測量過程中所產生的系統誤差(systematic error)，ε_i 代表隨機機差(random error)被假設為平均(mean)為 0 變方(variance)為 σ^2 的常態隨機變數(normal random variables)，τ 為未知測量物的真正值(true value)。

我們推導模式中未知參數 α 、β 、σ^2 及 $\tau_1, \tau_2, \cdots, \tau_m$ (假設有 m 個未知測量物)的最大概似估式(MLE)及其大樣本(asymptotic)的變方。針對測量總數 $N = a_0 + a_1 + \sum_{i=1}^{m} n_i$ ，a_0 及 a_1 表示兩個標準物的測量次數；n_i 表示第 i 個未知測量物的測量次數，$i = 1,2,\cdots,m$ ，提出區域 A-最佳化(locally A-optimal)測量試驗及貝式 A-最佳化(Bayesian A-optimal) 測量試驗。並且利用統計模擬研究，以印證結果的精準度。詳細研究成果請參照英文報告。

Optimal Allocation of Measurements in a Linear Calibration Process

C. T. Liao
Division of Biometry
Institute of Agronomy
National Taiwan University
Taipei, Taiwan

Abstract

A problem of allocation of measurements for a linear calibration process is considered in this article. It is assumed that a total of N measurements are made some of which may be measurements on two distinct standards while the remaining measurements are of m different unknown specimens. We discuss the optimal allocations of the N measurements for the two standards and m unknown specimens based on the A-optimal criterion applied to the asymptotic variances of the maximum likelihood estimates of the true values of the m unknown specimens. Since the optimal allocation depends on the true values of unknown specimens, the investigator is able to get a locally or Bayesian A-optimal design based on some reasonable prior information on the values of the unknown specimens. The situation that the measurement process is budgeted within a given cost is also discussed.
Keywords: systematic errors, random errors, A-optimal, measurement designs.

1 Introduction

A measurement process is typically subject to errors which are generally classified as systematic, or random, or a combination of both. The random errors are defined to have a zero expected value, and the systematic errors are defined to be due to biases in the measurement process. Typically, the instrument is considered to have a linear systematic error in which the measured value of an unknown specimen ("*unknown*" for short) is described by the following model

$$y_i = \alpha + \beta\tau + \epsilon_i \tag{1.1}$$

where α and β are parameters of the systematic error; τ represents the true value of the unknown being measured; and ϵ_i represents the random error.

Since "*standards*" have known true values, the errors associated with the measurement process are observed whenever a standard is measured. Therefore, the known standards are always used to estimate the parameters α and β. The calibration problem for the estimation procedure has been extensively studied. The main results can be seen in Fuller (1987) and Brown (1993). Literature pertaining to calibration problems involving the allocation of the measurements appears to begin with Pepper (1973). He discussed the measurement designs for the process described by a simple additive model of the form

$$y_i = \tau + (b_i + \eta_i)$$

where the quantity $(b_i + \eta_i)$ represents random errors. b_i arise from a random walk process and η_i are identical and independent normal random variables with 0 mean and a common constant variance. Perng and Tong (1977) considered the linear calibration model of (1.1). They presented a sequential procedure, observing either standard or unknown at each time, for the optimal allocation of the measurements for the interval estimation of the single one unknown so that the probability of coverage is maximized. Liao, Taylor and Iyer (2000) discussed optimum balanced designs for the additive model with correlated errors arising from a first order autoregressive process (AR(1)).

In this paper, we simply consider the linear calibration model of (1.1) with random errors being assumed to be identical and independent normal random variables with 0 mean and a common variance. We discuss the optimal allocation of the N measurements for the two standards and m unknowns when N is large. In the next section, we describe our problem. In section 3, we present the maximum likelihood estimation for the parameters α, β and the true values of the unknowns. In sections 4 and 5, we discuss some possible optimal measurement designs for the problem treated in the paper. Finally, we extend the results to the situation that a cost constraint is considered in the measurement process.

2 The problem

Let the number of measurements of the two distinct standards S_0 and S_1 be denoted by a_0 and a_1, respectively, and the number of measurements of unknown j is denoted by n_j, for $j = 1, 2, \cdots, m$. Thus, the total number of measurements $N = a_0 + a_1 + \sum_{j=1}^{m} n_j$. Let μ_0 and μ_1 be the true values of the standards S_0 and S_1, respectively. μ_0 and μ_1 are certainly known. Also let τ_j denote the true value of unknown j, for $j = 1, 2, \cdots, m$. Define the indicator functions $\delta_i^{S_j}$ and $\delta_i^{U_j}$ by

$$\delta_i^{S_j} = \begin{cases} 1 & \text{if } j = 0 \text{ and observation } i \text{ is of the standard } S_0 \\ 1 & \text{if } j = 1 \text{ and observation } i \text{ is of the standard } S_1 \\ 0 & \text{otherwise.} \end{cases}$$

and

$$\delta_i^{U_j} = \begin{cases} 1 & \text{if } j \in \{1, 2, \cdots, m\} \text{ and observation } i \text{ is of unknown } j \\ 0 & \text{otherwise.} \end{cases}$$

Then, the model described in (1.1) can be rewritten as

$$y_i = \alpha + \beta(\delta_i^{S_0}\mu_0 + \delta_i^{S_1}\mu_1 + \sum_{j=1}^{m} \delta_i^{U_j}\tau_j) + \epsilon_i, \quad \text{for } i = 1, 2, \ldots, N. \tag{2.1}$$

The random errors ϵ_i are now assumed to be independent and arise from a $N(0, \sigma^2)$ distribution. This paper is concerned with obtaining optimal allocation of a_0, a_1, n_1, n_2, ..., n_m such that the true values of unknowns $\tau_1, \tau_2, \ldots, \tau_m$ can be estimated most efficiently. In particular, we consider the A-optimality criterion according to which the average of asymptotic variances of the MLEs of the true values is minimized.

3 MLEs and their asymptotic variances

The log-likelihood function corresponding to the observed data of (2.1) is given by

$$l = -\frac{N}{2}log(2\pi) - \frac{N}{2}log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}[y_i - \alpha - \beta(\delta_i^{S_0}\mu_0 + \delta_i^{S_1}\mu_1 + \sum_{j=1}^{m}\delta_i^{U_j}\tau_j)]^2. \tag{3.1}$$

After differentiating the log-likelihood function and equating the derivatives to zero, we have the following likelihood estimators.

$$\hat{\alpha} = \bar{y} - \hat{\beta}\mu \tag{3.2}$$

$$\hat{\beta} = \frac{\sum_{i=1}^{N}\delta_i^{S_0}(y_i - \bar{y})(\mu_0 - \mu) + \sum_{i=1}^{N}\delta_i^{S_1}(y_i - \bar{y})(\mu_1 - \mu)}{a_0(\mu_0 - \mu)^2 + a_1(\mu_1 - \mu)^2} \tag{3.3}$$

$$\hat{\tau}_j = \frac{\bar{y}_j - \hat{\alpha}}{\hat{\beta}}, \quad \text{for } j = 1, 2, \ldots, m, \tag{3.4}$$

where

$$\bar{y} = \frac{1}{a_0 + a_1}(\sum_{i=1}^{N}\delta_i^{S_0}y_i + \sum_{i=1}^{N}\delta_i^{S_1}y_i),$$

$$\mu = \frac{a_0\mu_0 + a_1\mu_1}{a_0 + a_1}$$

and

$$\bar{y}_j = \frac{\sum_{i=1}^{N}\delta_i^{U_j}y_i}{n_j}, \quad \text{for } j = 1, 2, \ldots, m.$$

Clearly, the MLE for τ_j given in (3.4) is the well known classical estimator computed from $\hat{\alpha}$ of (3.2) and $\hat{\beta}$ of (3.3), which only depend on the observed data of the standards.

Computing the second derivative of the log-likelihood function of (3.1) with respect to parameters α, β, τ_1, τ_2, ..., τ_m and σ^2, and taking expectations, we can obtain the Fisher information matrix. By using the standard properties of partitioned matrices (Graybill, 1983), we can invert the Fisher information and obtain the asymptotic covariance matrix for the MLEs of τ_1, τ_2, ..., τ_m, which are of primary interest, as follows

$$Var(\hat{\boldsymbol{\tau}}) = \{v_{jk}\},\ j = 1, 2, \ldots, m;\ k = 1, 2, \ldots, m,$$

where $\hat{\boldsymbol{\tau}} = [\hat{\tau}_1, \hat{\tau}_2, \ldots, \hat{\tau}_m]'$ and

$$v_{jk} = \begin{cases} \frac{1}{\beta^2}\left[\frac{1}{n_j} + \frac{a_0(\mu_0-\tau_j)^2 + a_1(\mu_1-\tau_j)^2}{a_0 a_1(\mu_0-\mu_1)^2}\right] & \text{if } j = k \\ \frac{1}{\beta^2}\left[\frac{a_0(\mu_0-\tau_j)(\mu_0-\tau_k) + a_1(\mu_1-\tau_j)(\mu_1-\tau_k)}{a_0 a_1(\mu_0-\mu_1)^2}\right] & \text{if } j \neq k. \end{cases}$$

4 Locally A-optimal measurement designs

It is easy to see that an explicit expression for the trace of the covariance matrix of $Var(\hat{\boldsymbol{\tau}})$ is given by

$$Trace(Var(\hat{\boldsymbol{\tau}})) = \frac{1}{\beta^2}\left[\sum_{j=1}^{m}\frac{1}{n_j} + \frac{1}{a_1}\sum_{j=1}^{m}\frac{(\mu_0-\tau_j)^2}{(\mu_0-\mu_1)^2} + \frac{1}{a_0}\sum_{j=1}^{m}\frac{(\mu_1-\tau_j)^2}{(\mu_0-\mu_1)^2}\right].$$

Let $\theta_0 = \sum_{j=1}^{m}\frac{(\mu_0-\tau_j)^2}{(\mu_0-\mu_1)^2}$ and $\theta_1 = \sum_{j=1}^{m}\frac{(\mu_1-\tau_j)^2}{(\mu_0-\mu_1)^2}$. Also let $b_0 = a_0/N$, $b_1 = a_1/N$ and $r_j = n_j/N$, for $j = 1, 2, \ldots, m$. For obtaining A-optimal measurement designs for estimation of τ_1, τ_2, ..., τ_m, it is needed to minimize the following function

$$f(b_0, b_1, r_1, r_2, \ldots, r_m, \theta_0, \theta_1) = \frac{\theta_1}{b_0} + \frac{\theta_0}{b_1} + \sum_{j=1}^{m}\frac{1}{r_j}, \tag{4.1}$$

subject to the constraint

$$b_0 + b_1 + \sum_{j=1}^{m} r_j = 1.$$

For given values of θ_0 and θ_1, by using Lagrange multipliers, we have the following optimal allocation of the measurements for the standards and unknowns

$$b_0^* = \frac{\sqrt{\theta_1}}{\sqrt{\theta_0} + \sqrt{\theta_1} + m}; \tag{4.2}$$

$$b_1^* = \frac{\sqrt{\theta_0}}{\sqrt{\theta_0} + \sqrt{\theta_1} + m} \tag{4.3}$$

and

$$r_j^* = \frac{1}{\sqrt{\theta_0} + \sqrt{\theta_1} + m}, \quad \text{for } j = 1, 2, \ldots, m. \tag{4.4}$$

Since the optimal ratios of b_0^*, b_1^*, r_1^*, r_2^*, $\cdots$, r_m^* given above depend on the parameters τ_1, τ_2, $\cdots$, τ_m, the problem of allocation cannot be solved unless one has some prior information on these parameters.

Therefore, we first consider the following particular case, which will be called "*uniform scheme*", for obtaining a locally A-optimal measurement design. There is a possibility to designate the two distinct standards such that the range of $[\mu_0, \mu_1]$ is large enough to cover the values of the unknowns. Suppose the values of τ_1, τ_2, $\ldots$, τ_m are discrete and uniformly distributed among the range of $[\mu_0, \mu_1]$, i.e.

$$\tau_j = \mu_0 + (\frac{\mu_1 - \mu_0}{m+1})j, \quad \text{for } j = 1, 2, \ldots, m. \tag{4.5}$$

Then substitute the values of (4.5) in θ_0 and θ_1, we have

$$\theta_0 = \theta_1 = \frac{m(2m+1)}{6(m+1)}.$$

Thus, for fixed m, the allocation ratios of (4.2), (4.3) and (4.4) can be computed by substituting these θ_0 and θ_1. Table 4.1 displays some locally A-optimal allocations based on the uniform scheme.

Table 4.1. The locally A-optimal allocations based on the uniform scheme for $1 \leq m \leq 5$.

m	b_0^*	b_1^*	r_j^*
1	.250	.250	.500
2	.214	.214	.286
3	.192	.192	.205
4	.176	.176	.162
5	.165	.165	.134

To evaluate how good the above locally A-optimal design is in estimating the true values of the unknowns. The following simulation study is carried out. It is sufficient to consider the single unknown case, i.e. $m = 1$. The simulation results are quite similar for various

values of α and β. Hence, we only report the case that $u_0 = 0$, $u_1 = 1$, $\alpha = .2$, $\beta = .4$ and $\sigma = .1$. Table 4.2 displays the average of $\hat{\tau}_1$ values and their standard deviation of 500 simulated experiments for each various values of $\tau_1 = .1, .3, .5, .7, .9$ and $N = 20, 40, 60, 80$ and 100.

Table 4.2. The average of $\hat{\tau}_1$ values and their standard deviation of the simulation study.

τ_1		$N = 20$	$N = 40$	$N = 60$	$N = 80$	$N = 100$
.1	Average	.0886	.0979	.0987	.1012	.0990
	S.D.	.1383	.0935	.0766	.0669	.0574
.3	Average	.2938	.2996	.3004	.3020	.3014
	S.D.	.1217	.0838	.0696	.0605	.0521
.5	Average	.4991	.5013	.5021	.5028	.5037
	S.D.	.1138	.0804	.0673	.0581	.0505
.7	Average	.7043	.7030	.7039	.7037	.7061
	S.D.	.1163	.0840	.0701	.0602	.0530
.9	Average	.9095	.9047	.9056	.9045	.9085
	S.D.	.1286	.0939	.0774	.0663	.0590

The results show that the performance improves significantly as N larger than 20. For fixed N, the minimum standard deviation occurs at $\tau_1 = .5$ since the used design is the locally optimal design for τ_1 being .5. More importantly, the locally A-optimal design is likely to be robust for various τ_1 values, i.e. for fixed N, the standard deviations for various τ_1 values do not fluctuate drastically. Therefore, we suggest that the locally A-optimal allocation based on the uniform scheme could be a practical choice if the investigator does not have any prior knowledge on the true values of the unknowns.

5 Bayesian A-optimal measurement designs

It is well known that optimal designs for the non-linear models, such as model (2.1), depend on the values of unknown parameters. In this section, instead of the locally optimal designs, a prior distribution for $\tau_1, \tau_2, \ldots, \tau_m$ will be incorporated into the optimum criterion of (4.1) to yield a Bayesian optimal design. A general theory regarding the optimum Bayesian designs refers to Atkinson and Donev (1992).

It is reasonably assumed that $\tau_1, \tau_2, \ldots, \tau_m$ are independently and arise from the continuous uniform distribution over the range $[\mu_0, \mu_1]$. Then we take expectation of (4.1) over

this prior distribution to have the following

$$E_{\theta_0,\theta_1}[f(b_0, b_1, r_1, r_2, \ldots, r_m, \theta_0, \theta_1)] = \frac{m}{3b_0} + \frac{m}{3b_1} + \sum_{j=1}^{m} \frac{1}{r_j}.$$

We thus have the following optimal allocation of the measurements for the standards and unknowns

$$b_0^* = b_1^* = \frac{\sqrt{m/3}}{\sqrt{m/3} + \sqrt{m/3 + m}}$$

and

$$r_j^* = \frac{1}{\sqrt{m/3} + \sqrt{m/3 + m}}, \quad \text{for } j = 1, 2, \ldots, m.$$

The following table displays the Bayesian A-optimal measurement designs for $1 \leq m \leq 5$.

Table 5.1. The Bayesian A-optimal allocations based on the continuous uniform distribution over the range $[\mu_0, \mu_1]$, for $1 \leq m \leq 5$.

m	b_0^*	b_1^*	r_j^*
1	.268	.268	.464
2	.225	.225	.275
3	.200	.200	.200
4	.183	.183	.158
5	.170	.170	.132

It is of interest to note that the results of Table 4.1 and Table 5.1 are pretty close, but they still show the common-sense property that the weights of the standards increase as the prior information for the unknowns become more dispersed.

6 Concluding remarks

The results discussed in sections 4 and 5 can be easily extended to the case that a cost constraint is considered in the measurement process. Suppose that each measurement of the standards S_0 and S_1 cost c_0 and c_1 dollars, respectively; that each measurement of any

unknown costs c dollars; and that the total cost is budgeted within B dollars. Then the problem is to find a_0, a_1 and n_j, for $j = 1, 2, \ldots, m$, such that the objective function

$$f(a_0, a_1, n_1, n_2, \ldots, n_m, \theta_0, \theta_1) = \frac{\theta_1}{a_0} + \frac{\theta_0}{a_1} + \sum_{j=1}^{m} \frac{1}{n_j},$$

is minimized and subject to the constraint

$$c_0 a_0 + c_1 a_1 + c \sum_{j=1}^{m} n_j \leq B.$$

Similarly, we have the following optimal allocation for the budget being exactly equal to B dollars.

$$a_0^* = \frac{B\sqrt{\theta_1}}{\sqrt{c_0}(\sqrt{c_1\theta_0} + \sqrt{c_0\theta_1} + m\sqrt{c})},$$

$$a_1^* = \frac{B\sqrt{\theta_0}}{\sqrt{c_1}(\sqrt{c_1\theta_0} + \sqrt{c_0\theta_1} + m\sqrt{c})}$$

and

$$n_j^* = \frac{B}{\sqrt{c}(\sqrt{c_1\theta_0} + \sqrt{c_0\theta_1} + m\sqrt{c})}.$$

In practice, a_0, a_1 and n_j can be taken to be the largest integers less than or equal to a_0^*, a_1^* and n_j^*, respectively. The optimal allocation still depends on the true values of $\tau_1, \tau_2, \cdots, \tau_m$. Therefore, the investigator is able to get some locally or Bayesian A-optimal measurement design according to some reasonable prior information on the values of the unknowns.

References

[1] Atkinson, A. C. and Donev, A. N. (1992). Optimum experimental designs. New York: Oxford University Press.

[2] Brown, R. B. (1993). Measurement, regression and calibration. New York: Oxford University Press.

[3] Fuller, W. A. (1987). Measurement error models. New York: John Wiley and Sons.

[4] Graybill, F. A. (1983). Matrices with applications in statistics. Belmont, California: Wadsworth International Group.

[5] Liao, C. T., Taylor, C. H. and Iyer, H. K. (2000). Optimal measurement designs when errors are correlated. *J. Statist. Plann. Inference*, **84**, pp. 295-321.

[6] Pepper, M. P. G. (1973). A calibration of instruments with non-random errors. *Technometrics*, **15**, pp. 587-599.

[7] Perng, S. K. and Tong, Y. L. (1977). Optimal allocation of observations in inverse linear regression. *Ann. Statist.*, **5**, pp. 191-196.