Ann. Inst. Statist. Math.
30 (1978), Part A, 1-8

ON LEVEL OF SIGNIFICANCE OF THE PRELIMINARY TEST IN POOLING MEANS

KATUOMI HIRANO

(Received Feb. 17, 1977; revised Oct. 5, 1977)

1. Introduction and summary

In estimating the mean of a normal distribution $N(\boldsymbol{\mu}, \Sigma)$ a random sample of size n is taken. Suppose a second independent random sample of size m from $N(\boldsymbol{\xi}, \Sigma)$ is available. If $\boldsymbol{\mu}=\boldsymbol{\xi}$, it is advantageous to pool the two samples and use the pooled estimator. In some practical cases, one may perform a preliminary test of hypothesis $\boldsymbol{\mu}=\boldsymbol{\xi}$. If the hypothesis is accepted, a pooled estimator for $\boldsymbol{\mu}$ will be used; if the hypothesis is rejected, the experimenter uses only the first sample in his estimation procedure. Since this estimator always depends on the significance level, one can not uniquely decide the estimator for $\boldsymbol{\mu}$. The preliminary test estimation procedures always have this defect. For one sample case, Hirano [7] has obtained the optimal significance level of the preliminary test based on the shrinkage technique (see Thompson [13]) and/or Akaike's information criterion [1]. For two sample case, Hirano [8] has specified the necessary level of significance of the preliminary test in estimating the variance in a problem discussed by Bancroft [3]. The preliminary test procedures for the mean are considered by Kitagawa [10], Bennett [5], Asano [2], Bancroft and Han [4] among others. In this paper we specify the necessary significance level of the preliminary test in estimating the mean for the two sample case.

Let $\hat{\boldsymbol{\theta}}$ be the maximum likelihood estimate based on a random sample having probability density function $f(\boldsymbol{x}|\boldsymbol{\theta})$, where the unknown parameter $\boldsymbol{\theta}$ consists of k independent components. Then Akaike's information criterion [1] is given by $\mathrm{AIC}=-2\log_e L(\hat{\boldsymbol{\theta}})+2k$, where $L(\boldsymbol{\theta})$ denotes the likelihood. A proof of the asymptotic optimality of a decision procedure based on this criterion has been given by Inagaki [9].

In this paper the n observations $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ and the m observations $\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_m$ are taken from the normal distributions $N(\boldsymbol{\mu}, \Sigma_1)$ and $N(\boldsymbol{\xi}, \Sigma_2)$, respectively, where $\boldsymbol{\mu}$ and $\boldsymbol{\xi}$ and Σ_1 and Σ_2 are $p\times 1$ vectors and $p\times p$ positive definite matrices respectively, and they are assumed to be mutually independent. Let $\bar{\boldsymbol{x}}=\sum_{i=1}^{n}\boldsymbol{x}_i/n$, $\bar{\boldsymbol{y}}=\sum_{j=1}^{m}\boldsymbol{y}_j/m$ and $\bar{\boldsymbol{z}}=(n\bar{\boldsymbol{x}}+m\bar{\boldsymbol{y}})/(n+m)$ be the sample means and the sample mean of the pooled sample, respectively. When, especially, $p=1$, we denote $\boldsymbol{\mu}$, $\boldsymbol{\xi}$, $\bar{\boldsymbol{x}}$, $\bar{\boldsymbol{y}}$ and $\bar{\boldsymbol{z}}$ by μ, ξ, $\bar{x}$, $\bar{y}$ and $\bar{z}$, respectively.

2. Treatments of two sample problem by classical method and information criterion, when $\Sigma_1=\sigma_1^2 I_p$ and $\Sigma_2=\sigma_2^2 I_p$ are known

In this section, we suppose that $\Sigma_1=\sigma_1^2 I_p$ and $\Sigma_2=\sigma_2^2 I_p$ are known. First, we assume $\boldsymbol{\mu}=\boldsymbol{\xi}$ and consider the procedure of estimating the mean $\boldsymbol{\mu}$. We denote this situation by Model H_0. The AIC of Model H_0 is easily given by

$$\text{(2.1)}\quad \text{AIC}\,(H_0)=p(n+m)\log_e 2\pi+pn\log_e\sigma_1^2+pm\log_e\sigma_2^2 +\frac{1}{\sigma_1^2}\sum_{i=1}^{n}(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})'(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})+\frac{1}{\sigma_2^2}\sum_{j=1}^{m}(\boldsymbol{y}_j-\hat{\boldsymbol{\mu}})'(\boldsymbol{y}_j-\hat{\boldsymbol{\mu}})+2p$$

where $\hat{\boldsymbol{\mu}}=(1-w)\bar{\boldsymbol{x}}+w\bar{\boldsymbol{y}}$ and $w=m\sigma_1^2/(m\sigma_1^2+n\sigma_2^2)$.

Next, we assume $\boldsymbol{\mu}\neq\boldsymbol{\xi}$ and consider the problem of estimating $\boldsymbol{\mu}$ and $\boldsymbol{\xi}$. We denote this situation by Model H_1. The AIC of Model H_1 is given by

$$\text{(2.2)}\quad \text{AIC}\,(H_1)=p(n+m)\log_e 2\pi+pn\log_e\sigma_1^2+pm\log_e\sigma_2^2 +\frac{1}{\sigma_1^2}\sum_{i=1}^{n}(\boldsymbol{x}_i-\bar{\boldsymbol{x}})'(\boldsymbol{x}_i-\bar{\boldsymbol{x}})+\frac{1}{\sigma_2^2}\sum_{j=1}^{m}(\boldsymbol{y}_j-\bar{\boldsymbol{y}})'(\boldsymbol{y}_j-\bar{\boldsymbol{y}})+4p\,.$$

Therefore we may summarize the condition for the preference of Model H_0 by the minimum AIC procedure as follows:

Relation 1.

$$\text{(2.3)}\qquad \text{AIC}\,(H_0)-\text{AIC}\,(H_1)<0 \Longleftrightarrow (\bar{\boldsymbol{x}}-\bar{\boldsymbol{y}})'(\bar{\boldsymbol{x}}-\bar{\boldsymbol{y}})\Big/\left(\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}\right)<2p\,.$$

It should be noted that the left-hand side of the second inequality in (2.3) has the non-cental chi-square distribution with p degrees of freedom and with the non-centrality parameter $\nu=nm(\boldsymbol{\mu}-\boldsymbol{\xi})'(\boldsymbol{\mu}-\boldsymbol{\xi})/(m\sigma_1^2+n\sigma_2^2)$. The probability of $\text{AIC}\,(H_0)-\text{AIC}\,(H_1)<0$ under H_0 is equal to $\Pr\{\chi_p^2<2p\}$ for the χ^2-variates with p degrees of freedom. When $p=1$, this is equal to $0.8427\cdots$.

3. Examples of application of Relation 1

3.1. *For preliminary test estimation procedure*

First we consider the case $p=1$. Here we denote $\hat{\boldsymbol{\mu}}$ by $\hat{\mu}$ when

$p=1$. According to Kitagawa [10] and Bennett [5], a common form of estimation procedure has been provided by

$$PT=\begin{cases}(1-w)\bar{x}+w\bar{y} & \text{if } \dfrac{|\bar{x}-\bar{y}|}{\sqrt{\sigma_1^2/n+\sigma_2^2/m}}<\varepsilon_\alpha \\ \bar{x} & \text{otherwise},\end{cases} \tag{3.1}$$

where ε_α is the upper $100(\alpha/2)\%$ point of the standard normal distribution, and is assigned beforehand. But by Relation 1, we can decide on the critical level of the preliminary test, that is, $\varepsilon_\alpha=\sqrt{2}$ and the level of significance $\alpha=0.15729\cdots$.

Now, when the hypothesis H_0 and the class H_1 of alternatives are given by (i) H_0: $\mu=\xi$, H_1: $\mu\neq\xi$, (ii) H_0: $\mu=\xi$, H_1': $\mu>\xi$ and (iii) H_0: $\mu=\xi$, H_1'': $\mu<\xi$, we consider three preliminary test estimators for μ. Of course one of them is PT with (i). We specify the necessary significance levels of the preliminary tests (ii) and (iii). Let AIC (H_1') and AIC (H_1'') denote the AIC's of the models in estimating the means μ and ξ under the hypotheses H_1' and H_1'', respectively. Since, under H_1', $\max(\bar{x}, \hat{\mu})$ and $\min(\bar{y}, \hat{\mu})$ are the maximum likelihood estimators for μ and ξ, respectively, we obtain under the condition $\bar{x}>\bar{y}$

$$\text{AIC}(H_0)-\text{AIC}(H_1')<0 \Longleftrightarrow 0<\frac{\bar{x}-\bar{y}}{\sqrt{\sigma_1^2/n+\sigma_2^2/m}}<\sqrt{2}$$

and under the condition $\bar{x}\leqq\bar{y}$

$$\text{AIC}(H_0)-\text{AIC}(H_1')<0$$

is always satisfied. Therefore, we obtain under H_0

$$\begin{aligned}1-\alpha&=\Pr\{\text{AIC}(H_0)-\text{AIC}(H_1')<0\}\\ &=\Pr\{\text{AIC}(H_0)-\text{AIC}(H_1')<0\cap\bar{x}>\bar{y}\}\\ &\quad+\Pr\{\text{AIC}(H_0)-\text{AIC}(H_1')<0\cap\bar{x}\leqq\bar{y}\}\\ &=0.9213\cdots.\end{aligned} \tag{3.2}$$

Similarly, we also obtain under H_0, $1-\alpha=\Pr\{\text{AIC}(H_0)-\text{AIC}(H_1'')<0\}=0.9213\cdots$. Hence the necessary significance levels are $0.0786\cdots$ in both cases (ii) and (iii). This is the half of it in two-sided case.

Next, we assume that $\Sigma_1=\Sigma_2$. Hence without loss of generality we may let $\Sigma_1=I_p$ (i.e. $\sigma_1^2=\sigma_2^2=1$). Bancroft and Han [4] have given the preliminary test estimator for $\boldsymbol{\mu}$ defined as

$$\tilde{\boldsymbol{\mu}}=\begin{cases}\bar{\boldsymbol{z}} & \text{if } \dfrac{nm}{n+m}(\bar{\boldsymbol{x}}-\bar{\boldsymbol{y}})'(\bar{\boldsymbol{x}}-\bar{\boldsymbol{y}})<\eta_\alpha \\ \bar{\boldsymbol{x}} & \text{otherwise},\end{cases} \tag{3.3}$$

where η_α is the upper $100(1-\alpha)\%$ point of the chi-square distribution with p degrees of freedom and is pre-assigned. But by Relation 1, we can decide on the critical level of the preliminary test, that is $\eta_\alpha=\sqrt{2p}$.

3.2. *For relation between shrinkage technique and estimating the mean difference*

We assume $p=1$ and let α be the significance level for testing a null hypothesis H_0: $\mu-\xi=\delta$ against H_1: $\mu-\xi\neq\delta$. A preliminary test estimator for $\mu-\xi$ can be given by

$$PT=\begin{cases}\delta & \text{if } H_0 \text{ is accepted}\,, \\ \bar{x}-\bar{y} & \text{otherwise}\end{cases} \tag{3.4}$$

(see Hirano [7]). By Relation 1, the case "if H_0 is accepted" in the definition of PT of (3.4), can be replaced by

$$\delta-\sqrt{2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}<\bar{x}-\bar{y}<\delta+\sqrt{2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\,. \tag{3.5}$$

In this case α is equal to $0.15729\cdots$ and the mean squared error of PT is given by (33) of [7], if μ_0, μ and σ^2/n are replaced by δ, $\mu-\xi$ and $\sigma_1^2/n+\sigma_2^2/m$, respectively.

Next, the shrinkage estimate [13] of $\mu-\xi$ is easily given by

$$Sh=\frac{(\bar{x}-\bar{y}-\delta)^3}{(\bar{x}-\bar{y}-\delta)^2+\sigma_1^2/n+\sigma_2^2/m}+\delta\,.$$

From [7], Sh is more accurate than $\bar{x}-\bar{y}$ in the sense of the mean squared error when the inequality (3.5) is satisfied. The difference of the mean squared errors of PT and Sh is slight, and one may say that these approximately equal. The significance levels given by two methods may approximately coincide. But we can state as follows.

(i) The shrinkage technique is the method that one decides $\mu-\xi\neq\delta$ even if H_0 is accepted and uses Sh as an estimate of $\mu-\xi$. He is doubtful whether μ is really equal to $\xi+\delta$ even if H_0 is accepted.

(ii) The preliminary test estimation procedure is the method that one decides $\mu-\xi=\delta$ if H_0 is accepted and uses δ as an estimate of $\mu-\xi$. He believes that μ is really equal to $\xi+\delta$ if H_0 is accepted.

We can state the same notes as these when σ_1^2 and σ_2^2 are unknown.

4. Case that Σ_1 and Σ_2 are unknown but assumed to be equal

In this section we consider the case where $\Sigma_1=\sigma_1^2 I_p$ and $\Sigma_2=\sigma_2^2 I_p$ are unknown but $\Sigma_1=\Sigma_2$ i.e. $\sigma_1^2=\sigma_2^2$ ($=\sigma^2$, say). The AIC of Model H_0 is given by

$$\text{(4.1)}\quad \text{AIC}\,(H_0)=p(n+m)\log_e 2\pi+p(n+m)\log\hat{\sigma}^2+p(n+m)+2(p+1)\,,$$

where

$$\text{(4.2)}\qquad \hat{\sigma}^2=\frac{1}{p(n+m)}\left\{\sum_{i=1}^{n}(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})'(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})+\sum_{j=1}^{m}(\boldsymbol{y}_j-\hat{\boldsymbol{\mu}})'(\boldsymbol{y}_j-\hat{\boldsymbol{\mu}})\right\}$$

and

$$\text{(4.3)}\qquad \hat{\boldsymbol{\mu}}=\frac{n\bar{\boldsymbol{x}}+m\bar{\boldsymbol{y}}}{n+m}\,.$$

Next, the AIC of Model H_1 is given by

$$\text{(4.4)}\quad \text{AIC}\,(H_1)=p(n+m)\log_e 2\pi+p(n+m)\log\tilde{\sigma}^2+p(n+m)+2(2p+1)\,,$$

where

$$\text{(4.5)}\qquad \tilde{\sigma}^2=\frac{1}{p(n+m)}\left\{\sum_{i=1}^{n}(\boldsymbol{x}_i-\bar{\boldsymbol{x}})'(\boldsymbol{x}_i-\bar{\boldsymbol{x}})+\sum_{j=1}^{m}(\boldsymbol{y}_j-\bar{\boldsymbol{y}})'(\boldsymbol{y}_j-\bar{\boldsymbol{y}})\right\}\,.$$

Therefore we may summarize the condition for the preference of Model H_0 by the minimum AIC procedure as follows:

Relation 2. For $n,\ m\geqq 2$,

$$\text{(4.6)}\qquad \text{AIC}\,(H_0)-\text{AIC}\,(H_1)<0 \Longleftrightarrow t^2<p(n+m-2)\{e^{2/(n+m)}-1\}$$

where

$$t^2=\frac{(\bar{\boldsymbol{x}}-\bar{\boldsymbol{y}})'(\bar{\boldsymbol{x}}-\bar{\boldsymbol{y}})}{\dfrac{1}{n}+\dfrac{1}{m}}\Bigg/\frac{\sum_{i=1}^{n}(\boldsymbol{x}_i-\bar{\boldsymbol{x}})'(\boldsymbol{x}_i-\bar{\boldsymbol{x}})+\sum_{j=1}^{m}(\boldsymbol{y}_j-\bar{\boldsymbol{y}})'(\boldsymbol{y}_j-\bar{\boldsymbol{y}})}{p(n+m-2)}\,.$$

It is well known that the distribution of t^2 is the non-central F-distribution with $(p,\ p(n+m-2))$ degrees of freedom and with the non-centrality parameter $\zeta=\{nm/(n+m)\}\,\{(\boldsymbol{\mu}-\boldsymbol{\xi})'(\boldsymbol{\mu}-\boldsymbol{\xi})/\sigma^2\}$.

5. Examples of application of Relation 2

By applying Relation 2 to the estimation problem of the mean μ on the basis of the preliminary test of significance, we can specify the necessary significance level. First we consider the special case $p=1$.

Kitagawa [10] and Bennett [5] give the following estimator PT_α for μ

$$PT_\alpha=\begin{cases}\bar{z} & \text{if } |t_{n+m-2}|<t_\alpha\\ \bar{x} & \text{otherwise}\,,\end{cases}$$

where t_{n+m-2} has the t-distribution with $n+m-2$ degrees of freedom and t_α is the upper $100(\alpha/2)\%$ point of it. The estimator PT_α always depends on the α. From Relation 2, Akaike's information criterion gives an estimator PT for μ, namely, it implies to take $t_\alpha=\sqrt{(n+m-2)}\times\sqrt{\{\exp(2/(n+m))-1\}}$ in PT_α, where $\alpha=\Pr\{|t_{n+m-2}|\geqq t_\alpha\}$, under H_0. In Table, the values of α are given for various choices of n and m.

Table. $\alpha=\Pr\{|t_{n+m-2}|\geqq\sqrt{(n+m-2)(e^{2/(n+m)}-1)}\}$

n \ m	1	2	3	4	5	7	10	14	∞
5	0.27684	0.25480	0.23957	0.22842	0.21991	0.20780	0.19643	0.18738	0.15729
10	0.21321	0.20780	0.20334	0.19961	0.19643	0.19131	0.18573	0.18063	0.15729
15	0.19369	0.19131	0.18923	0.18738	0.18573	0.18293	0.17963	0.17635	0.15729
20	0.18426	0.18293	0.18173	0.18063	0.17963	0.17786	0.17568	0.17340	0.15729
30	0.17505	0.17446	0.17391	0.17340	0.17291	0.17203	0.17087	0.16959	0.15729
40	0.17053	0.17020	0.16989	0.16959	0.16930	0.16877	0.16806	0.16724	0.15729
50	0.16784	0.16763	0.16743	0.16724	0.16705	0.16670	0.16621	0.16564	0.15729
70	0.16480	0.16469	0.16458	0.16448	0.16439	0.16420	0.16393	0.16361	0.15729
100	0.16253	0.16248	0.16243	0.16238	0.16233	0.16223	0.16209	0.16192	0.15729
∞	0.15729	0.15729	0.15729	0.15729	0.15729	0.15729	0.15729	0.15729	0.15729

Remark 1. We can also specify the necessary significance levels of the preliminary tests (ii) and (iii). We obtain the exact significance levels for each m and n if one utilizes Table and the same representation as (3.2). Further by the similar ways stated in Section 3.1, we obtain

$$1-\alpha=\Pr\{\mathrm{AIC}(H_0)-\mathrm{AIC}(H_1')<0\}\to 0.9213\cdots, \quad \text{as } n \text{ or } m\to\infty, \tag{5.1}$$

and

$$1-\alpha=\Pr\{\mathrm{AIC}(H_0)-\mathrm{AIC}(H_1'')<0\}\to 0.9213\cdots, \quad \text{as } n \text{ or } m\to\infty. \tag{5.2}$$

Hence the significance levels are asymptotically equal to $0.0786\cdots$.

Remark 2. Han and Bancroft [6] have obtained the criterion stated below for selecting the estimator, or equivalently the α level of the preliminary test: If the experimenter does not know the size of $|\mu-\xi|/\sigma$ (say v) and is willing to accept an estimator which has a relative efficiency of PT_α to $\bar{x}$ (say $e(\alpha, v)$) of no less than e_0, the preassingned relative efficiency, then among the set of estimators with $\alpha\in A$, where $A=\{\alpha|e(\alpha, v)\geqq e_0$ for all $v\}$, the estimator is chosen to maximize $e(\alpha, v)$ with respect to α and v. But, it should be noted that since to specify α is essentially equivalent to specify e_0, this procedure is not essentially a decision procedure on the level α of the preliminary test.

Remark 3. If one uses the usual test for H_0: $\mu=\xi$ against H_1: $\mu\neq\xi$, Table, (3.2), (5.1) and (5.2) show that one should not use the significance levels 0.01, 0.05 or 0.1 if he is going to use the criterion of this paper.

Next, we consider the preliminary test estimator for $\boldsymbol{\mu}$ defined as

$$PT=\begin{cases}\bar{\boldsymbol{z}} & \text{if } t^2<p(n+m-2)\{e^{2/(n+m)}-1\}\\ \bar{\boldsymbol{x}} & \text{otherwise}.\end{cases}$$

Though, the estimator PT is an extension of the univariate estimator given by [10] and [5] in the case that the variance is unknown, PT depends only on the sample sizes n and m and the dimension p, and does not require the specification of the significance level.

Remark 4. The results given in this paper can be extended to the multiple decision problem situation if one uses some results given in Sugiura [12].

6. Discussions

Lehmann [11] (p. 61) states as follows: "The choice of a level of significance α will usually be somewhat arbitrary since in most situations there is no precise limit to the probability of an error of the first kind that can be tolerated. It has become customary to choose for α one of a number of standard values such as 0.005, 0.01 or 0.05. There is some convenience in such standardizations since it permits a reduction in certain table needed for carrying out various tests. Otherwise there appears to be no particular reason for selecting these values." This demonstrates that we can not *theoretically* select the value of a level of significance by the theory of testing hypothesis. Our ultimate objective is to estimate the unknown parameter based on the samples. The problem demands not only to decide whether the null hypothesis is rejected or not, but also to estimate the parameters. In this context the theory of testing should be discussed as a part of the theory of estimation. The preliminary test estimation procedure is a combination of a test and estimation.

If we consider the procedure of preliminary test estimation as an estimation procedure, the choice of the level of significance will not be arbitrary. The procedure of minimizing the information criterion AIC which is a measure of goodness of fit of an estimating model, uniquely determines the necessary level of significance. In general, the method stated in this paper is applicable to many other distributions.

Acknowledgements

The author is thankful to Dr. K. Takahasi for his useful discussions. His thanks are also extended to Dr. H. Akaike for reading the draft version of the paper, and to the referee for kindly pointing out the error and his valuable comments on the paper.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory*, B. N. Petrov and F. Csáki, eds., Akademiai, Budapest, 267-281.

[2] Asano, C. (1960). Estimations after preliminary tests of significance and their applications to biometrical researches, *Bull. Math. Statist.*, **9**, 1-24.

[3] Bancroft, T. A. (1944). On biases in estimation due to the use of preliminary test of significance, *Ann. Math. Statist.*, **15**, 190-204.

[4] Bancroft, T. A. and Han, C. P. (1976). On pooling means in multivariate normal distributions, *Essays in Probability and Statistics*, The Committee of EPS., 353-363.

[5] Bennett, B. M. (1952). Estimation of means on the basis of preliminary tests of significance, *Ann. Inst. Statist. Math.*, **4**, 31-43.

[6] Han, C. P. and Bancroft, T. A. (1968). On pooling means when variance is unknown, *J. Amer. Statist. Ass.*, **63**., 1333-1342.

[7] Hirano, K. (1977). Estimation procedures based on preliminary test, shrinkage technique and information criterion, *Ann. Inst. Statist. Math.*, **29**, A, 21-34.
Hirano, K. (1977). Corrections to "Estimation procedures based on preliminary test, shrinkage technique and information criterion", *Ann. Inst. Statist. Math.*, **29**, A, 317.

[8] Hirano, K. (1977). A note on the level of significance of the preliminary test in pooling variances, to appear in *J. Japan Statist. Soc.*

[9] Inagaki, N. (1977). Two errors in statistical model fitting, *Ann. Inst. Statist. Math.*, **29**, A, 131-152.

[10] Kitagawa, T. (1950). Successive process of statistical inference, *Mem. Fac. Sci. Kyushu University*, A, **15**, 139-180.

[11] Lehmann, E. L. (1959). *Testing Statistical Hypotheses*, John Wiley, New York.

[12] Sugiura, N. (1976). Finite correction of Akaike's information criterion, Reported at the meeting of Math. Soc. Japan, October.

[13] Thompson, J. R. (1968). Some shrinkage techniques for estimating the mean, *J. Amer. Statist. Ass.*, **63**, 113-122.