Title	CONTRIBUTION TO THE THEORY OF SUCCESSIVE INFERENCE FROM A PRACTICAL AND A THEORETICAL VIEW POINT
Author(s)	永田, 靖
Citation	大阪大学, 1985, 博士論文
Version Type	VoR
URL	https://hdl.handle.net/11094/2131
rights	
Note	

CONTRIBUTION TO THE THEORY OF SUCCESSIVE INFERENCE FROM A PRACTICAL AND A THEORETICAL VIEW POINT

JUNE 1985

BY

YASUSHI NAGATA

SUMMARY

This thesis is concerned with the theory of successive inference. When there is an ambiguous information about the assumed model or values of parameters included in it, a process that checks its validity and then makes inference according to the outcome is discussed. The theory is approached from the two aspects:(I) Which model is to be selected? (II) How well inference is carried out under the model?

Two problems are dealt with in the thesis. The first is pooling problem in analysis of variance. Interval estimation of cell effects in a two-way layout fixed model is discussed. The relative efficiency of the pooling method with respect to the never pooling method and the coverage probability are calculated, from which it is recommended to use the pooling method. The efficiency is defined as the ratio of the expected lengthes of the methods. Furthermore a method of testing main effect which is often used in practice by modifying the proper pooling method is justified by considering its size and power.

The second problem is to make inference about the normal mean through the framework of statistical dicision theory. Using Inagaki's loss function which evaluates both an error of model fitting and an error of estimation, admissibility and minimaxity of some procedures which use Akaike Information Criterion and maximum likelihood estimators are proved. Furthermore a notion of "data-compatible model selection" is given and discussed.

CONTENTS

CHAPTER 1

INTRODUCTION

Let $X_1,\ldots,X_n$ be random variables from one-dimensional normal distribution $N_1(\theta,\sigma^2)$, and let us consider the problem to estimate σ^2 under the qradratic loss function. When the value of θ is known (without loss of generality we can assume $\theta=0$), $\hat{\sigma}_1^2=\sum_{i=1}^n X_i^2/(n+2)$ is the best scale invariant estimator and an admissible estimator (cf. Karlin[17]). However, with unknown θ, the best location and scale invariant estimator $\hat{\sigma}_2^2=\sum_{i=1}^n(X_i-\bar{X})^2/(n+1)$ is inadmissible and is dominated by the estimator $\hat{\sigma}_0^2=\min[\hat{\sigma}_1^2,\hat{\sigma}_2^2]$ (cf. Stein[48]). The latter estimator has the following interpretation: we use Student t-test of the null hypothesis $H_0:\theta=0$ and if it is accepted at some significance level (i.e. $t^2=n\bar{X}^2/(\sum_{i=1}^n(X_i-\bar{X})^2)\leqq 1/(n+1)$), then the estimator $\hat{\sigma}_1^2$ which has a good performance with $\theta=0$ is adopted, while $\hat{\sigma}_2^2$ is used, otherwise. That is, it is a kind of preliminary test estimator. There are some cases, in which similar improvement of usual estimators can be done (for example, see Arnold[1] or Nagata[24] which will be worked out in some details in Appendix).

However, statistical procedures with the preliminary tests cannot always improve usual ones uniformly, or rather abovestated cases are somewhat exceptional. In most cases they cause trade-off, that is, they are better than usual ones in some set of the parameter space but worse in the complementary set.

The theory of statistical procedure with a preliminary test is often called "the successive process of statistical inference". (We abbreviate it as "successive inference".) It involves many issues in mathematical statistics, such as pooling problem in analysis of variance (ANOVA), Behrens-Fisher problem and variable selection problem in regression analysis. Since statistical procedures are developed based on several assumptions, it is reasonable that we should check them by using data at hand and proceed along the outcomes. From this point of view we can state that successive inference contains vast area to discuss steadily. Bibliographies in Bancroft and Han[3] and Kitagawa[19] are good sources of references.

Preliminary test is used when there is an ambiguous prior information about assumed models or parameters. If this prior information can be formulated in exact expressions, there are different approaches (e.g. Bayes statistic and see also Nagata[25] in relation to it), which will not be treated here. In this thesis we discuss the successive inference procedure from the standpoint that we check the information by preliminary test and make inference as the final purpose. Thus, we have two kinds of things to discuss:(1) Which model is chosen?;(2) Is the inference taken appropriately under the model? The example in the first paragraph is a very fortunate one and therefore it is not the problem to be discussed any more.

In Chapter 2 we treat the pooling problem in ANOVA model. Since Bancroft[2], it has been one of central parts in the theory of successive inference and many literatures have been published. Various models and various situations have been discussed, but we deal with in particular a two-way layout fixed model. First we review known results on test of main effect and point estimation of error variance. Next we deal with interval estimation of a cell mean. By calculating the relative efficiency of the expected length of the interval with pooling method to that with never pooling method and the coverage probability, we find that there is a trade-off between the two kinds of performance. We give recommendation for the practical purpose that it may be used at 25 percent significance level of the preliminary test because the method with such a level does not cause disturbance on the nominal coverage probability. Further we discuss a method used by many experimenters in practice. In testing main effect they do not use the ordinary method faithfully but modify it conveniently. We call the modified one the practical pooling method, formulate it and justify it by calculating the size and the power numerically. We find that it is very close to the ordinary one with 25 percent of significance level of the preliminary test.

In Chapter 3 we formulate the successive inference of the normal mean with known variance mathematically through the framework of statistical decision theory, in which there have been few interesting results obtained so far. We make use of the loss function introduced by Inagaki[15], which is based on Kullback-Leibler information measure and evaluates both an error of model fitting and an error of estimate, so that it is suitable for our standpoint to successive inference. Under this loss function the procedure that selects a model by using Akaike Information Criterion (AIC) and estimates parameters by maximum likelihood estimator (m.l.e.) is considered. For one-dimensional case both admissibility and minimaxity of the procedure are proved and an extension to multi-variate case is also considered. Furthermore the notion of "data-compatible model selection" is introduced and its relation to minimaxity is considered.

CHAPTER 2

POOLING PROBLEM IN ANALYSIS OF VARIANCE

2.1. General remarks

In this chapter we discuss the pooling problem in the ANOVA. It involves three kinds of procedures; test of main effect, point estimation and interval estimation. They can be treated separately, but it is more desirable for practical experimenters that total recommendation would be given. We consider a fixed model in a two-factor analysis of variance with repetition throughout this chapter. Let the full model be

$$(2.1) \qquad x_{ijk} = \mu + a_i + b_j + (ab)_{ij} + e_{ijk},$$

where $\sum_{i=1}^{l} a_i = \sum_{j=1}^{m} b_j = \sum_{i=1}^{l} (ab)_{ij} = \sum_{j=1}^{m} (ab)_{ij} = 0$, $e_{ijk} \sim NID(0, \sigma_2^2)$, $i=1,2,\ldots,l$, $j=1,2,\ldots,m$ and $k=1,2,\ldots,n$. If the effects of interactions are suspected to be nonexistent, we usually test those effects preliminarily. If the effects of interactions can be ignored, then the further analysis would be carried out according to the simplified model,

$$(2.2) \qquad x_{ijk} = \mu + a_i + b_j + e_{ijk}.$$

Otherwise, the analysis follows the original model (2.1).

There are many studies on such a preliminary test procedure on this or other models where we test doubtful error first and then further analysis is carried out on the basis of the outcome of the preliminary test, e.g.;(1)Srivastava and Gupta[46], Toyoda and Wallace[50], Hirano[13] and Ohtani and Toyoda[32] on the point estimation of the variance σ_2^2 of error term;(2) Paull[34], Bozivich, Bancroft and Hartley[6] and Mead, Bancroft and Han[22] on the size and power of test of main effect. There are also studies on the case where several doubtful errors exist; Gupta and Srivastava[10], Singh [43] and Srivastava[47] on the point estimation; Srivastava and Bozivich[44], Srivastava[45], Gupta and Srivastava[9] and Saxena and Srivastava[39] on the size and power of the test of main effect.

The pooling procedure is summarized as follows. Let V_1, V_2 and V_3 denote the mean square error (m.s.e.) of the interaction A×B, of the error term and of the main effect A, respectively. (The main effect B can be treated similarly.) Let n_i be the number of degrees of freedom (d.f.) of V_i and $E(V_i)=\sigma_i^2$ (i=1,2 and 3). Let the null hypotheses of nonexistence of the main effect and of the interaction A×B as a doubtful error be $H_0:\sigma_3^2=\sigma_2^2$ and $H_{00}:\sigma_1^2=\sigma_2^2$, respectively. In advance of testing H_0, we will test preliminarily H_{00} by using the statistic V_1/V_2. And if the hypothesis H_{00} is rejected we will use the statistic V_3/V_2 to test H_0, otherwise H_{00} is judged to hold and the statistic V_3/V will be used, where V is a pooled error defined as $V=(n_1V_1+n_2V_2)/(n_1+n_2)$.

Since it may be possible to increase the degree of freedom of error term by using this procedure, one may expect more powerful test. However, in fact the size of the main test becomes disturbed. So a question will arise as to how to control the significance level of the preliminary test by taking the size disturbance and the gain of power into consideration. Mead et al.[22] gave a warning in regard to the indiscriminate use of preliminary tests because the accuracy is not always better. On the other hand many authors have reported that if we consider that the purpose of pooling is not only to get an accurate procedure but to simplify the model if possible, preliminary test with significance level about $\alpha=0.25$ do not disturb the size so much and may be used in many practical situation.

When we estimate error variance σ_2^2, the same problem also arises. But in this case it is similar to the example in Chapter 1 and we had better use preliminary test at some significance level (cf. abovestated literatures (1)).

In Section 2.2 we will treat interval estimation. In particular we clarify the influence of pooling on interval estimation of a cell mean by calculating the coverage probability and the expected length numerically. And in Section 2.3 returning to test of main effect, we notice that experimenters do not follow the pooling procedure in this section exactly in practice and give the formulation of this procedure and the justification.

2.2. INTERVAL ESTIMATION OF A CELL MEAN UNDER A PRELIMINARY TEST

In this section we discuss the interval estimation of a cell mean following Nagata and Araki[29]. In order to estimate $\mu+a_i+b_j+(ab)_{ij}$ when the interaction effect is judged to exist, or $\mu+a_i+b_j$ otherwise, the following pooling procedure is considered:

$$(2.3) \quad \begin{cases} \bar{x}_{ij.} \pm t_1\sqrt{V_2/n}, & \text{if } H_{00} \text{ is rejected,} \\ \bar{x}_{i..} + \bar{x}_{.j.} - \bar{x}_{...} \pm t_2\sqrt{V/n_e}, & \text{otherwise,} \end{cases}$$

where t_1 and t_2 denote the upper critical point of some nominal levels with n_2 and n_1+n_2 degrees of freedom, respectively and $n_e(=lmn/(l+m-1))$ is the effective number of replication.

FORMULAS FOR COVERAGE PROBABILITY. Let P denote the coverage probability, then it is written as the sum of the probability P_1 and P_2, where

$$(2.4) \quad P_1 = \Pr\{V_1/V_2 \geqq \lambda, \ |\bar{x}_{ij.} - (\mu + a_i + b_j + (ab)_{ij})| \leqq t_1\sqrt{V_2/n}\},$$

and

$$(2.5) \quad P_2 = \Pr\{V_1/V_2 < \lambda, |\bar{x}_{i..} + \bar{x}_{.j.} - \bar{x}_{...} - (\mu + a_i + b_j)| \leqq t_2\sqrt{V/n_e}\},$$

where λ is a critical value of the preliminary test.

In order to evaluate the probabilities P_1 and P_2, we follow arguments similar to Mead et al.[22] using the Patnaik's approximation [33] in which the sum of squares n_1V_1 with a noncentral chi-square distribution (with noncentrality parameter $\eta=n_1(\sigma_1^2-\sigma_2^2)/2\sigma_2^2$) is approximated by $\sigma_2^2C_1\chi_{\nu_1}^2$ where $\nu_1=n_1+4\eta^2/(n_1+4\eta)$, $C_1=1+2\eta/(n_1+2\eta)$ and $\chi_{\nu_1}^2$ is a central chi-square distribution with ν_1 degrees of freedom. This approximation was evaluated by Gurland et al.[11] numerically in the study of Behrens-Fisher problem.

Define $\theta=\sigma_1^2/\sigma_2^2$ ($\geqq 1$) and $V_4=n\{\bar{x}_{ij.}-(\mu+a_i+b_j+(ab)_{ij})\}^2$. Then V_4/σ_2^2 is distributed as central chi-square distribution with one degree of freedom independently of V_1 and V_2. The joint density of V_1, V_2 and V_4 is

$$(2.6)\qquad \frac{1}{2^{(\nu_1+n_2+1)/2}\Gamma(1/2)\Gamma(\nu_1/2)\Gamma(n_2/2)}(n_1/C_1\sigma_2^2)^{\nu_1/2}$$

$$\times(n_2/\sigma_2^2)^{n_2/2}(1/\sigma_2^2)^{1/2}V_1^{\nu_1/2-1}V_2^{n_2/2-1}V_4^{1/2-1}$$

$$\times\exp\{-(n_1V_1/C_1+n_2V_2+V_4)/2\sigma_2^2\}.$$

Now we define the new variables $u_1=n_1V_1/n_2C_1V_2$, $u_2=V_4/n_2V_2$ and $w=n_2V_2/2\sigma_2^2$. Then the joint density of u_1, u_2 and w is

$$(2.7)\qquad \frac{1}{\Gamma(1/2)\Gamma(\nu_1/2)\Gamma(n_2/2)}w^{(\nu_1+n_2+1)/2-1}u_1^{\nu_1/2-1}u_2^{1/2-1}$$

$$\times\exp\{-w(1+u_1+u_2)\}.$$

Integrating out w, we obtain the joint density $f(u_1,u_2)$ of u_1 and u_2 as

$$(2.8) \qquad f(u_1,u_2) = \frac{\Gamma((\nu_1+n_2+1)/2)}{\Gamma(1/2)\Gamma(\nu_1/2)\Gamma(n_2/2)} u_1^{\nu_1/2-1} u_2^{1/2-1} \times(1+u_1+u_2)^{-(\nu_1+n_2+1)/2}.$$

In terms of new variables, P_1 can be rewritten as follows:

$$(2.9) \qquad P_1 = \Pr\{u_1 \geqq u_1^0,\ u_2 \leqq u_2^0\},$$

where $u_1^0=n_1\lambda/n_2C_1$ and $u_2^0=t_1^2/n_2$.

Furthermore we define $V_4=n_e\{\bar{x}_{i..}+\bar{x}_{.j.}-\bar{x}_{...}-(\mu+a_i+b_j)\}^2$ newly, and similarly we obtain P_2 as follows:

$$(2.10) \qquad P_2 = \Pr\{u_1 \leqq u_1^0,\ u_2 \leqq (1 + C_1u_1)u_3^0\},$$

where $u_3^0=t_2^2/(n_1+n_2)$. Making use of the joint density of u_1 and u_2, P_1 may be rewritten in an integral form:

$$(2.11) \qquad P_1 = \frac{\Gamma((\nu_1+n_2+1)/2)}{\Gamma(1/2)\Gamma(\nu_1/2)\Gamma(n_2/2)} \int_0^{u_2^0} \int_{u_1^0}^{\infty} u_2^{1/2-1} u_1^{\nu_1/2-1} \times(1+u_1+u_2)^{-(\nu_1+n_2+1)/2} du_1 du_2.$$

Taking the integration by parts, we get the following set of recursion formulas:

$$(2.12)\qquad P_1(a) = P_1(a+1) - \frac{\Gamma(n_2/2+a+1)}{\Gamma(n_2/2)\Gamma(a+2)} \left(\frac{u_1^0}{1+u_1^0}\right)^{a+1} (1+u_1^0)^{-n_2/2}$$

$$\times(1 - I_A(n_2/2+a+1,1/2)),$$

where $a=\nu_1/2-1$, $A=(1+u_1^0)/(1+u_1^0+u_2^0)$ and $I_X(c,d)$ denotes the incomplete Beta function defined by

$$I_X(c,d) = (1/B(c,d)) \int_0^X y^{c-1}(1-y)^{d-1}\,dy.$$

The initial value is given by

$$(2.13)\qquad P_1(0) = (1 + u_1^0)^{-n_2/2}(1 - I_A(n_2/2,1/2)).$$

For P_2, it holds that

$$(2.14)\qquad P_2 = \Pr\{u_1 \geqq 0,\ u_2 \leqq u_3^0\} - \Pr\{u_1 \geqq u_1^0,\ u_2 \leqq u_3^0\}$$

$$+ \Pr\{u_1 \leqq u_1^0,\ u_3^0 \leqq u_2 \leqq (1+C_1u_1)u_3^0\}.$$

The first two terms in the right-hand-side of (2.14) are evaluated from the formulas for P_1. Let P_{23} denote the third term, then

$$(2.15)\qquad P_{23} = \Pr\{(u_2-u_3^0)/C_1u_3^0 \leqq u_1 \leqq u_1^0,\ u_3^0 \leqq u_2 \leqq (1+C_1u_1^0)u_3^0\}.$$

The corresponding recursion formulas for P_{23} are

$$(2.16)\quad P_{23}(a) = P_{23}(a+1) + \frac{\Gamma((n_2+1)/2+a+1)}{\Gamma(1/2)\Gamma(n_2/2)\Gamma(a+2)}\Big[\int_{u_3^0}^{B} u_2^{1/2-1}$$

$$\times\{u_1^{0\,a+1}(1+u_2+u_1^0)^{-((n_2+1)/2+a+1)}$$

$$-\left(\frac{u_2-u_3^0}{C_1u_3^0}\right)^{a+1}\left(1+u_2+\frac{u_2-u_3^0}{C_1u_3^0}\right)^{-((n_2+1)/2+a+1)}\}du_2\Big],$$

where $a=\nu_1/2-1$ and $B=(1+C_1u_1^0)u_3^0$. The initial value is

$$(2.17)\quad P_{23}(0) = \frac{\Gamma((n_2+1)/2)}{\Gamma(1/2)\Gamma(n_2/2)}\int_{u_3^0}^{B} u_2^{1/2-1}\Big[\left(1+u_2+\frac{u_1-u_3^0}{C_1u_3^0}\right)^{-(n_2+1)/2}$$

$$-(1+u_2+u_1^0)^{-(n_2+1)/2}\Big]\,du_2.$$

The probability P can be evaluated by using these recursion formulas. Note that a in (2.12) and (2.16) is not necessarily an integer. Figures 1 and 2 are drawn by connecting smoothly several points corresponding to integral values for a.

FORMULAS FOR THE EXPECTED LENGTH OF CONFIDENCE INTERVAL. We call the estimation procedure with confidence limits (2.3) sometimes-pooling procedure (SPP). On the other hand the following two procedures not incorporating pooling procedure are often used. One is the procedure with the confidence limits:

$$(2.18)\quad \bar{x}_{ij.} \pm t_1\sqrt{V_2/n},$$

which is used in the situation where interaction is always regarded to exist. We call it never-pooling procedure I (NPP-I). The other has the following confidence limits:

$$(2.19)\quad \begin{cases} \bar{x}_{ij.} \pm t_1\sqrt{V_2/n}, & \text{if } H_0 \text{ is rejected,} \\ \bar{x}_{i..} + \bar{x}_{.j.} - \bar{x}_{...} \pm t_1\sqrt{V_2/n_e}, & \text{otherwise.} \end{cases}$$

In this procedure one judges the existence or the nonexistence of the interactions by the preliminary test, and reflects the outcome only on the point estimation, but does not pool the doubtful error term into the original error term. We call it never-pooling procedure II (NPP-II). The feature of these two procedures is that they give the interval with the exact prescribed nominal probability. And NPP-II is always shorter than NPP-I since $n_e > n$. Now we compare the expected length of SPP with that of NPP-II, using as a criterion the relative efficiency of expected length(%):

(2.20) R = 100{(expected length of the confidence interval of NPP-II) - (expected length of the confidence interval of SPP)}/(expected length of the confidence interval of NPP-II) .

Let $f(u_1)$ be the density function of u_1, and $f(w|u_1)$ be the conditional density function of w for given u_2. By virtue of (2.7) and (2.8) the numerator of R (say Num) is written as

$$(2.21)\quad \text{Num} = \int_0^{u_1^0}\Big[\int_0^\infty\{2t_1\sqrt{2\sigma_2^2 w/n_2 n_e} - 2t_2\sqrt{2\sigma_2^2(1+C_1u_1)w/(n_1+n_2)n_e}\}$$

$$\times f(w|u_1)\ dw\Big]f(u_1)\ du_1\times 100$$

$$= \frac{2\sqrt{2\sigma_2^2}\Gamma((\nu_1+n_2+1)/2)n_1^{\nu_1/2}n_2^{n_2/2}}{\sqrt{n_e}\Gamma(\nu_1/2)\Gamma(n_2/2)C_1^{\nu_1/2}}$$

$$\times\int_0^\lambda(t_1-t_2(\frac{n_1y+n_2}{n_1+n_2})^{1/2})y^{\nu_1/2-1}(n_1y/C^1+n_2)^{-(\nu_1+n_2+1)/2}$$

$$\times dy\times 100.$$

The denominator of R (say Den) is

$$(2.22)\quad \text{Den} = \int_{u_1^0}^\infty\int_0^\infty\{2t_1\sqrt{2\sigma_2^2 w/n_2 n}\}f(w|u_1)\ dwf(u_1)\ du_1$$

$$+\int_0^{u_1^0}\int_0^\infty\{2t_1\sqrt{2\sigma_2^2 w/n_2 n_e}\}f(w|u_1)\ dwf(u_1)\ du_1$$

$$= 2t_1\sqrt{2\sigma_2^2/n_1 n_e}\ \frac{\Gamma((\nu_1+n_2+1)/2)}{\Gamma(\nu_1/2)\Gamma(n_2/2)}\ \Big[\sqrt{n_e/n}\frac{\Gamma(\nu_1/2)\Gamma((n_2+1)/2}{\Gamma((\nu_1+n_2+1)/2)}$$

$$+(1-\sqrt{n_e/n})\frac{n_1^{\nu_1/2}n_2^{(n_2+1)/2}}{C_1^{\nu_1/2}}\int_0^\lambda y^{\nu_1/2-1}$$

$$\times(n_1y/C_1+n_2)^{-(\nu_1+n_2+1)/2}\quad dy\Big].$$

Then R is described

$$(2.23)\qquad R = \frac{n_1^{\nu_1/2}\, n_2^{(n_2+1)/2}}{C_1^{\nu_1/2}} \int_0^{\lambda} (1-(t_2/t_1)(\frac{n_1 y+n_2}{n_1+n_2})^{1/2})\, y^{\nu_1/2}$$

$$\times (n_1 y/C_1+n_2)^{-(\nu_1+n_2+1)/2}\, dy \times 100$$

$$/[\sqrt{n_e/n}\, \frac{\Gamma(\nu_1/2)\Gamma((n_2+1)/2)}{\Gamma((\nu_1+n_2+1)/2)} + (1-\sqrt{n_e/n}) \frac{n_1^{\nu_1/2}\, n_2^{(n_2+1)/2}}{C_1^{\nu_1/2}}$$

$$\times \int_0^{\lambda} y^{\nu_1/2-1} (n_1 y/C_1+n_2)^{-(\nu_1+n_2+1)/2}\, dy].$$

From (2.23) we can get the numerical results for the relative efficiency.

SOME THEOREMS. For fixed θ we give some theorems on the behavior of coverage probability and interval length as a function of λ. For the coverage probability we consider the volume $D(\lambda)$ defined as $D(\lambda)$=(coverage probability of NPP-II)-(coverage probability of SPP). Note that (coverage probability of NPP-II)=$1-\tau$, where τ is a prescribed constant. $D(\lambda)$ is written as

$$(2.24)\qquad D(\lambda) = \Pr\{u_1 \leq u_1^0,\ u_2 \leq u_2^0\} - \Pr\{u_1 \leq u_1^0,\ u_2 \leq (1+C_1u_1)u_3^0\}$$

$$= \int_0^{u_1^0} [\int_0^{u_2^0} - \int_0^{(1+C_1u_1)u_3^0}]\, f(u_1,u_2)\, du_2 du_1.$$

Recalling that $u_1^0 = n_1\lambda/n_2C_1$, $D(\lambda)$ may be differentiated in term of λ as follows.

$$(2.25)\quad dD(\lambda)/d\lambda = (n_1/n_2C_1)\left[\int_0^{u_2^0} - \int_0^{(1+n_1\lambda/n_2)u_3^0}\right]f(u_1^0,u_2)\,du_2.$$

Let λ_0 denote a λ which satisfies the equation; $u_2^0=(1+n_1\lambda/n_2)u_3^0$, then $D(\lambda)$ is zero at $\lambda=\lambda_0$. Now we obtain the following theorem.

THEOREM 2.1. *$D(\lambda)$ attains its maximum at $\lambda=\lambda_0=\{(n_1+n_2)t_1^2-n_2t_2^2\}/n_1t_2^2$, and is monotone increasing when $\lambda<\lambda_0$ and monotone decreasing when $\lambda>\lambda_0$.*

As for the expected length, we define $G(\lambda)$ as $G(\lambda)$=(expected length of the confidence interval of NPP-II)-(expected length of the confidence interval of SPP)=(Numerator of (2.23)), differentiate it and obtain the following theorem.

THEOREM 2.2. *$G(\lambda)$ attains its maximum at $\lambda=\lambda_0$, and is monotone increasing when $\lambda<\lambda_0$ and monotone decreasing when $\lambda>\lambda_0$.*

These theorems imply that there is trade-off concerning the coverage probability and the expected length when we use SPP. We next fix the significance of the preliminary test and observe the extent of the trade-off numerically.

NUMERICAL RESULTS AND DISCUSSION. The coverage probability and the relative efficiency of expected length are obtained from the derived formulas. Numerical results for the two cases $(n_1,n_2)=(6,12)$ and $(12,40)$ which correspond to the model (2.1) with $(l,m,n)=(4,3,2)$ and $(5,4,3)$, respectively, are given in Figures 1 and 2, where α is a level of significance of the preliminary test. The coverage probability is small in a neighborhood of $\theta=1.0$, while it is conservative about the nominal level 0.95 for $\theta>1.5$. On the other hand, the relative efficiency R is negative for $\alpha=0.10$ when θ is larger than about 1.5. In other words the expected length of the interval for sometimes-pooling procedure is longer than that for never-pooling procedure. For $\alpha=0.25$, R is positive when θ is smaller than about 2, and R may be sometimes negative when θ is larger than 2 but R has only about -0.1% as minimum value. And for $\alpha=0.50$ R is positive for all θ. The behaviors of coverage probability and expected length for $\alpha=0.25$ are very close to those for $\alpha=0.50$. We may deduce that the sometimes-pooling procedure is robust for the level of significances from 0.25 to 0.50. Computation was carried out at other values of $(l,m,n)=(3,3,2),(5,4,2),(5,5,2),(3,3,3),(4,3,3)$ and $(5,5,3)$ and similar results were obtained. We can see that the larger the degrees of freedom are, the less the differences of the expected length get.

Pooling is a statistical procedure in which we start from a full model, take a process of model-building and then make the main test or estimation. Hence if we are not interested in model-building, it is appropriate to use NPP-I and we don't have to discuss any more. In the process of model-building we consider that SPP is more reasonable than NPP-II. And fortunately the loss of coverage probability of SPP is just slight. By considering the trade-off of the loss of

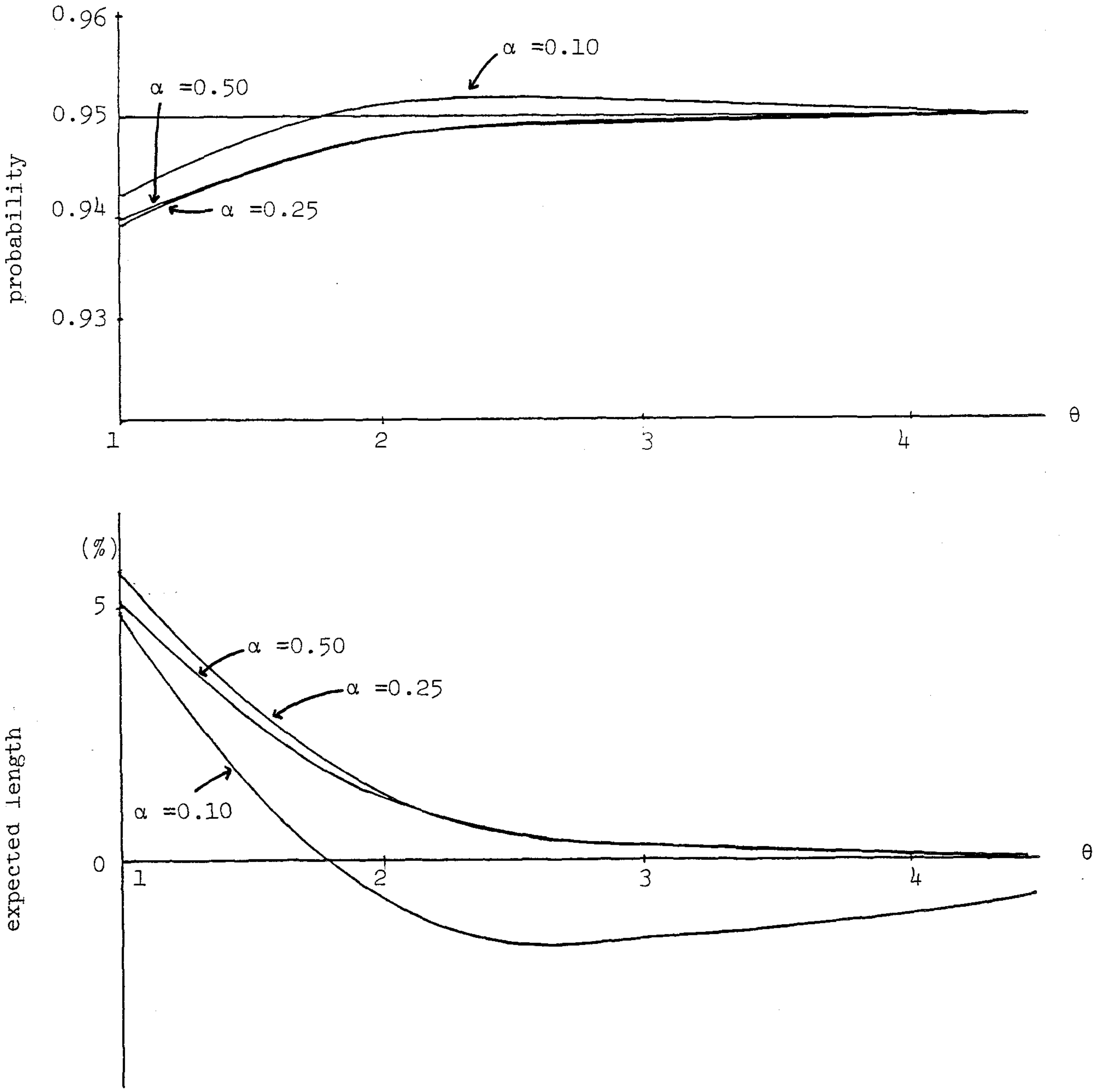

Figure 1

Coverage probability curve and relative efficiency curve of expected length for $(n_1,n_2)=(6,12)$.

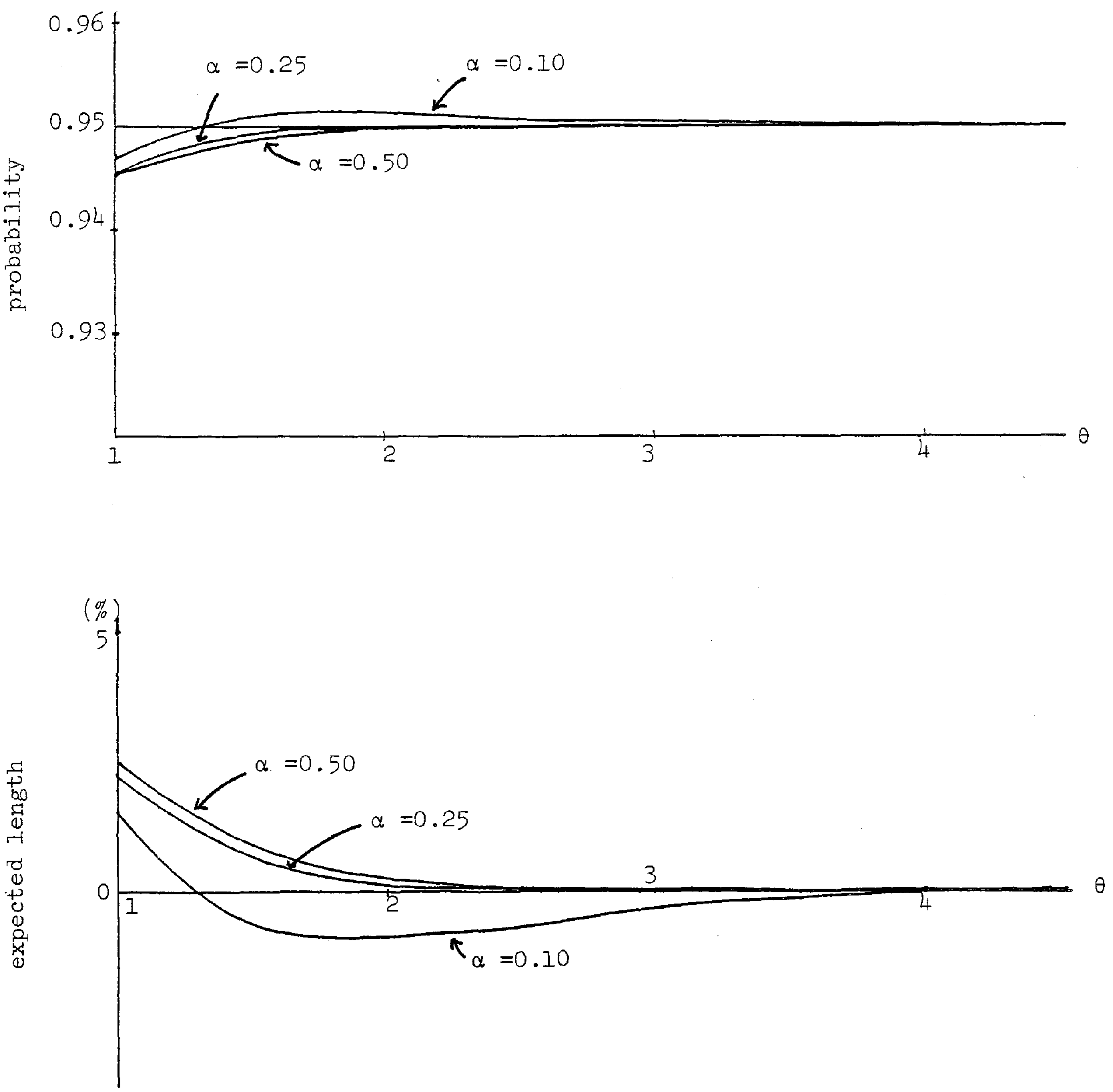

Figure 2

Coverage probability curve and relative efficiency curve of expected length for $(n_1,n_2)=(12,40)$.

coverage probability and the relative efficiency of expected length, it is desirable to set α in the range from 0.25 to 0.50. We recommend SPP with $\alpha=0.25$ as the significance level of the preliminary test, for a smaller significance level will increase the possibilities of pooling the doubtful error and working with simpler model.

2.3. Practical pooling method

Nagata and Araki[30] noticed that many experimenters usually do not apply the pooling method defined in Section 2.1, which will be called PM, precisely. They modify the PM in practice unconsciously as follows. They make the ANOVA table on the basis of the full model (cf. ANOVA TABLE I), and they will test the main effect A as well as doubtful error

ANOVA TABLE I

Source of variation	d.f.	m.s.e.	test statistic	e.m.s.
Treatment A	n_3	V_3	V_3/V_2	$\sigma_3^2=\sigma_2^2(1+2\eta_3/n_3)$
Doubtful error A×B	n_1	V_1	V_1/V_2	$\sigma_1^2=\sigma_2^2(1+2\eta_1/n_1)$
Error	n_2	V_2		σ_2^2

A×B. If the hypothesis H_{00} is accepted, they will pool A×B and error terms and test H_0 by making the pooled ANOVA table (cf. ANOVA TABLE II). At this time there is a possibility that the outcome of the test of the main effect is significant in ANOVA TABLE I but not significant in ANOVA TABLE II. In such a case experimenter who has made two ANOVA tables will decide that the main effect is "significant", but if PM was

ANOVA TABLE II

Source of variation	d.f.	m.s.e.	test statistic
Treatment A	n_3	V_3	V_3/V
Error	n_1+n_2	V	

taken, opposite result would be obtained. This testing procedure corresponds to the method that when the main effect A is significant in ANOVA TABLE I, experimenter does not have to test A further. And if A×B can be pooled, the modified model (2.2) should be used with the sole object of estimation. (If A is not significant in ANOVA TABLE I and A×B can be pooled, we should test A in ANOVA TABLE II.) This testing procedure will be called "practical pooling method" (PPM).

FORMULAS FOR THE SIZE AND POWER. Let $F(p,q;\tau)$ denote the upper 100τ percent point of central F distribution with (p,q) degrees of freedom and τ be a prescribed significance level. PPM rejects the null hypothesis $H_0:\sigma_3^2=\sigma_2^2$ if either

$$(2.26) \quad \{V_3/V_2 \geqq \lambda_2\} \text{ or } \{V_3/V_2<\lambda_2,\ V_1/V_2<\lambda_1 \text{ and } V_3/V \geqq \lambda_3\},$$

where $\lambda_1=F(n_1,n_2;\alpha)$, $\lambda_2=F(n_3,n_2;\tau_1)$ and $\lambda_3=F(n_3,n_1+n_2;\tau_2)$.

Then the probability P of rejecting H_0 is the sum of the two probabilities of the disjoint events in (2.1) and is rewritten as follows;

$$(2.27) \quad P = Pr\{V_3/V_2 \geqq \lambda_2\} + Pr\{V_3/V_2 < \lambda_2, V_1/V_2 < \lambda_1 \text{ and } V_3/V \geqq \lambda_3\}$$

$$= Pr\{V_1/V_2 \geqq \lambda_1 \text{ and } V_3/V_2 \geqq \lambda_2\} + Pr\{V_1/V_2 < \lambda_1 \text{ and } V_3/V \geqq \lambda_3\}$$

$$+ Pr\{V_3/V_2 \geqq \lambda_2, V_1/V_2 \leqq \lambda_1 \text{ and } V_3/V < \lambda_3\}.$$

Let P_1, P_2 and P_3 denote the first, second and third term of extreme right-hand-side of (2.27), respectively. Note that the probability of rejecting H_0 by PM is P_1+P_2, and that both the size and power of PPM are, therefore, larger by P_3 than those of PM. To evaluate the probabilities we take parallel arguments to Mead et al.[22] using the Patnaik's[33] approximation as in Section 2.2. Each of the sum of squares n_iV_i (i=1 and 3) with a noncentral chi-square distribution and with n_i degrees of freedom and noncentral parameter η_i shown in ANOVA TABLE I can be approximated by $\sigma_2^2 C_i \chi_{\nu_i}^2$, where $\nu_i = n_i + 4\eta_i^2/(n_i+4\eta_i)$, $C_i = 1+2\eta_i/(n_i+2\eta_i)$ and $\chi_{\nu_i}^2$ is a central chi-square distribution with ν_i degrees of freedom. Define the variables $u_1 = n_3V_3/(n_2V_2C_3)$, $u_2 = n_1V_1/(n_2V_2C_1)$ and $w = n_2V_2/(2\sigma_2^2)$, and then the P_i's are reduced as follows;

$$(2.28) \quad P_1 = Pr\{u_1 \geqq u_1^0 \text{ and } u_2 \geqq u_2^0\}$$

$$(2.29) \quad P_2 = Pr\{u_1 \geqq u_3^0(C_1u_2+1) \text{ and } u_2 \leqq u_2^0\}$$

$$(2.30) \qquad P_3 = \Pr\{u_1^0 \leqq u_1 \leqq u_3^0(C_1u_2+1) \text{ and } u_2 \leqq u_2^0\},$$

where $u_1^0=n_3\lambda_2/(n_2C_3)$, $u_2^0=n_1\lambda_1/(n_2C_1)$ and $u_3^0=n_3\lambda_3/\{(n_1+n_2)C_3\}$. To evaluate P_3 we consider the following three exclusive and exhaustive cases;

(i) Case 1 $[u_1^0 \leqq u_3^0$, i.e. $\lambda_2/n_2 \leqq \lambda_3/(n_1+n_2)]$. In this case P_3 is written as

$$(2.31) \qquad P_3 = \Pr\{u_1 \geqq u_1^0 \text{ and } u_2^0 \geqq 0\} - \Pr\{u_1 \geqq u_1^0 \text{ and } u_2 \geqq u_2^0\}$$

$$- \Pr\{u_1 \geqq u_3^0(C_1u_2+1) \text{ and } u_2 \leqq u_2^0\}$$

$$= \Pr\{u_1 \geqq u_1^0 \text{ and } u_2 \geqq 0\} - P_1 - P_2$$

and we obtain $P=P_1+P_2+P_3=\Pr\{u_1 \geqq u_1^0 \text{ and } u_2 \geqq 0\}$. Therefore this case corresponds to the never pooling method. But since the experimenter would usually set $\tau_1=\tau_2$ at some nominal level, it follows that $\lambda_2 \geqq \lambda_3$ and Case 1 would not arise.

(ii) Case 2 $[u_3^0 \leqq u_1^0 \leqq u_3^0(C_1u_2^0+1)$, i.e. $\lambda_3/(n_1+n_2) \leqq \lambda_2/n_2 \leqq \{\lambda_3/(n_1+n_2)\}(n_1\lambda_1/n_2+1)]$. In this case P_3 is represented as the shaded area in Figure 3. Let z denote the solution u_2 of the equation $u_1^0=u_3^0(C_1u_2+1)$, or

$$(2.32) \qquad z = (u_1^0 - u_3^0)/(u_3^0C_1).$$

P_3 is written as

(2.33) $P_3 = \Pr\{u_1 \geqq u_1^0 \text{ and } u_2 \geqq z\} - P_1 - P_2$

$$+\Pr\{u_1 \geqq u_3^0(C_1u_2+1) \text{ and } u_2 \leqq z\}.$$

Then the probability P is reduced to

(2.34) $P = \Pr\{u_1 \geqq u_1^0 \text{ and } u_2 \geqq z\} + \Pr\{u_1 \geqq u_3^0(C_1u_2+1) \text{ and } u_2 \leqq z\}.$

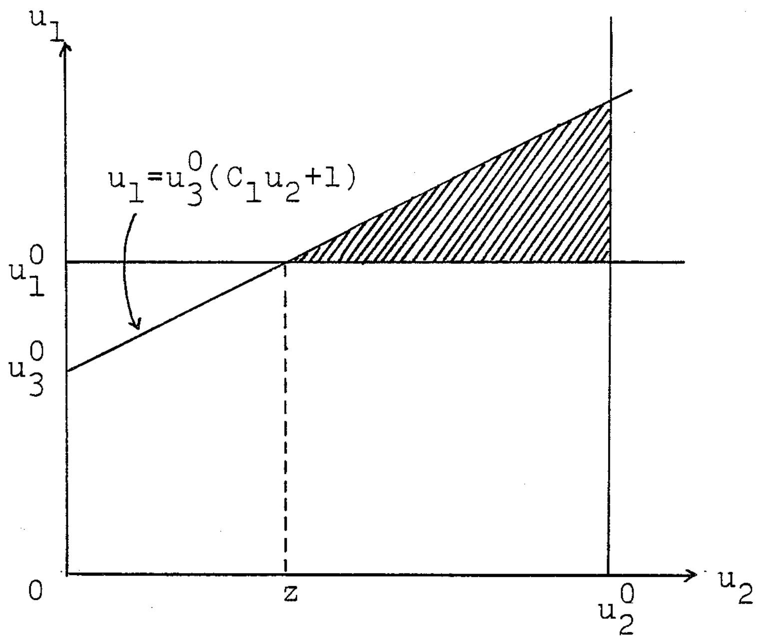

Figure 3

By (2.32) and (2.34) it can be seen that in Case 2 P does not depend on the significance level α of the preliminary test, and that this probability P is equal to the probability of rejecting H_0 under PM by setting u_2^0 at z. That is in Case 2 PPM corresponds to PM with the controlled significance level of preliminary test so as to maximize the probability of rejecting H_0. Note that if we set the significance level

α at 0.25, Case 2 is usually realized. By setting $z=u_2^0$, it holds that $\lambda_1=\{(n_2/n_1+1)\lambda_2/\lambda_3-n_2/n_1\}$ ($=\lambda_1^*$, say).

(iii) Case 3 $[u_3^0(C_1u_2^0+1)\leqq u_1^0$, i.e. $\{\lambda_3/(n_1+n_2)\}(n_1\lambda_1/n_2+1)\leqq \lambda_2/n_2]$. In this case, $P_3=0$ and P is equal to that of PM. Note that if we set the significance level α at 0.50, this case holds usually.

By above arguments it is shown that PPM is considered as either PM with $\lambda_1=\lambda_1^*$ (in Case 2) or PM (in Case 3). Table 1 compares λ_1^* with λ_1's for $\alpha=0.10$, 0.25 and 0.50 under several combinations of (n_1,n_2,n_3) and $\tau_1=\tau_2=0.05$. By Table 1 we

Table 1

$\lambda_1^*=\{(n_2/n_1+1)\lambda_2/\lambda_3-n_2/n_1\}$ and λ_1's for $\alpha=0.10$, 0.25 and 0.50.

	(n_1,n_2,n_3)	(8,15,4)	(12,20,4)	(16,25,4)	(20,10,4)	(24,10,4)	(12,21,6)	(18,28,6)	(8,18,8)
	λ_1^*	1.268	1.198	1.156	1.440	1.443	1.211	1.146	1.266
λ_1	$\alpha=0.10$	2.119	1.892	1.758	2.201	2.178	1.875	1.704	2.038
	$\alpha=0.25$	1.463	1.387	1.338	1.524	1.518	1.380	1.317	1.431
	$\alpha=0.50$	0.961	0.978	0.985	1.035	1.041	0.976	0.987	0.953

may say that λ_1^* is near to λ_1 at $\alpha=0.25$. To evaluate P in Cases 2 and 3 we shall give recursion formulas for P_1 and P_2. The formulas for P_1 are the same as in Mead et al.[22]. The formulas for P_2 are slightly different from those in Mead et al. but they enable us to evaluate P_2 for any value of $\theta_1=\sigma_1^2/\sigma_2^2$, whereas formulas in Mead et al. are applicable only at specified values of θ_1. Since those formulas can be derived similarly as in Mead et al., we will give only the results. Let $a=\nu_3/2-1$. The recursion formulas for P_1 are

$$(2.35)\qquad P_1(a+1) = P_1(a) + \frac{1}{(a+1)B(a+1,n_2/2)}\left(\frac{1-X_2}{X_1}\right)^{a+1}$$
$$\times\left(\frac{X_1+X_2-1}{X_1}\right)^{n_2/2} I_{X_1}(a+1+n_2/2,b+1),$$

where $b=\nu_1/2-1$, $X_1=(1+u_1^0)/(1+u_1^0+u_2^0)$, $X_2=(1+u_2^0)/(1+u_1^0+u_2^0)$, $B(.,.)$ denotes the complete Beta function and $I_X(.,.)$ denotes the incomplete Beta function. The initial value is

$$(2.36)\qquad P_1(0) = \left(\frac{X_1+X_2-1}{X_1}\right)^{n_2/2} I_{X_1}(n_2/2,b+1).$$

The corresponding formulas for P_2 are

$$(2.37)\qquad P_2(a) = P_2(a-1) + \frac{\Gamma((\nu_1+n_2)/2+a)}{\Gamma(\nu_1/2)\Gamma(n_2/2)\Gamma(a+1)}$$
$$\times\int_0^{u_2^0} u_2^{\nu_1/2-1}\{(C_1u_2+1)u_3^0\}^a$$
$$\times\{1+u_3^0+(C_1u_3^0+1)u_2\}^{-((\nu_1+n_2)/2+a)}\,du_2$$

and the initial value is

$$(2.38) \qquad P_2(0) = \frac{\Gamma((\nu_1+n_2)/2)}{\Gamma(\nu_1/2)\Gamma(n_2/2)} \int_0^{u_2^0} u_2^{\nu_1/2-1} \times\{1+u_3^0+(C_1u_3^0+1)u_2\}^{-(\nu_1+n_2)/2}\, du_2.$$

By using these formulas we can evaluate the size and power of PPM.

NUMERICAL RESULTS AND DISCUSSION. Let us discuss the numerical results obtained by (2.35) to (2.38). Tables 2 and 3 give illustrative examples. In Tables 2 and 3 the probabilities of rejecting H_0 by using PPM and the differences of the two probabilities of rejecting H_0 by PPM and PM under the combinations $(n_1,n_2,n_3)=(20,10,4)$ and $(8,15,4)$ with $\alpha=0.10$ and 0.25 are given. Former combination of degrees of freedom can occur in two-way layout model with unequal and proportional subclass frequencies and latter corresponds to $l=5$, $m=3$ and $n=2$ in (2.1). The values λ_1 for $\alpha=0.10$ and 0.25 result in Case 2, and the probability P obained by PPM for $\alpha=0.10$ is equal to that for $\alpha=0.25$. We are interested in the power gain and size increase by PPM compared with PM and it is reflected in the value P_3. For $\alpha=0.50$ the probability P by PPM coincides with that by PM and we will leave the discussion on this case to other references.

In Tables 2 and 3 the first row on each cell gives the probability of rejecting H_0 by PPM in Case 2, the second row

Table 2

The probabilities of rejecting H_0 by using PPM (first row in each cell) and differences of the two probabilities of rejecting H_0 by PPM and PM (second row for $\alpha=0.10$ and third row for $\alpha=0.25$) for $(n_1,n_2,n_3)=(20,10,4)$ and $\tau_1=\tau_2=0.05$. $\theta_1=\sigma_1^2/\sigma_2^2$ and $\theta_3=\sigma_3^2/\sigma_2^2$.

(%)

θ_1 \ θ_3	1.000	2.336	3.414	4.436	5.449
1.000	7.452 0.362 0.012	41.412 1.075 0.039	65.826 1.051 0.041	82.278 0.758 0.032	91.693 0.450 0.020
1.500	5.501 0.259 0.006	32.654 1.505 0.043	55.391 1.997 0.060	73.260 1.890 0.061	85.327 1.450 0.049
2.000	5.091 0.123 0.002	29.465 1.225 0.027	50.590 2.029 0.049	68.186 2.320 0.059	81.011 2.117 0.056
2.500	5.015 0.045 0.000	28.476 0.743 0.014	48.767 1.481 0.028	65.905 1.978 0.040	78.748 2.080 0.044
3.000	5.003 0.005 0.001	28.203 0.371 0.005	48.166 0.873 0.013	65.033 1.334 0.021	77.759 1.585 0.027
3.500	5.001 0.004 0.000	28.135 0.162 0.002	47.989 0.442 0.006	64.739 0.762 0.010	77.383 1.009 0.014
4.000	5.000 0.001 0.000	28.119 0.063 0.000	47.940 0.119 0.001	64.648 0.383 0.003	77.254 0.560 0.006

Table 3

The probabilities of rejecting H_0 by using PPM (first row in each cell) and differences of the two probabilities of rejecting H_0 by PPM and PM (second row for $\alpha=0.10$ and third row for $\alpha=0.25$) for $(n_1,n_2,n_3)=(8,15,4)$ and $\tau_1=\tau_2=0.05$. $\theta_1=\sigma_1^2/\sigma_2^2$ and $\theta_3=\sigma_3^2/\sigma_2^2$.

(%)

θ_1 \ θ_3	1.000	2.336	3.414	4.436	5.449
1.000	6.522 0.443 0.039	38.767 1.361 0.134	62.905 1.394 0.146	79.831 1.065 0.115	89.993 0.577 0.077
1.500	5.547 0.452 0.031	34.953 1.976 0.148	58.523 1.387 0.186	76.115 2.114 0.170	87.388 1.543 0.128
2.000	5.180 0.343 0.019	33.159 1.955 0.117	56.242 2.647 0.164	74.003 2.585 0.165	85.782 2.065 0.135
2.500	5.056 0.219 0.010	32.392 1.572 0.077	55.168 2.264 0.118	72.924 1.485 0.128	84.898 2.138 0.113
3.000	5.016 0.122 0.004	32.088 1.097 0.043	54.703 1.886 0.073	72.420 2.037 0.085	84.456 1.872 0.090
3.500	5.005 0.063 0.002	31.976 0.690 0.022	54.515 1.217 0.040	72.202 1.585 0.050	84.252 1.450 0.050
4.000	5.001 0.029 0.001	39.936 0.399 0.010	54.444 0.761 0.020	72.113 0.988 0.027	84.164 1.021 0.029

gives the values P_3 (i.e. the power gain when $\theta_3=\sigma_3^2/\sigma_2^2\neq 1$ and size increase when $\theta_3=1$) for $\alpha=0.10$ and the third row gives P_3 for $\alpha=0.25$. By Table 2 it can be seen that the maximal power gain by PPM is 2.302% for $\alpha=0.10$ and only 0.061% for $\alpha=0.25$. This can be expected from the fact that in Table 1 λ_1^* is close to λ_1 for $\alpha=0.25$. Thus PPM corresponds practically to PM with about $\alpha=0.25$ in this case. When $(n_1,n_2,n_3)=(8,15,4)$, λ_1^* is not so close to λ_1 for $\alpha=0.25$ compared with abovestated case, but we can see from Table 3 similar behavior. That is, the maximal power gains by PPM are 2.647% and 0.186% for $\alpha=0.10$ and 0.25, respectively. When $\alpha=0.25$, it can be seen that there is little size increase. The size increase are only 0.012% and 0.039% under the combinations $(n_1,n_2,n_3)=(20,10,4)$ and $(8,15,4)$, respectively. We have evaluated the probabilities for other combinations listed in Table 1 and we have obtained similar results. By using PPM we suffer little size disturbance and power gain seems to exceed the loss.

As a conclusion, when we would take the pooling method with the main objective of making the model simpler, PPM may be used. That is, if the main effect is significant in the first ANOVA table, we need not to make the pooled ANOVA table and should use pooled model for estimation object.

2.4. Conclusion

We have discussed the pooling procedure in a two-factor analysis of variance model from several directions in this chapter. The procedures discussed, in particular, the test of main effect and the interval estimation cause the trade-off and they are not recommended only from view points of accuracy. However, when we consider that the simpler model makes experimenters take actions more easily, it is reasonable to pool the doubtful error if possible, that is, if it does not cause much disturbance of the nominal level or the coverage probability. If we wish to treat all procedures discussed in this chapter together, we can justify the pooling procedure with the significance level of the preliminary test $\alpha=0.25$. And the PPM is also justified in the sense that it differs from the PM with $\alpha=0.25$ only slightly.

In the next chapter we will discuss the preliminary test procedure more mathematically.

CHAPTER 3

DECISION THEORETIC APPROACH TO MODEL SELECTION AND ESTIMATION PROCEDURE FOR THE NORMAL MEAN

3.1. General remarks

In this chapter we discuss the statistical decision theoretic approach to the successive inference through the inference of the normal mean. Let a random variable X follow one-dimensional normal distribution $N_1(\theta,\sigma^2)$, where θ is an unknown parameter and σ^2 is known (without loss of generality we assume $\sigma^2=1$). We wish to estimate the unknown parameter θ in the situation where we have vague information about θ that it is equal to zero. In such a case the following preliminary test estimator has been considered:

$$(3.1) \qquad d_c(X) = \begin{cases} 0, & \text{if } |X|\leqq c, \\ X, & \text{otherwise.} \end{cases}$$

It looks appealing at a glance, but there are some inadequacies. First of all it is inadmissible under many loss functions for estimation, including a quadratic loss function. For an admissible estimator must be a (proper) Bayes estimator or its limit and therefore it must be a smooth function in X. But the estimator (3.1) is not smooth in X, neither a Bayes estimator nor its limit, which implies its inadmissibility.

Secondly it is not minimax under quadratic loss function (cf. Sclove et al.[40]). Furthermore this procedure is different from the example in Chapter 1 in the sense that its relative efficiency to the usual estimator $d(X)=X$ (e.g. the ratio of mean square errors) is large in some restricted region of the parameter space but small in the complementary region. Therefore, many statisticians have come to consider that the preliminary test estimator cannot accomplish the purpose of improving inference.

Some statisticians considered the preliminary test estimation as a model selection problem, introduced some criteria and derived the optimal critical values (e.g. optimal value c in (3.1)) (see Sawa and Hiromatsu[36], Toyoda and Wallace[50] and [51]). Shibata[42] stated the following two standpoints of model selection: (I) to select the true model accurately, assuming its existence and (II) to select a model, considering the accuracy of sequent estimation. Above authors discussed the determination of optimal value c from the standpoint (II) and they chose one in the class $D_0=\{d_c(X); c\varepsilon[0,\infty)\}$, where $d_c(X)$ is defined in (3.1). The standpoint (II) may be reasonable in the situation where there are many data sets after determining the model. But the proposed model selection procedure often depends on the forms of sequent estimator. Furthermore, when the whole inference must be taken by using only one data set, we must give both "a model" and "an estimate under the model" as outputs of our statistical analysis, which necessarily requires us to consider the performances of two kinds of procedure, that is, model selection and estimation, simultaneously.

On the other hand Kitagawa[18] discussed that the preliminary test estimation does not fit the usual decision theoretical framework of estimation and that therefore it should be approached from some other directions. Following this suggestion, Cohen[7], Meeden and Arnold[23] and Stone[49] studied the admissibility of (3.1) under the suitable loss functions as hybrid problems. Their loss functions are indeed one approach to preliminary test estimation, but are not based on the idea described in the last paragraph. We will approach preliminary test estimation from the view points which have been stated through this thesis. In the next section we will define the loss function introduced by Inagaki[15] which incorporates model fitting and evaluation of an estimate simultaneously. It is based on Kullback-Leibler information measure. And we will discuss the admissibility and minimaxity of the preliminary test estimation in the subsequent sections. It is noted that we will treat the optimality in the class of whole estimators not only in the one-parameter family D_0.

3.2. INAGAKI'S LOSS FUNCTION

In this section we will describe the loss function due to Inagaki[15] based on the Kullback-Leibler information measure. Let X be a random variable with probability density function (p.d.f.) $f(x:\theta)\varepsilon F=\{f(x:\theta);\theta\varepsilon\Theta\}$, where Θ is a parameter space. Suppose that $F_\gamma=\{f_\gamma(x:\zeta);\zeta\varepsilon\Theta_\gamma\}$ is a model for F and Θ_γ a parameter space indexed by γ, and that $\zeta_\gamma(\theta)$ is defined by the following equation:

$$(3.2)\quad \int \log\{f(x:\theta)/f_\gamma(x:\zeta_\gamma(\theta))\}f(x:\theta)\,dx$$

$$:= \min_{\zeta\varepsilon\Theta_\gamma} \int \log\{f(x:\theta)/f_\gamma(x:\zeta)\}f(x:\theta)\,dx.$$

Sawa[38] called $\zeta_\gamma(\theta)$ the pseudo true parameter. That is, when we determine the model F_γ, we set $\zeta_\gamma(\theta)$ as a target of our estimation as if it were a true parameter. Inagaki's loss function has the following form:

$$(3.3)\quad L((k,d),\theta) = \log\{f(x:\theta)/f_k(x:\zeta_k(\theta))\}$$

$$+ \int \log\{f_k(y:\zeta_k(\theta))/f_k(y:\zeta_k(d))\}f_k(y:\zeta_k(\theta))dy,$$

where $k(X)$, $d(X)$ and $\zeta_\gamma(d(X))$ are estimators of the index γ, the unknown parameter θ and $\zeta_\gamma(\theta)$, respectively. He introduced the first term, the log-likelihood ratio, as a smooth loss for the model fitting and the second term as a loss incurred by an estimate. It is noted that this loss function (3.3) is not always nonnegative, but the first term is decomposed into the sum of the following two parts, J_0 and J_1, J_0 being common to all γ and J_1 nonnegative:

$$(3.4)\quad J_0(\theta) = \log\{f(x:\theta)/f^0(x:\theta)\}$$

$$J_1(k,\theta) = \log\{f^0(x:\theta)/f_k(x:\zeta_k(\theta))\}$$

where $f^0(x:\theta)=\sup_\gamma f_\gamma(x:\zeta_\gamma(\theta))$. The second term is, of cource, nonnegative. We can proceed as a usual statistical decision problem, since the risk function, which is the expectation of (3.3) with respect to $f(x:\theta)$, is always nonnegative.

3.3. ONE-DIMENSIONAL CASE

Now, let X follow a one-dimensional normal distribution $N_1(\theta,1)$ with p.d.f. $f(x:\theta)$. We consider two models $F_0=\{f(x:0)\}$ and $F_1=\{f(x:\theta);\theta\varepsilon(-\infty,\infty)\}$. In this problem Hirano[12] proposed the following preliminary test estimator by using Akaike information Criterion (AIC) procedure,

$$(3.5)\qquad d_{\sqrt{2}}(X) = \begin{cases} 0, & \text{if } |X|\leqq\sqrt{2}, \\ X, & \text{otherwise.} \end{cases}$$

This means that the model F_0 is selected when $|X|\leqq\sqrt{2}$, while F_1 is selected and the unknown parameter θ is estimated as X when $|X|>\sqrt{2}$. Inagaki's loss function (3.3) then becomes

$$(3.6)\qquad L((k,d),\theta) = \begin{cases} \{x^2 - (x-\theta)^2\}/2, & \text{if } k=0, \\ (d-\theta)^2/2, & \text{if } k=1. \end{cases}$$

We shall first show the admissibility of the procedure (3.5).

THEOREM 3.1 (Nagata[26]). *The procedure (3.5) is admissible under the loss function (3.6).*

We prepare the following lemma.

LEMMA 3.1. *If* θ *follows a prior distribution* $N_1(0,\tau^2)$, *then the Bayes estimator of* θ *under the loss function (3.6) is*

$$(3.7) \qquad d_\tau(X) = \begin{cases} 0, & \textit{if } |X| \leqq \sqrt{2/(2-a(\tau))}, \\ a(\tau)X, & \textit{otherwise}, \end{cases}$$

where $a(\tau)=\tau^2/(1+\tau^2)$.

PROOF. The posterior distribution of θ given X is $N_1(a(\tau)X, a(\tau))$. Let $g(\theta|X)$ be the conditional p.d.f. given X. Then the posterior risk $\rho((k,d),\tau)$ can be written as

$$(3.8) \qquad \rho((0,d),\tau) = \int L((0,d(x)),\theta)g(\theta|x)\, d\theta$$

$$= (1/2)((2a(\tau) - a(\tau)^2)x^2 - a(\tau)),$$

and

$$(3.9) \qquad \rho((1,d),\tau) = (1/2)\int (d(x) - \theta)^2 g(\theta|x)\, d\theta,$$

which is minimized by putting $d(x)=a(\tau)x$ and we obtain

$$(3.10) \qquad \min_d \rho((1,d),\tau) = a(\tau)/2.$$

Therefore comparing (3.8) with (3.10), we obtain (3.7) as a Bayes solution. Q.E.D.

Note that since $a(\tau)\to 1$ as $\tau\to\infty$, the procedure (3.5) is the limit of (3.7).

PROOF OF THEOREM 3.1. We shall use the method of Blyth [5]. By straightforward calculation the respective risk functions of $d_{\sqrt{2}}$ and d_τ become

$$(3.11)\quad R(\theta,d_{\sqrt{2}}) = \int L((k(x),d_{\sqrt{2}}(x)),\theta)f(x:\theta)\,dx$$

$$= 1/2 + (1/2)\int_{-\sqrt{2}}^{\sqrt{2}} (4x\theta - x^2 - 2\theta^2)f(x:\theta)\,dx$$

and

$$(3.12)\quad R(\theta,d_\tau) = \int L((k(x),d_\tau(x)),\theta)f(x:\theta)\,dx$$

$$= (1/2)\{a(\tau)^2 + \theta^2(a(\tau) - 1)^2\}$$

$$+ (1/2)\int_{-\sqrt{2/(2-a(\tau))}}^{\sqrt{2/(2-a(\tau))}}\{2x\theta(1 + a(\tau))$$

$$- a(\tau)^2x^2 - 2\theta^2\}f(x:\theta)\,dx.$$

Next we calculate the Bayes risk functions of $d_{\sqrt{2}}$ and d_τ with respect to $N_1(0,\tau^2)$, obtaining

$$(3.13)\quad r(\tau^2,d_{\sqrt{2}}) = 1/2 + (1/2)\int_{-\sqrt{2}}^{\sqrt{2}}\{x^2(4a(\tau) - 1 - 2a(\tau)^2)$$

$$- 2a(\tau)\}f_\tau(x)\,dx,$$

where $f_\tau(x)$ is the p.d.f. of marginal distribution of X, $N_1(0,1+\tau^2)$, and

$$(3.14)\qquad r(\tau^2,d_\tau) = a(\tau)/2 + (1/2)\int_{-\sqrt{2/(2-a(\tau))}}^{\sqrt{2/(2-a(\tau))}} \{x^2(2a(\tau)-a(\tau)^2) - 2a(\tau)\}f_\tau(x)\,dx.$$

Suppose that $d_{\sqrt{2}}$ is inadmissible. Then there exists another estimator d^* such that for all $\theta\varepsilon\Theta$

$$(3.15)\qquad R(\theta,d^*) \leqq R(\theta,d_{\sqrt{2}})$$

and for some $\theta_0\varepsilon\Theta$

$$(3.16)\qquad R(\theta_0,d^*) < R(\theta_0,d_{\sqrt{2}}).$$

Since the loss function (3.6) is continuous in θ for a fixed d, there exists ε (>0) and δ (>0) such that for all $\theta\varepsilon(\theta_0-\delta, \theta_0+\delta)$,

$$(3.17)\qquad R(\theta,d^*) < R(\theta,d_{\sqrt{2}}) - \varepsilon.$$

Therefore from (3.15) and (3.17) we obtain

$$(3.18)\qquad r(\tau^2,d_{\sqrt{2}}) - r(\tau^2,d^*) > \varepsilon\int_{\theta_0-\delta}^{\theta_0+\delta}(1/\sqrt{2\pi}\tau)\exp(-\theta^2/2\tau^2)\,d\theta$$

$$= \varepsilon K/\tau,$$

where K is a positive constant. From (3.13) and (3.14) it follows that

$$(3.19)\quad r(\tau^2,d_{\sqrt{2}}) - r(\tau^2,d_\tau) = (1/2)\int_{-\sqrt{2}}^{\sqrt{2}}\{x^2(4a(\tau) - 1 - 2a(\tau)^2) - 2a(\tau)\}f_\tau(x)\,dx - (1/2)\int_{-\sqrt{2/(2-a(\tau))}}^{\sqrt{2/(2-a(\tau))}}\{x^2(2a(\tau) - a(\tau)^2) - 2a(\tau)\}f_\tau(x)\,dx + 1/(2(1+\tau^2)).$$

Using (3.18) and (3.19) and Lebesgue's dominating convergence theorem, we have

$$(3.20)\quad \{r(\tau^2,d_{\sqrt{2}}) - r(\tau^2,d^*)\}/\{r(\tau^2,d_{\sqrt{2}}) - r(\tau^2,d_\tau)\} \to \infty$$

as $\tau\to\infty$. So the left-hand-side of (3.20) is larger than one for a large τ, which implies

$$(3.21)\quad r(\tau^2,d^*) < r(\tau^2,d_\tau).$$

This contradicts the fact that d_τ is a Bayes solution. Q.E.D.

Next we shall show the minimaxity.

THEOREM 3.2 (Nagata and Inaba[27]). *The procedure (3.5) is minimax under the loss function (3.6).*

In order to prove the theorem we need the following well-known lemma, which is stated without proof. (See Lehmann[21], Theorem 4.2.2.)

LEMMA 3.2. *Suppose that there exists a class* $\{\pi_\tau\}$ *of distributions such that the Bayes risk* $r(\tau, d_\tau)$ *of the Bayes solution* d_τ *of* θ *with respect to* π_τ *converges to some constant* r *as* τ *tends to infinity. If the risk,* $R(\theta, d_0)$*, of* d_0 *satisfies that* $R(\theta, d_0) \leqq r$ *for all* θ*, then* d_0 *is a minimax estimator of* θ.

From (3.14) the following lemma holds.

LEMMA 3.3. $r(\tau, d_\tau)$ *converges to* $1/2$ *as* $\tau \to \infty$.

PROOF. This lemma is a simple consequence of the facts that $a(\tau) \to 1$ as $\tau \to \infty$ and that

$$(3.22) \qquad |\text{the second term of } (3.14)| \leqq \frac{K}{(1+\tau^2)^{1/2}} \times \int_{-\sqrt{2}}^{\sqrt{2}} |x^2(2a(\tau) - a(\tau)^2) - 2a(\tau)|\, dx,$$

where K is a positive constant. Note that the right-hand-side of (3.22) clearly converges to zero. Q.E.D.

For the proof of the theorem we have only to show that the second term of the right-hand-side of (3.11) is non-positive. Putting

$$(3.23) \qquad g(\theta) = \int_{-\sqrt{2}}^{\sqrt{2}} \{x^2 - 2(x-\theta)^2\} \exp\{-(x-\theta)^2/2\}\, dx,$$

we shall prove the next lemma.

LEMMA 3.4. *It holds that for all* $\theta \in (-\infty, \infty)$, (3.24) $\quad g(\theta) \leqq 0.$

PROOF. Since we can easily show that $g(-\theta)=g(\theta)$, we may confine ourselves to the case where $\theta\varepsilon[0,\infty)$. Furthermore, for $\theta\varepsilon(\sqrt{2}+1,\infty)$ the quadratic function of the integrand in (3.23) is always negative in the domain of integration $(-\sqrt{2},\sqrt{2})$. Hence we have only to prove (3.24) for $\theta\varepsilon[0,\sqrt{2}+1]$. Now carrying out the integral (3.23) we obtain

$$(3.25) \quad g(\theta) = (\sqrt{2}-3\theta)\exp\{-(\sqrt{2}-\theta)^2/2\} + (\sqrt{2}+3\theta)\exp\{-(\sqrt{2}+\theta)^2/2\}$$
$$+ (\theta^2-1)\int_{-\sqrt{2}-\theta}^{\sqrt{2}-\theta} \exp(-x^2/2)\,dx.$$

We note that $g(0)=-\int_{-\sqrt{2}}^{\sqrt{2}}x^2\exp(-x^2/2)dx<0$ from (3.23) and that $g(1)=\exp(-3/2)\{(\sqrt{2}-3)\exp\sqrt{2}+(\sqrt{2}+3)\exp(-\sqrt{2})\}<0$ from (3.25). We shall separete the following two cases.

(i) Case 1 when $\theta\varepsilon[0,1)$. We shall show that $g_1(\theta)=g(\theta)/(1-\theta^2)\leqq 0$ for $\theta\varepsilon[0,1)$. We have

$$(3.26) \quad g_1(\theta) = \frac{\sqrt{2}-3\theta}{1-\theta^2}\exp\{-(\sqrt{2}-\theta)^2/2\} + \frac{\sqrt{2}+3\theta}{1-\theta^2}\exp\{-(\sqrt{2}+\theta)^2/2\}$$
$$- \int_{-\sqrt{2}-\theta}^{\sqrt{2}-\theta} \exp(-x^2/2)\,dx.$$

Differentiating (3.26) and simplifying, we obtain

$$(3.27) \quad g_1'(\theta) = A(\theta)\{(-\sqrt{2} - 2\theta + 2\sqrt{2}\theta^2 - \theta^3)$$

$$+ (-\sqrt{2} + 2\theta + 2\sqrt{2}\theta^2 + \theta^3)\exp(-2\sqrt{2}\theta)\},$$

where $A(\theta)=2\theta\exp\{-(\sqrt{2}-\theta)^2/2\}/(1-\theta^2)^2$ (>0). Clearly for $[0,1/\sqrt{2}]$ $g_1'(\theta)<0$. Since for $\theta\varepsilon[1/\sqrt{2},1)$ the expression in the second parentheses in the bracket of (3.27) is positive and $\exp(-2\sqrt{2}\theta)<\exp(-2)<1/5$, it follows

$$(3.28)\quad g_1'(\theta) < (A(\theta)/5)g_2(\theta),$$

where $g_2(\theta)=-6\sqrt{2}-8\theta+12\sqrt{2}\theta^2-4\theta^3$. Examining the behavior of $g_2(\theta)$ by differentiating, we can see that it increases in $[1/\sqrt{2},1)$. Since $g_2(1)=6\sqrt{2}-12<0$, $g_2(\theta)$ is negative in $[1/\sqrt{2},1)$. Hence from (3.28), $g_1'(\theta)<0$ for $\theta\varepsilon[1/\sqrt{2},1)$. Now as $g_1'(\theta)<0$ for $\theta\varepsilon[0,1)$, $g_1(\theta)$ is decreasing. Therefore noting $g_1(0)=g(0)<0$, we conclude that $g_1(\theta)\leqq 0$ for $\theta\varepsilon[0,1)$.

(ii) Case 2 when $\theta\varepsilon(1,\sqrt{2}+1]$. We shall show that $g_3(\theta)=g(\theta)/(\theta^2-1)\leqq 0$ for $\theta\varepsilon(1,\sqrt{2}+1]$. Since $g_3(\theta)=-g_1(\theta)$, similarly to (3.27) we obtain,

$$(3.29)\quad g_3'(\theta) = A(\theta)\{(\sqrt{2} + 2\theta - 2\sqrt{2}\theta^2 + \theta^3)$$

$$+ (\sqrt{2} - 2\theta - 2\sqrt{2}\theta^2 - \theta^3)\exp(-2\sqrt{2}\theta)\}.$$

Since for $\theta\varepsilon(1,\sqrt{2}+1]$ the second parentheses in the bracket of (3.29) is negative and $\exp(-2\sqrt{2}\theta)<\exp(-2\sqrt{2})<1/16$, it follows

(3.30) $\quad g_3'(\theta) > (A(\theta)/16)g_4(\theta),$

where $g_4(\theta)=17\sqrt{2}+30\theta-34\sqrt{2}\theta^2+15\theta^3$. Examining the behavior of $g_4(\theta)$, we can see that it has a local maximum at $\alpha=(34\sqrt{2}-\sqrt{962})/45$ and a local minimum at $\beta=(34\sqrt{2}+\sqrt{962})/45$. Since $\alpha<1<\beta<\sqrt{2}+1$ and $g_4(\beta)>0$, $g_4(\theta)$ is positive for $\theta\varepsilon(1,\sqrt{2}+1]$. Hence from (3.30), $g_3'(\theta)>0$ for $\theta\varepsilon(1,\sqrt{2}+1]$, which implies that $g_3(\theta)$ is increasing. Therefore noting $g_3(\sqrt{2}+1)=g(\sqrt{2}+1)/(2+2\sqrt{2})<0$ ($g(\sqrt{2}+1)$ is negative recalling that the quadratic function of the integrand in (3.23) is always negative in the domain of integration $(-\sqrt{2},\sqrt{2})$), we conclude that $g_3(\theta)<0$ for $\theta\varepsilon(1,\sqrt{2}+1]$. Q.E.D.

Our theorem follows immediately from Lemmas 3.2, 3.3 and 3.4.

COROLLARY 3.1. *Under the loss function (3.6), usual procedure $d(X)=X$ is also minimax but is inadmissible.*

PROOF. Corollary 3.1 is obtained clearly by the facts that the risk of $d(X)=X$ is 1/2 for all θ and it is equal to or greater than the risk of $d_{\sqrt{2}}$, (3.11), from Lemma 3.4. That is, it is dominated by $d_{\sqrt{2}}$. Q.E.D.

3.4. EXTENTION TO MULTI-DIMENSIONAL CASE

In this section let a p-dimensional random vector $\underset{\sim}{X}$ follows $N_p(\underset{\sim}{\theta},I_p)$, where $\underset{\sim}{\theta}=(\theta_1,\ldots,\theta_p)'$ is the unknown vector and I_p is the identity matrix of order p. We wish to construct an estimation procedure $\underset{\sim}{d}(\underset{\sim}{X})=(d_1(\underset{\sim}{X}),\ldots,d_p(\underset{\sim}{X}))'$ for $\underset{\sim}{\theta}$. We consider two situations (Cases A and B).

CASE A. We discuss the estimation of the unknown mean vector $\underset{\sim}{\theta}$ when there is an ambiguous information that some components of $\underset{\sim}{\theta}$ are zero. In such a case, we often decide which components are zero (model selection) and then estimate the remaining components under the model. We consider 2^p competing models, or $F_1=\{f(\underset{\sim}{x}:\underset{\sim}{\zeta}^1);\underset{\sim}{\zeta}^1=(0,0,\ldots,0)'\}$, $F_2=\{f(\underset{\sim}{x}:\underset{\sim}{\zeta}^2);\underset{\sim}{\zeta}^2=(\zeta_1,0,\ldots,0)',\zeta_1\varepsilon R^1\}$, $F_3=\{f(\underset{\sim}{x}:\underset{\sim}{\zeta}^3);\underset{\sim}{\zeta}^3=(0,\zeta_2,0,\ldots,0)',\zeta_2\varepsilon R^1\}$, $\ldots$, $F_{p+1}=\{f(\underset{\sim}{x}:\underset{\sim}{\zeta}^{p+1});\underset{\sim}{\zeta}^{p+1}=(0,0,\ldots,0,\zeta_p)',\zeta_p\varepsilon R^1\}$, $F_{p+2}=\{f(\underset{\sim}{x}:\underset{\sim}{\zeta}^{p+2});\underset{\sim}{\zeta}^{p+2}=(\zeta_1,\zeta_2,0,\ldots,0)',\zeta_1,\zeta_2\varepsilon R^1\},\ldots$, $F_{2^p}=\{f(\underset{\sim}{x}:\underset{\sim}{\zeta}^{2^p});\underset{\sim}{\zeta}^{2^p}=(\zeta_1,\zeta_2,\ldots,\zeta_p)'\varepsilon R^p\}$. Then, for example, $\underset{\sim}{\zeta}_1(\underset{\sim}{\theta})=(0,0,\ldots,0)'$ for the model F_1, $\underset{\sim}{\zeta}_2(\underset{\sim}{\theta})=(\theta_1,0,\ldots,0)'$ for the model F_2, $\underset{\sim}{\zeta}_3(\underset{\sim}{\theta})=(0,\theta_2,0,\ldots,0)'$ for the model F_3 and $\underset{\sim}{\zeta}_{p+2}(\underset{\sim}{\theta})=(\theta_1,\theta_2,0,\ldots,0)'$ for the model F_{p+2}, etc. The loss function (3.3) becomes

$$(3.31)\qquad L((k,\underset{\sim}{d}),\underset{\sim}{\theta}) = [\textstyle\sum'^k\{x_j^{\,2} - (x_j - \theta_j)^2\} + \sum''^k(d_j(\underset{\sim}{x}) - \theta_j)^2]/2,$$

where for any chosen model F_k, $\sum'^k$ means the summation over subscripts of components in $\underset{\sim}{\zeta}^k$ which are equal to zero, while $\sum''^k$ is the summation over the remaining subscripts. Under this loss function, we discuss the procedure with AIC for model selection and m.l.e. for estimation.

Since AIC=-2log(maximum likelihood under the model) + 2(number of free parameters of the model), AIC of the model F_γ is

$$(3.32) \qquad \mathrm{AIC}(\gamma) = C + \sum{}'^{\gamma} x_j^{\,2} + 2\sum{}''^{\gamma} 1$$

$$= C + 2p + \sum{}'^{\gamma}(x_j^{\,2} - 2),$$

where C is a constant common to all models. The model is chosen by minimizing AIC(γ), so that we have the following procedure with model selection and estimation

$$(3.33) \qquad \underset{\sim}{d}_0(\underset{\sim}{X}) = T_k\underset{\sim}{X} \quad \text{if } (1-2\,{}_kt_j)X_j^{\,2} \leqq (1-2\,{}_kt_j)2 \ (j=1,2,\ldots,p),$$

where $T_k=\mathrm{diag}({}_kt_1, {}_kt_2, \ldots, {}_kt_p)$ $(k=1,2,\ldots,2^p)$, ${}_kt_j=0$ if j-th component of $\underset{\sim}{\zeta}^k$ of the chosen model is zero, =1 otherwise. Then we obtain the following results.

THEOREM 3.3 (Nagata[28]). *The procedure (3.33) is inadmissible if $p \geqq 3$ under the loss function (3.31).*

PROOF. Theorem 3.3 is a kind of Stein problem (see James and Stein[16]) and the procedure (3.33) is dominated by the following one:

$$(3.34) \qquad \underset{\sim}{d}_*(\underset{\sim}{X}) = \begin{cases} T_k\underset{\sim}{X}, & \text{if } (1-2\,{}_kt_j)X_j^{\,2} \leqq (1-2\,{}_kt_j)2 \ (j=1,2,\ldots,p), \\ (1 - c/\underset{\sim}{X}'\underset{\sim}{X})\underset{\sim}{X}, & \text{if } X_j^{\,2}>2 \ (j=1,2,\ldots,p), \end{cases}$$

where $k=1,2,\ldots,2^p-1$ and c is a constant satisfying $0<c<2(p-2)$. Theorem 3.3 is proved by integration by parts. We consider the difference of risk functions of (3.33) and (3.34) and define the region $A=\{\underset{\sim}{x}=(x_1,x_2,\ldots,x_p)';|x_i|>\sqrt{2}\ (i=1,2,\ldots,p)\}$. It follows that

$$(3.35)\qquad R(\underset{\sim}{\theta},\underset{\sim}{d}_0) - R(\underset{\sim}{\theta},\underset{\sim}{d}_*) = (1/2)[\int_A(\underset{\sim}{x} - \underset{\sim}{\theta})'(\underset{\sim}{x} - \underset{\sim}{\theta})f(\underset{\sim}{x}:\underset{\sim}{\theta})\ d\underset{\sim}{x}$$

$$- \int_A(\underset{\sim}{x} - \underset{\sim}{\theta} - c\underset{\sim}{x}/\underset{\sim}{x}'\underset{\sim}{x})'(\underset{\sim}{x} - \underset{\sim}{\theta} - c\underset{\sim}{x}/\underset{\sim}{x}'\underset{\sim}{x})f(\underset{\sim}{x}:\underset{\sim}{\theta})\ d\underset{\sim}{x}]$$

$$=(1/2)[2\int_A(\underset{\sim}{x} - \underset{\sim}{\theta})'(c\underset{\sim}{x}/\underset{\sim}{x}'\underset{\sim}{x})f(\underset{\sim}{x}:\underset{\sim}{\theta})\ d\underset{\sim}{x} - \int_A(c^2/\underset{\sim}{x}'\underset{\sim}{x})f(\underset{\sim}{x}:\underset{\sim}{\theta})\ d\underset{\sim}{x}].$$

Now, the i-th term of the right-hand-side of (3.35) becomes, by integration by parts

$$(3.36)\qquad \int_A(x_i - \theta_i)(cx_i/\underset{\sim}{x}'\underset{\sim}{x})f(\underset{\sim}{x}:\underset{\sim}{\theta})\ d\underset{\sim}{x}$$

$$= \int_{A'}[\{c\sqrt{2}/(\underset{\sim}{x}'\underset{\sim}{x} - x_i^2 + 2)\}f(\sqrt{2}:\theta_i) + \{c\sqrt{2}/(\underset{\sim}{x}'\underset{\sim}{x}$$

$$- x_i^2 + 2)\}f(-\sqrt{2}:\theta_i)]f(\underset{\sim}{x}:\underset{\sim}{\theta})/f(x_i:\theta_i)d\underset{\sim}{x}^i$$

$$+ \int_A\{(\partial/\partial x_i)(cx_i/\underset{\sim}{x}'\underset{\sim}{x})\}f(\underset{\sim}{x}:\underset{\sim}{\theta})\ d\underset{\sim}{x}$$

$$\geqq \int_A\{(\partial/\partial x_i)(cx_i/\underset{\sim}{x}'\underset{\sim}{x})\}f(\underset{\sim}{x}:\underset{\sim}{\theta})\ d\underset{\sim}{x},$$

where $A'=\{\underset{\sim}{x}^i=(x_1,x_2,\ldots,x_{i-1},x_{i+1},\ldots,x_p)';|x_j|\geqq\sqrt{2}\ (j=1,2,\ldots,i-1,i+1,\ldots,p)\}$. Therefore we obtain from (3.35) and (3.36)

$$(3.37)\qquad R(\underset{\sim}{\theta},\underset{\sim}{d}_0) - R(\underset{\sim}{\theta},\underset{\sim}{d}_*) \geqq (1/2)\int_A\{(2pc - 4c - c^2)/\underset{\sim}{x}'\underset{\sim}{x}\}f(\underset{\sim}{x}:\underset{\sim}{\theta})d\underset{\sim}{x}>0$$

and Theorem 3.3 is proved. Q.E.D.

We remark that $\underset{\sim}{d}_*$ in (3.34) improves $\underset{\sim}{d}_0$ only in the region A, and that better procedures may be constructed by considering other regions.

We also consider minimaxity again.

THEOREM 3.4 (Nagata[28]). *The procedure (3.33) is minimax for all p under the loss function (3.31).*

To prove this theorem we shall again make use of Lemma 3.2. First we prepare the following lemma.

LEMMA 3.5. *If* $\underset{\sim}{\theta}$ *follows* $N_p(\underset{\sim}{0}, \tau^2 I_p)$, *the Bayes solution of* $\underset{\sim}{\theta}$ *under the loss function (3.31) is*

$$(3.38)\qquad \underset{\sim}{d}_\tau(\underset{\sim}{X}) = a(\tau)T_k\underset{\sim}{X}, \quad \text{if } (1-2_k t_j)X_j^2 \leqq (1-2_k t_j)2/(2-a(\tau)),$$

where T_k *and* ${}_k t_j$ *are defined after (3.33)* $(j=1,2,\ldots,p)$.

PROOF. The conditional distribution of $\underset{\sim}{\theta}$ given $\underset{\sim}{X}$ is $N_p(a(\tau)\underset{\sim}{X}, a(\tau)I_p)$ with p.d.f. $g(\underset{\sim}{\theta}|\underset{\sim}{X})$. The posterior risk $\rho((k,\underset{\sim}{d}),\tau)$ can be written as

$$(3.39)\qquad \rho((k,\underset{\sim}{d}),\tau) = (1/2)\int L((k,\underset{\sim}{d}),\underset{\sim}{\theta})g(\underset{\sim}{\theta}|\underset{\sim}{x})\, d\underset{\sim}{\theta}$$

$$= (1/2)[\textstyle\sum'^{k}\{(2a(\tau) - a(\tau)^2)x_j^2 - a(\tau)\}$$

$$+ \textstyle\sum''^{k}\int(d_j(\underset{\sim}{x}) - \theta_j)^2 g(\underset{\sim}{\theta}|\underset{\sim}{x})\, d\underset{\sim}{\theta}],$$

which is minimized by putting $d_j(\underset{\sim}{x})=a(\tau)x_j$, and we obtain

$$(3.40)\quad \min_{\underset{\sim}{d}} \rho((k,\underset{\sim}{d}),\tau) = (a(\tau)/2)[\textstyle\sum'^k\{(2 - a(\tau))x_j^2 - 1\} + \sum''^k 1]$$

$$= (a(\tau)/2)[p + \textstyle\sum'^k\{(2 - a(\tau))x_j^2 - 2\}].$$

Noting that this form is similar to that of (3.32), this lemma holds. Q.E.D.

Note that (3.38) converges to (3.33) as $\tau\to\infty$.

PROOF OF THEOREM 3.4. Equation (3.31) implies that the risk function corresponding to the procedure (3.33) is

$$(3.41)\quad R(\underset{\sim}{\theta},\underset{\sim}{d}_0) = (1/2)[\textstyle\sum_{i=1}^p\int_{-\sqrt{2}}^{\sqrt{2}}\{x_i^2 - (x_i - \theta_i)^2\}f(x_i:\theta_i)\, dx_i$$

$$+ \textstyle\sum_{i=1}^p\int_{|x_i|\geqq\sqrt{2}}(x_i - \theta_i)^2 f(x_i:\theta_i)\, dx_i]$$

$$= (1/2)[p + \textstyle\sum_{i=1}^p g(\theta_i)],$$

where

$$(3.42)\quad g(\theta_i) = \int_{-\sqrt{2}}^{\sqrt{2}}\{x_i^2 - 2(x_i - \theta_i)^2\}f(x_i:\theta_i)\, dx_i$$

and $f(x_i:\theta_i)$ is the p.d.f. of one-dimensional normal distribution $N_1(\theta_i,1)$. Furthermore the risk function of $\underset{\sim}{d}_\tau$ given by (3.38) is

$$(3.43)\quad R(\underset{\sim}{\theta},\underset{\sim}{d}_\tau) = (1/2)[\textstyle\sum_{i=1}^p\int_{-\sqrt{2/(2-a(\tau))}}^{\sqrt{2/(2-a(\tau))}}\{x_i^2$$

$$- (x_i - \theta_i)^2\}f(x_i:\theta_i)\ dx_i$$

$$+ \sum_{i=1}^{p}\int_{|x_i|\geqq\sqrt{2/(2-a(\tau))}}(a(\tau)x_i - \theta_i)^2 f(x_i:\theta_i)\ dx_i],$$

where the Bayes risk of $\underset{\sim}{d}_\tau$ is

$$(3.44)\qquad r(\tau,\underset{\sim}{d}_\tau) = pa(\tau)/2 + \sum_{i=1}^{p}\int_{-\sqrt{2/(2-a(\tau))}}^{\sqrt{2/(2-a(\tau))}}\{(2a(\tau) - a(\tau)^2)$$

$$\times x_i^2 - 2a(\tau)\}f_\tau(x_i)dx_i/2,$$

where $f_\tau(x_i)$ is p.d.f. of $N_1(0,1+\tau^2)$. Therefore we obtain from (3.44) that $r(\tau,\underset{\sim}{d}_\tau)$ converges to $p/2$ as $\tau\to\infty$. Since it was proved in Lemma 3.4 that $g(\theta_i)\leqq 0$ for all θ_i $(i=1,2,\ldots,p)$, the condition of Lemma 3.2 are satisfied and Theorem 3.4 holds. Q.E.D.

Now let us divide the sample space R^p into the following 3^p disjoint areas:

$$(3.45)\qquad \underset{\sim}{A}_m = A_1^{a_1}\times A_2^{a_2}\times\ldots\times A_p^{a_p}\ (m=1,2,\ldots,3^p),$$

where $a_i=0$ or ±1, $A_i^{-1}=\{x_i;x_i\leqq-\sqrt{2}\}$, $A_i^{0}=\{x_i;-\sqrt{2}<x_i\leqq\sqrt{2}\}$ and $A_i^{1}=\{x_i;x_i>\sqrt{2}\}$ are subsets of R^1 and $\times$ in (3.45) means Cartesian product. Observing which area a sample belongs to, we decide a model. This partition seems somewhat artificial but may be justified in the following way. For p=1 it is rational from our situation of problem and symmetry about zero to divide R^1 into three parts of forms $\{x; x \leq -c\}$, $\{x; -c < x \leq c\}$ and $\{x; x > c\}$, where $c > 0$. Furthermore if we set $c = \sqrt{2}$, generalized Bayes estimator with respect to Lebesgue measure under our loss function is well related (recall the comment following the proof of Lemma 3.5). (Even if we choose arbitrary finite value of c, Theorem 3.6 which we describe below still holds. However Lemma 3.4 may be false, so that the forthcoming Theorem 3.5 may not be generalized.) Since we consider 2^p competing models, forms of Cartesian product (3.45) are natural.

DEFINITION 3.1. *Under the above formulation, a model selection rule is said to be data-incompatible if in at least one area* $\underset{\sim}{A}_m$ *the rule selects a model of which i-th component is zero when* $a_i = \pm 1$ *for at least one index i. It is said to be data-compatible if it is not data-incompatible.*

For illustration we take the case p=2. In this case we have four models: $F_1 = \{f(\underset{\sim}{x} : \underset{\sim}{\zeta}^1); \underset{\sim}{\zeta}^1 = (0,0)'\}$, $F_2 = \{f(\underset{\sim}{x} : \underset{\sim}{\zeta}^2); \underset{\sim}{\zeta}^2 = (\zeta_1, 0)', \zeta_1 \varepsilon R^1\}$, $F_3 = \{f(\underset{\sim}{x} : \underset{\sim}{\zeta}^3); \underset{\sim}{\zeta}^3 = (0, \zeta_2)', \zeta_2 \varepsilon R^1\}$ and $F_4 = \{f(\underset{\sim}{x} : \underset{\sim}{\zeta}^4); \underset{\sim}{\zeta}^4 = (\zeta_1, \zeta_2)' \varepsilon R^2\}$. We divide R^2 into $3^2 = 9$ areas. There are 4^9 model selection rules and some of them are exhibited below.

4	3	4
2	1	2
4	3	4

(I)

4	4	4
2	2	2
4	4	4

(II)

4	4	4
2	4	4
4	3	4

(III)

4	3	4
3	1	3
4	3	4

(IV)

4	3	4
1	1	1
4	3	4

(V)

Both x_1- and x_2-axes are omitted and a number in each area indicates which model is selected. Model selection rules (I), (II) and (III) are data-compatible and (IV) and (V) are data-incompatible. Exhibit (I) corresponds to the procedure (3.33). A family of data-compatible model selection rules is quite large, since some member selects a model which allows effects for some components remain from the practical viewpoint even if the corresponding components of the data are small in absolute values (cf. Exhibit (II)). Furthermore it contains rules which do not have clear statistical implication (cf. Exhibit (III)).

Now we state following results.

THEOREM 3.5 (Nagata[31]). *Every procedure that selects a model data-compatibly and estimates the remaining parameters under the chosen model by m.l.e. is minimax under the loss function (3.31).*

To show this theorem following lemma is used.

LEMMA 3.6. *The model selection and estimation procedure* *is minimax under the loss function (3.31) if and only if its risk function is less than or equal to p/2.*

PROOF. The "if" part clearly follows from the proof of Theorem 3.4. The "only if" part is easily obtained by considering that the procedure $\underset{\sim}{d}(\underset{\sim}{X})=\underset{\sim}{X}$ (it means that one always selects the full model F_{2^p}) has a constant risk function: $R(\underset{\sim}{\theta},\underset{\sim}{d})=p/2$. Q.E.D.

PROOF OF THEOREM 3.5. Let $\underset{\sim}{d}^*$ be an arbitrary procedure that selects a model data-compatibly and estimates by m.l.e. It can be obtained by modifying the procedure (3.33) on some areas of the form (3.45). Now we assume that $\underset{\sim}{d}^*$ is a modification of the procedure (3.33) on only one area of (3.45), say $\tilde{A}_{m_1}=A_1^{a_{m_1}}\times A_2^{a_{m_2}}\times\ldots\times A_p^{a_{m_p}}$. And assume without loss of generality that, among the p superscripts a_{m_i}'s, the first k_1 are equal to zero, the succeeding k_2 are -1 and the remaining $k_3=p-k_1-k_2$ are 1. Following the procedure (3.33), one selects a model with first k_1 components zero and remaining k_2+k_3 components nonzero when $\underset{\sim}{x}\varepsilon\tilde{A}_{m_1}$, so we may assume $k_1\neq 0$. Furthermore assume that the model selection rule of $\underset{\sim}{d}^*$ is different from (3.33) concerning only the first coordinate on $\tilde{A}_{m_1}$, that is, following $\underset{\sim}{d}^*$ one selects a model with $\underset{\sim}{(\zeta_1,0,\ldots,0,\zeta_{k_1+1},\ldots,\zeta_p)'}$. Then, the risk function of (3.33) restricted to area $\tilde{A}_{m_1}$ is

$$(3.46)\qquad r_1 = (1/2)\underbrace{\int_{-\infty}^{-\sqrt{2}}\cdots\int_{-\infty}^{-\sqrt{2}}}_{k_2}\underbrace{\int_{\sqrt{2}}^{\infty}\cdots\int_{\sqrt{2}}^{\infty}}_{k_3}\underbrace{\int_{-\sqrt{2}}^{\sqrt{2}}\cdots\int_{-\sqrt{2}}^{\sqrt{2}}}_{k_1}[\textstyle\sum_{i=1}^{k_1}\{x_i^{\,2}$$

$$- (x_i - \theta_i)^2\} + \sum_{i=k_1+1}^{p}(x_i - \theta_i)^2]\Pi_{i=1}^{p} f(x_i:\theta_i)\ dx_i$$

and that of $\underset{\sim}{d}^*$ is

$$(3.47) \qquad r_1^* = (1/2)\int_{-\infty}^{-\sqrt{2}}\cdots\int_{-\infty}^{-\sqrt{2}}\int_{\sqrt{2}}^{\infty}\cdots\int_{\sqrt{2}}^{\infty}\int_{-\sqrt{2}}^{\sqrt{2}}\cdots\int_{-\sqrt{2}}^{\sqrt{2}}[(x_1 - \theta_1)^2$$

$$+ \sum_{i=2}^{k_1}\{x_i^{\,2} - (x_i - \theta_i)^2\} + \sum_{i=k_1+1}^{p}(x_i - \theta_i)^2]$$

$$\times\Pi_{i=1}^{p} f(x_i:\theta_i)\ dx_i.$$

Therefore, recalling (3.41), the risk function of $\underset{\sim}{d}^*$ is

$$(3.48) \qquad R(\underset{\sim}{\theta},\underset{\sim}{d}^*) = (1/2)[p + \sum_{i=1}^{p} g(\theta_i)] - r_1 + r_1^*$$

$$= (1/2)[p + (1 - C_1)g(\theta_1) + \sum_{i=2}^{p} g(\theta_i)],$$

where

$$(3.49) \qquad C_1 = \underbrace{\int_{-\infty}^{-\sqrt{2}}\cdots\int_{-\infty}^{-\sqrt{2}}}_{k_2}\underbrace{\int_{\sqrt{2}}^{\infty}\cdots\int_{\sqrt{2}}^{\infty}}_{k_3}\underbrace{\int_{-\sqrt{2}}^{\sqrt{2}}\cdots\int_{-\sqrt{2}}^{\sqrt{2}}}_{k_1-1}\Pi_{i=2}^{p} f(x_i:\theta_i)\ dx_i.$$

Since $0<C_1<1$ and thus $R(\underset{\sim}{\theta},\underset{\sim}{d}^*)\leqq p/2$ from Lemma 3.4, $\underset{\sim}{d}^*$ is minimax by virtue of Lemma 3.6.

When the model selection rule of $\underset{\sim}{d}^*$ implies that on $\tilde{A}_{m_1}$ one selects a model with several components of first k_1 coordinates nonzero, minimaxity of $\underset{\sim}{d}^*$ holds similarly: it only changes the coefficients of corresponding $g(\theta_i)$ just as (3.48).

Next when $\underset{\sim}{d}^*$ is a modification of (3.33) on not only $\underset{\sim}{A}_{m_1}$ but other areas, minimaxity of $\underset{\sim}{d}^*$ is also valid. In such a case coefficients of $g(\theta_i)$ changes to, say $1-C_1-C_2$. C_2 is a same type integral of C_1 but has different integral domain, so $0 \leqq 1-C_1-C_2 < 1$, which implies minimaxity of $\underset{\sim}{d}^*$. Q.E.D.

COROLLARY 3.2. *The risk function of the procedure (3.33) is smaller than or equal to that of every procedure stated in Theorem 3.5 uniformly in* θ.

PROOF. Corollary 3.2 is easily obtained from the proof of Theorem 3.5. Q.E.D.

THEOREM 3.6 (Nagata[31]). *Every procedure that selects a model data-incompatibly and estimates remaining parameters by m.l.e. is not minimax under the loss function (3.31).*

Note that the following lemma holds.

LEMMA 3.7. $g(\theta) \to 0$ *as* $\theta \to \pm\infty$, *where* $g(\theta)$ *is defined in (3.23).*

PROOF. If $\theta > \sqrt{2}$, we obtain

$$(3.50) \qquad |g(\theta)| \leqq \int_{-\sqrt{2}}^{\sqrt{2}} x^2 f(x:\theta)\, dx + 2\int_{-\sqrt{2}}^{\sqrt{2}} (x-\theta)^2 f(x:\theta)\, dx$$

$$\leqq (2/\sqrt{\pi})\{2 + (\sqrt{2}+\theta)^2\}\exp\{-(\sqrt{2}-\theta)^2/2\} \to 0,$$

as $\theta\to\infty$. The case $\theta\to-\infty$ is reduced to (3.50) by using the symmetry about zero. Q.E.D.

PROOF OF THEOREM 3.6. Let $\underset{\sim}{d}^{**}$ be an arbitrary procedure that selects a model data-incompatibly and estimates by m.l.e. Assume that $\underset{\sim}{d}^{**}$ contains a data-incompatible model selection on only one area, say $\underset{\sim}{A}_{n_1}=A_1^{a_{n_1}}\times A_2^{a_{n_2}}\times\ldots\times A_p^{a_{n_p}}$. And we may assume that the first k_1 a_{n_i}'s are -1, next k_2 a_{n_i}'s 1 and remaining $k_3=p-k_1-k_2$ a_{n_i}'s zero, and that $k_1\neq 0$ since $k_1+k_2>0$. Furthermore assume that one selects a model data-incompatibly concerning only first component on $\underset{\sim}{A}_{n_1}$, that is, a model with $(0,\zeta_2,\ldots,\zeta_{k'},0,\ldots,0)'$, where $k'\geqq k_1+k_2$ rearranging coordinates. By modifying it on this area, we can change $\underset{\sim}{d}^{**}$ to the procedure with a data-compatible model selection, say $\underset{\sim}{d}'$, whose risk function is from the proof of Theorem 3.5

$$(3.51)\qquad R(\underset{\sim}{\theta},\underset{\sim}{d}') = (1/2)[p + \textstyle\sum_{i=1}^{p} w_i g(\theta_i)],$$

where $0\leqq \underset{\sim}{w}_i\leqq 1$ $(i=1,2,\ldots,p)$. The risk function of $\underset{\sim}{d}'$ restricted to area A_{n_1} is

$$(3.52)\qquad s_1 = (1/2)\underbrace{\int_{-\sqrt{2}}^{\sqrt{2}}\cdots\int_{-\sqrt{2}}^{\sqrt{2}}}_{k_3}\underbrace{\int_{\sqrt{2}}^{\infty}\cdots\int_{\sqrt{2}}^{\infty}}_{k_2}\underbrace{\int_{-\infty}^{-\sqrt{2}}\cdots\int_{-\infty}^{-\sqrt{2}}}_{k_1}[\textstyle\sum_{i=1}^{k'}$$

$$(x_i-\theta_i)^2 + \textstyle\sum_{k'+1}^{p}\{x_i^2-(x_i-\theta_i)^2\}]$$

$$\times\Pi_{i=1}^{p} f(x_i:\theta_i)\,dx_i$$

and that of $\underset{\sim}{d}^{**}$ is

$$(3.53)\qquad s_1^* = (1/2)\int_{-\sqrt{2}}^{\sqrt{2}}\cdots\int_{-\sqrt{2}}^{\sqrt{2}}\int_{\sqrt{2}}^{\infty}\cdots\int_{\sqrt{2}}^{\infty}\int_{-\infty}^{-\sqrt{2}}\cdots\int_{-\infty}^{-\sqrt{2}}[x_1^2 -$$

$$(x_1 - \theta_1)^2 + \textstyle\sum_{i=2}^{k'}(x_i - \theta_i)^2 + \sum_{i=k'+1}^{p}$$

$$\{x_i^2 - (x_i - \theta_i)^2\}]\Pi_{i=1}^{p} f(x_i:\theta_i)\ dx_i.$$

Therefore the risk function $\underset{\sim}{d}^{**}$ is

$$(3.54)\qquad R(\underset{\sim}{\theta},\underset{\sim}{d}^{**}) = (1/2)[p + \textstyle\sum_{i=1}^{p} w_i g(\theta_i)] - s_1 + s_1^*$$

$$= (1/2)[p + \textstyle\sum_{i=1}^{p} w_i g(\theta_i)] + D_1 h(\theta_1),$$

where

$$(3.55)\qquad D_1 = (1/2)\underbrace{\int_{-\sqrt{2}}^{\sqrt{2}}\cdots\int_{-\sqrt{2}}^{\sqrt{2}}}_{k_3}\underbrace{\int_{\sqrt{2}}^{\infty}\cdots\int_{\sqrt{2}}^{\infty}}_{k_2}\underbrace{\int_{-\infty}^{-\sqrt{2}}\cdots\int_{-\infty}^{-\sqrt{2}}}_{k_1-1}\Pi_{i=2}^{p} f(x_i:\theta_i)\ dx_i$$

and

$$(3.56)\qquad h(\theta_1) = \int_{-\infty}^{-\sqrt{2}}\{x_1^2 - 2(x_1 - \theta_1)^2\} f(x_i:\theta_1)\ dx_1$$

$$= (\theta_1^2 - 1)\int_{-\infty}^{-\sqrt{2}} f(x_1:\theta_1)dx_i - (\sqrt{2} + 3\theta_1)$$

$$\times\exp\{-(\sqrt{2}+\theta_1)^2/2\}/\sqrt{2\pi}.$$

Fix the values of $\theta_2,\ldots,\theta_p$ and put $M=\sum_{i=2}^{p} w_i g(\theta_i)$. Using

Lemma 3.7, we can choose θ_1 to be negative and large absolute value, such that

$$(3.57) \quad h(\theta_1) > |w_1 g(\theta_1) + M|/D.$$

Thus there exists $\underset{\sim}{\theta}\varepsilon R^p$, such that $R(\underset{\sim}{\theta},\underset{\sim}{d}^{**})>p/2$, which implies from Lemma 3.6 that $\underset{\sim}{d}^{**}$ is not minimax.

When on A_{n_1} a model selection rule is data-incompatible concering two or more components, similar argument also works. For instance, if by $\underset{\sim}{d}^{**}$ one selects a model with $(0,0,\zeta_3,\ldots,\zeta_{k'},0,\ldots,0)'$, the risk function becomes

$$(3.58) \quad R(\underset{\sim}{\theta},\underset{\sim}{d}^{**}) = (1/2)[p + \textstyle\sum_{i=1}^{p} w_i g(\theta_i)] + D_1 h(\theta_1) + E_1 h(\theta_2).$$

E_1 depends on θ_1 but noting that $0<E_1<1$ or $|E_1 h(\theta_2)|<|h(\theta_2)|$, we can poceed just as (3.57). Furthermore when we must modify $\underset{\sim}{d}^{**}$ on several areas of forms (3.45), not only $\tilde{A}_{n_1}$, to change it the procedure with a data-compatible model selection, parallel arguments are valid. Q.E.D.

We should emphasize that the minimaxity in Theorem 3.5 holds among all model selection (not only based on (3.45)). and estimation (not only by m.l.e.) procedures. So the class of minimax procedures stated in Theorem 3.5 is, of course, not maximal. By Theorems 3.5 and 3.6 the procedures that select a model based on (3.45) and estimate by m.l.e. are separated in the sense of minimaxity according as each model selection is data-compatible or incompatible.

It is also remarked that through Definition 3.1 we consider a kind of unbiasedness of model selection rule. In the model selection theory unbiasedness seems to be intractable. Our procedures may not be unbiased in the sense of Lehmann (see Lehmann[20]) except trivial ones that select always only one model. (Another criterion of unbiasedness of model selection is given in Sawa and Takeuchi[37].)

CASE B. We consider the situation where there is an ambiguous information that $\underset{\sim}{\theta}=\underset{\sim}{0}$. We consider two models, $F_0=\{f(\underset{\sim}{x}:\underset{\sim}{0})\}$ and $F_1=\{f(\underset{\sim}{x}:\underset{\sim}{\theta});\underset{\sim}{\theta}\varepsilon R^p\}$. In this problem Hirano[12] discussed the following preliminary test estimator by using AIC procedure,

$$(3.59)\qquad \underset{\sim}{d}_{2p}(\underset{\sim}{X}) = \begin{cases} \underset{\sim}{0}, & \text{if } \underset{\sim}{X}'\underset{\sim}{X} \leqq 2p, \\ \underset{\sim}{X}, & \text{otherwise.} \end{cases}$$

And Inagaki's loss function (3.3) becomes

$$(3.60)\qquad L((k,\underset{\sim}{d}),\underset{\sim}{\theta}) = \begin{cases} \{\underset{\sim}{x}'\underset{\sim}{x} - (\underset{\sim}{x} - \underset{\sim}{\theta})'(\underset{\sim}{x} - \underset{\sim}{\theta})\}/2, & \text{if } k=0, \\ (\underset{\sim}{d} - \underset{\sim}{\theta})'(\underset{\sim}{d} - \underset{\sim}{\theta})/2, & \text{if } k=1. \end{cases}$$

Under this formulation of the problem, we obtain the following result.

THEOREM 3.7 (Inaba and Nagata[14]). *The procedure (3.59) is minimax for even $p\leqq 12$ under the loss function (3.60), but not for $p=14$.*

PROOF. The Bayes solution of $\underset{\sim}{\theta}$ with respect to the prior distribution $N_p(\underset{\sim}{0},\tau^2 I_p)$ is given by

$$(3.61)\qquad \underset{\sim}{d}_\tau(\underset{\sim}{X}) = \begin{cases} \underset{\sim}{0}, & \text{if } \underset{\sim}{X}'\underset{\sim}{X} \leq 2p/(2-a(\tau)), \\ a(\tau)\underset{\sim}{X}, & \text{otherwise}, \end{cases}$$

which is derived as in Lemmas 3.1 or 3.5. The risk function of the procedure (3.59) is

$$(3.62)\qquad R(\underset{\sim}{\theta},\underset{\sim}{d}_{2p}) = p/2 + (1/2)\int_{\underset{\sim}{x}'\underset{\sim}{x}\leq 2p}\{\underset{\sim}{x}'\underset{\sim}{x} - 2(\underset{\sim}{x}-\underset{\sim}{\theta})'(\underset{\sim}{x}-\underset{\sim}{\theta})\} \times f(\underset{\sim}{x}:\underset{\sim}{\theta})\, d\underset{\sim}{x}.$$

Noting that the conditional distribution of $\underset{\sim}{\theta}$ given $\underset{\sim}{X}$ is $N_p(a(\tau)\underset{\sim}{X}, a(\tau)I_p)$, the Bayes risk of the procedure (3.60) is

$$(3.63)\qquad r(\tau,\underset{\sim}{d}_\tau) = pa(\tau)/2 + (1/2)\int_{\underset{\sim}{x}'\underset{\sim}{x}\leq 2p/(2-a(\tau))}\{\underset{\sim}{x}'\underset{\sim}{x}(2a(\tau) - a(\tau)^2) - 2pa(\tau)\} f_\tau(\underset{\sim}{x})\, d\underset{\sim}{x},$$

where $f_\tau(\underset{\sim}{x})$ is p.d.f. of $N_p(\underset{\sim}{0},(1+\tau^2)I_p)$, the marginal distribution of $\underset{\sim}{X}$. Therefore $r(\tau,\underset{\sim}{d}_\tau)\to p/2$ as $\tau\to\infty$. In order to prove the minimaxity of the procedure (3.59) by using Lemma 3.2 we must establish that the second term of the right-hand-side of (3.62) is nonpositive for all $\underset{\sim}{\theta}\varepsilon R^p$. On the other hand, noting the form of the loss function (3.60), the risk function of the usual estimator $\underset{\sim}{d}(\underset{\sim}{X})=\underset{\sim}{X}$ is

$$(3.64) \qquad R(\underset{\sim}{\theta},\underset{\sim}{d}) = p/2 \quad \text{for all } \underset{\sim}{\theta}\varepsilon R^p,$$

which establishes the minimaxity of $\underset{\sim}{d}$ for all p. Putting

$$(3.65) \qquad g(\underset{\sim}{\theta}) = \int_{\underset{\sim}{x}'\underset{\sim}{x}\leqq 2p}\{\underset{\sim}{x}'\underset{\sim}{x} - 2(\underset{\sim}{x} - \underset{\sim}{\theta})'(\underset{\sim}{x} - \underset{\sim}{\theta})\}f(\underset{\sim}{x}:\underset{\sim}{\theta})\, d\underset{\sim}{x},$$

we obtain the following lemma from the above discussion as in Lemma 3.6.

LEMMA 3.8. *A necessary and sufficient condition for the procedure (3.59) to be minimax is* $g(\underset{\sim}{\theta})\leqq 0$ *for all* $\underset{\sim}{\theta}\varepsilon R^p$.

We need the following lemma at this stage of the proof of the theorem.

LEMMA 3.9. *Suppose the random vector* $\underset{\sim}{X}$ *follows* $N_p(\underset{\sim}{0},D)$ *with p.d.f.* $f(\underset{\sim}{x}:\underset{\sim}{0},D)$. *Let S be the set of* $\underset{\sim}{x}$ *such that* $(\underset{\sim}{x}+\underset{\sim}{a})'D^{-1}(\underset{\sim}{x}+\underset{\sim}{a})\geqq c$ *for a nonnegative constant c. Then*

$$(3.66) \qquad \int_S \underset{\sim}{x}f(\underset{\sim}{x}:\underset{\sim}{0},D)\, d\underset{\sim}{x} = \underset{\sim}{a}[\Pr\{\chi^2(p;\delta)\geqq c\} - \Pr\{\chi^2(p+2;\delta)\leqq c\}]$$

and

$$(3.67) \qquad \int_S \underset{\sim}{x}\underset{\sim}{x}'f(\underset{\sim}{x}:\underset{\sim}{0},D)\, d\underset{\sim}{x} = D[1 - \Pr\{\chi^2(p+2;\delta)\geqq c\}]$$
$$- \underset{\sim}{a}\underset{\sim}{a}'[\Pr\{\chi^2(p;\delta)\leqq c\} - 2\Pr\{\chi^2(p+2;\delta)\leqq c\}$$
$$+ \Pr\{\chi^2(p+4;\delta)\leqq c\}],$$

where $\underset{\sim}{a}$ *is a p-dimensional constant vector,* $\delta := \underset{\sim}{a}'D^{-1}\underset{\sim}{a}$ *and* $\chi^2(k;\delta)$ *is a random variable of a noncentral* χ^2*-distribution with degrees of freedom k and noncentrality parameter* δ.

These equalities are described in Sen[41] without proof. They can be proved by induction with respect to p. Now, $g(\underset{\sim}{\theta})$ can be rewritten as follows by using (3.66) and (3.67):

$$(3.68)\quad g(\underset{\sim}{\theta}) = -p\Pr\{\chi^2(p+2;\underset{\sim}{\theta}'\underset{\sim}{\theta})\leqq 2p\} + \underset{\sim}{\theta}'\underset{\sim}{\theta}[-2\Pr\{\chi^2(p;\underset{\sim}{\theta}'\underset{\sim}{\theta})\leqq 2p\}$$

$$+4\Pr\{\chi^2(p+2;\underset{\sim}{\theta}'\underset{\sim}{\theta})\leqq 2p\} - \Pr\{\chi^2(p+4;\underset{\sim}{\theta}'\underset{\sim}{\theta})\leqq 2p\}]$$

$$= e^{-r}\textstyle\sum_{j=1}^{\infty}(r^j/j!)[-p\beta(p/2+1) + 2r\{-2\beta(p/2+j-1)$$

$$+ 4\beta(p/2+j) - \beta(p/2+j+1)\}],$$

where $\beta(k)=\int_0^p e^{-x}x^k dx/\Gamma(k+1)$, $\Gamma(k+1)=\int_0^\infty e^{-x}x^k dx$ and $r=\underset{\sim}{\theta}'\underset{\sim}{\theta}/2$. We shall use the following lemma in order to modify (3.68).

LEMMA 3.10. *(I) It follows (i)* $\beta(k)-\beta(k-1)=-e^{-p}p^k/k!$, *(ii)* $\beta(k)=1-e^{-p}\sum_{i=0}^{k}p^i/i!$ *and (iii)* $\beta(k-1)/\beta(k)=1+[\sum_{m=1}^{\infty}\Pi_{i=1}^{m}(p/(k+i)]^{-1}$. *(II) If* $c>1$ *and* $i\geqq cp-1$, *then* $\beta(i-1)/\beta(i)>c$. *(III) Putting* $\eta(j)=\{2(j+1)-p\}\{3\beta(p/2+j)-2\beta(p/2+j-1)\}/(j+1)!$, *(i) if* $j\geqq p-1$, *then* $\eta(j)<0$, *(ii) if* $j=p-2$ *and* $p\geqq 4$, *then* $\eta(j)<0$ *and (iii) if* $j\leqq p/2-1$, *then* $\eta(j)\leqq 0$. *(IV) Putting* $\varepsilon(j)=\eta(j)p^{j+1}$, *if* $j\geqq 3p/2+1$, *then* $|\varepsilon(j)/\varepsilon(j-1)|<2/3$. *(V) It follows* $\sum_{j=3p/2+1}^{\infty}\varepsilon(j)>2\varepsilon(3p/2)$.

PROOF. Parts (I) and (IV) are obvious. Part (II) can be proved by considering the function $\beta(i-1)-c\beta(i)$. Part (III) (i) is established by considering $c=3/2$ in (II), whereas (ii) and (iii) are obtained by considering $p/(k+i)\leqq 2/3$ for $i\geqq 6$ and $\Pi_{i=1}^{m}p/(k+i)\geqq 0$ for $m\geqq 3$, respectively, in (I) (iii). Part (V) follows from (III) and (IV). Q.E.D.

Using $\eta(j)$ defined in Lemma 3.10 (III) and rearranging we can write $g(\underset{\sim}{\theta})$ in (3.68) in the following form:

$$(3.69)\qquad g(\underset{\sim}{\theta}) = e^{-r}[-p\beta(p/2) + \textstyle\sum_{j=0}^{\infty}\eta(j)r^{j+1}].$$

Now for $p=2$ or 4, we can see that $g(\underset{\sim}{\theta})\leqq 0$ for all $\underset{\sim}{\theta}$, since $\eta(j)\leqq 0$ for all j from Lemma 3.10 (III). Therefore the procedure (3.59) is minimax for $p=2$ or 4. For $p=6,8,10$ or 12 the procedure (3.59) can be proved to be also minimax by calculating $\eta(j)$'s for $j=0,1,\ldots,3p/2$ and showing that

$$(3.70)\qquad g(\underset{\sim}{\theta}) < e^{-r}[\textstyle\sum_{j=0}^{3p/2}\eta(j)r^{j+1}] < 0.$$

This line of the argument is similar to the case for $p=14$ which is discussed below, but in the latter case it is proved that there exists some $r_0=\underset{\sim}{\theta}_0'\underset{\sim}{\theta}_0/2$ such that $g(\underset{\sim}{\theta}_0)>0$ and that therefore the procedure (3.59) is not minimax. In order to prove the fact for $p=14$ described above, we take $r_0=14$. Making use of Lemma 3.10 (V), we obtain from (3.69)

$$(3.71)\qquad g(\underset{\sim}{\theta}_0) = e^{-14}[-14\beta(7) + \textstyle\sum_{j=0}^{\infty}\varepsilon(j)]$$

$$> e^{-14}[-14\beta(7) + \textstyle\sum_{j=0}^{21}\varepsilon(j) + 2\varepsilon(21)].$$

Now let us give the following notations:

(3.72) $$A(k) = \textstyle\sum_{i=0}^{k} 14^i/i!,\ B(k) = e^{14} - A(k)\ (=e^{14}\beta(k)),$$

$$C(j) = 3B(j+7) - 2B(j+6) \text{ and } D(j) = \{\frac{2(j+1)-14}{(j+1)!}\}14^{j-1}$$

$$\times C(j)\ (=e^{14}14^{-8}\varepsilon(j)),\ (k=0,1,\ldots,28 \text{ and } j=0,1,\ldots,21).$$

Using above notations, the extreme right-hand-side of (3.71) becomes

(3.73) $$g(\underset{\sim}{\theta}_0) > e^{-28}14^8[-14^{-7}B(7) + \textstyle\sum_{j=0}^{21}D(j) + 2D(21)].$$

Here, we obtain that $-14^{-7}B(7)\geqq-0.02$, $\sum_{j=0}^{21}D(j)\geqq5.46$ and $2D(21)$ $\geqq-0.27$, exactly. Hence, the bracket of the right-hand-side of (3.73) is positive. Therefore we conclude that the procedure (3.59) is not minimax for $p=14$, and the theorem has been proved. Q.E.D.

Theorem 3.7 shows that the validity of minimaxity in Case B depends on the dimension p. (It seems false also in the case that $p\geqq14$ and is even. The assumption that p is even serves for the convenience of the proof and it may not be essential.) Considering practically, it is not reasonable to deal with so many parameters and to decide whether they are all zero or not, and therefore we may as well follow Case A or Case B in lower dimension. However, it is somewhat strange mathematically.

THEOREM 3.8 (Nagata[26]). *The procedure (3.59) is inadmissible when* $p \geq 3$ *under loss function (3.60).*

PROOF. The procedure (3.59) is dominated by the following procedure:

$$(3.74) \qquad \underset{\sim}{d}_1(\underset{\sim}{X}) = \begin{cases} \underset{\sim}{0}, & \text{if } \underset{\sim}{X}'\underset{\sim}{X} \leq 2p, \\ (1 - c/\underset{\sim}{X}'\underset{\sim}{X})\underset{\sim}{X}, & \text{otherwise,} \end{cases}$$

where c is a constant satisfying $0<c<2(p-2)$. This fact can be shown similarly to the proof of Theorem 3.3. Q.E.D.

ACKNOWLEDGMENT

I wish to express my gratitude to Professor M.Okamoto of Osaka University for his recommendation to write this thesis and his encouragement. I am grateful to Professor N.Inagaki of Osaka University for helpful suggestions and essential comments. I am also indebted to Drs. S.Shirahata and Y.Toyooka and Mr. H.Nagahata of Osaka University for their suggestions and comments with instructive encouragement. Finally, I would like to thank Mr. T.Araki of Osaka University and Mr. T. Inaba of Nagoya University for acknowledging me to refer the results in joint works.

AUTHOR'S ADDRESS

Department of Applied Mathematics,
Faculty of Engineering Science,
Osaka University,
Toyonaka, Osaka,
JAPAN.

APPENDIX

In this appendix, we will give a similar result to the example in the Chapter 1 following Nagata[24]. We consider the estimation of the Pareto parameter of the Pareto distribution, which is often used for the distribution of income and has the following density;

$$\text{(A.1)} \quad f(x) = (ak^a/x^{a+1})I(x>k), \quad a,k>0,$$

where $I(x>k)$ is the indicator function of the set $\{x;x>k\}$, a being called the Pareto parameter and k the cut-off parameter.

Let $X_1,\ldots,X_n$ follow the distribution (A.1). In the following we discuss the optimal estimation for the Pareto parameter a with the quadratic loss function $L(d,a)=(d/a-1)^2$ and its admissibility in two cases: k is known (Case 1) and k is unknown (Case 2).

CASE 1. Clearly $\sum_{i=1}^{n}\log x_i$ is a sufficient statistic for a. Now we consider the class of estimators for a, $\{c_w=w/(\sum_{i=1}^{n}\log x_i - n\log k);w>0\}$, which contains the m.l.e. c_n as well as the minimum variance unbiased estimator due to Baxter[4], and derive the value of w which attains the minimum risk with quadratic loss.

According to Baxter[4], $T=2an/c_n$ is a χ^2-variable with

$2n$ degrees of freedom. Hence the risk of the estimator $c_w = w/(\sum_{i=1}^n \log x_i - n\log k)$ is given by

$$\text{(A.2)} \quad R(c_w, a) = E(c_w/a - 1)^2$$

$$= E((w/n)c_n/a - 1)^2$$

$$= (1/(n-1)(n-2))[(w - (n-2))^2 + n - 2], \text{ if } n>2,$$

so $c_{n-2} = (n-2)/(\sum_{i=1}^n \log x_i - n\log k)$ attains the minimum risk uniformly in a in the class under consideration. We note that the risks of c_{n-2}, c_{n-1} and c_n become larger in this order.

Now we remark that c_{n-2} is admissible by the proposition of Ghosh and Singh[8] or Ralescu,D. and Ralescu,S.[35]: *If $(x_1,\ldots,x_n)$ is a random sample from the distribution with the density $a\exp(-ax)I(x>0)$, then $(n-2)/\sum_{i=1}^n x_i$ is an admissible estimator for a.* In fact if we transform to $y_i = \log(x_i/k)$ $(i=1,\ldots,n)$, then $(y_1,\ldots,y_n)$ is a random sample from the above exponential distribution and hence $c_{n-2} = (n-2)/(\sum_{i=1}^n \log x_i - n\log k)$ is admissible.

CASE 2. When we put $s = (1/n)\sum_{i=1}^n \log x_i$ and $k' = \min(x_1,\ldots,x_n)$, (s,k') is a sufficient statistic for (a,k). If we transform to $y_i = \log x_i$ $(i=1,\ldots,n)$, $(y_1,\ldots,y_n)$ is a random sample from the exponential distribution with the density; $a\exp(-a(y-\log k)) \times I(y>\log k)$. So this problem is invariant under the transformation $y_i \to by_i + c$ $(i=1,\dots,n)$ and $(a, \log k) \to (a/b, b\log k + c)$, $0<b<\infty$, $-\infty<c<\infty$. An estimator d for a should be determined on the basis of the sufficient statistic $((1/n)\sum_{i=1}^n y_i, \min(y_1, \dots, y_n))$ and therefore should have the property of the equivariance, $d(b(1/n)\sum_{i=1}^n y_i + c, b\min(y_1,\dots,y_n)+c) = (1/b)d((1/n)\sum_{i=1}^n y_i, \min(y_1,\dots,y_n))$. Thus it will be reasonable to confine ourselves to the class $\{d_w = w/(\sum_{i=1}^n \log x_i - n\log k'); w>0\}$.

According to Baxter[4], d_n and k' are mutually independent and $T=2an/d_n$ is a χ^2-variable with 2n-2 degrees of freedom. Futhermore k' has a Pareto distribution with the Pareto parameter na and the cut-off parameter k. Hence the density of d_n is

$$\text{(A.3)} \qquad [(na)^{n-1}/(n-2)!](1/d_n{}^n)\exp(-na/d_n)I(d_n>0)$$

and the density of k' is

$$\text{(A.4)} \qquad (nak^{na}/k'^{na+1})I(k'>k).$$

With the quadratic loss function, the risk of the estimator $d_w = w/(\sum_{i=1}^n \log x_i - n\log k') = (w/n)d_n$ is given by

$$\begin{aligned}\text{(A.5)} \qquad R(d_w,(a,k)) &= E(d_w/a - 1)^2 \\ &= E((w/n)d_n/a - 1)^2 \\ &= (1/(n-2)(n-3))[(w - (n - 3))^2 + n - 3],\end{aligned}$$

if $n>3$. So $d_{n-3}=(n-3)/(\sum_{i=1}^{n}\log x_i - n\log k')$ attains the minimum risk uniformly in (a,k) in the class under consideration. We note that the risks of d_{n-3}, d_{n-2} and d_n become larger in this order.

Now we shall show that d_{n-3} is inadmissible. To see it, we should exhibit an estimator for a whose risk is better than or equal to that of d_{n-3} uniformly in (a,k) and is better at some (a_0,k_0) than the risk of d_{n-3}. We shall show that the following estimator has such a property:

$$(A.6) \quad d(d_n,k') = \begin{cases} \max[((n-3)/n)d_n,((n-2)/(n+nd_n\log k'))d_n], & \\ & \text{if } \log k'>0, \\ [(n-3)/n]d_n, & \text{otherwise.} \end{cases}$$

It should be remarked that d will not be shown to be admissible. Since d_n and k' are mutually independent, the joint density of (d_n,k') is

$$(A.7) \quad (na)^n k^{na}\exp(-na/d_n)I(d_n>0)I(k'>k)/[(n-2)!\,d_n{}^n k'^{na+1}].$$

Transforming to the random variable $(d_n,r)=(d_n,nd_n\log k')$, we obtain the joint density of (d_n,r);

$$(A.8) \quad f(d_n,r) = (na)^{na-1}ak^{na}\exp(-(n+r)a/d_n)I(d_n>0)$$
$$\times I(r/n>d_n\log k)/[(n-2)!\,d_n{}^{n+1}].$$

Hence the marginal density of r is

$$(A.9)\qquad f(r) = \int_{-\infty}^{\infty}(na)^{n-1}ak^{na}\exp(-(n+r)a/d_n)I(d_n>0)$$

$$\times I(r/n>d_n\log k)/[(n-2)!d_n{}^{n+1}]dd_n$$

$$= \begin{cases} (na)^{n-1}ak^{na}/(n-2)!\int_0^{r/n\log k}(1/d_n{}^{n+1}) \\ \qquad \times\exp(-(n+r)a/d_n)dd_n, \quad \text{if } \log k>0,\ r\geqq 0, \\ (na)^{n-1}ak^{na}/(n-2)!\int_0^{\infty}(1/d_n{}^{n+1})\exp(-(n+r)a/d_n)dd_n, \\ \qquad \text{if } \log k=0,\ r\geqq 0 \text{ or } \log k<0,\ r\geqq 0, \\ (na)^{n-1}ak^{na}/(n-2)!\int_{r/n\log k}^{\infty}(1/d_n{}^{n+1}) \\ \qquad \times\exp(-(n+r)a/d_n)dd_n, \quad \text{if } \log k<0,\ r<0, \\ 0, \qquad \text{otherwise.} \end{cases}$$

(transforming to $z=(n+r)a/d_n$)

$$= \begin{cases} n^{n-1}k^{na}/[(n-2)!(n+r)^n]\int_{(n+r)an\log k/r}^{\infty} z^{n-1} \\ \qquad \times\exp(-z)dz, \quad \text{if } \log k>0,\ r\geqq 0, \\ n^{n-1}k^{na}(n-1)/(n+r)^n, \text{ if } \log k=0,\ r\geqq 0 \text{ or } \log k<0,\ r\geqq 0, \\ n^{n-1}k^{na}/[(n-2)!(n+r)^n]\int_0^{(n+r)an\log k/r} z^{n-1} \\ \qquad \times\exp(-z)dz, \quad \text{if } \log k<0,\ r<0, \\ 0, \qquad \text{otherwise.} \end{cases}$$

Thus we obtain the following conditional density of d_n given r,

$$(A.10)\qquad f(d_n|r) = f(d_n,r)/f(r)$$

$$
= \begin{cases} A(1/d_n^{n+1})\exp(-(n+r)a/d_n)I(d_n>0) \\ \qquad \times I(r/n\log k>d_n), \quad \text{if } \log k>0,\ r\geqq 0, \\ B(1/d_n^{n+1})\exp(-(n+r)a/d_n)I(d_n>0) \\ \qquad \times I(r/n\log k<d_n), \quad \text{if } \log k<0,\ r\geqq 0, \\ C(1/d_n^{n+1})\exp(-(n+r)a/d_n)I(d_n>0) \\ \qquad \times I(r/n\log k<d_n), \quad \text{if } \log k<0,\ r<0, \\ D(1/d_n^{n+1})\exp(-(n+r)a/d_n)I(d_n>0), \\ \qquad\qquad \text{if } \log k=0,\ r\geqq 0, \\ 0, \qquad\qquad \text{otherwise,} \end{cases}
$$

where

$$A = a^n(n+r)^n/\int_{(n+r)an\log k/r}^{\infty} z^{n-1}\exp(-z)dz,$$

$$B = a^n(n+r)^n/(n-1)!,$$

$$C = a^n(n+r)^n/\int_0^{(n+r)an\log k/r} z^{n-1}\exp(-z)dz,$$

and

$$D = a^n(n+r)^n/(n-1)!.$$

By (A.10), if we denote by $u(r)$ the value of u that minimizes the conditional expectation

$$\text{(A.11)} \qquad E[(ud_n/a - 1)^2|r],$$

then

(A.12) $u(r) = aE(d_n|r)/E(d_n{}^2|r)$

$$= \begin{cases} \int_{(n+r)an\log k/r}^{\infty} z^{n-2}\exp(-z)dz/[(n+r)\int_{(n+r)an\log k/r}^{\infty} z^{n-3}\exp(-z)dz], & \text{if } \log k>0,\ r\geqq 0, \\ (n-2)/(n+r), & \text{if } \log k=0,\ r\geqq 0 \text{ or } \log k<0,\ r\geqq 0, \\ \int_{0}^{(n+r)an\log k/r} z^{n-2}\exp(-z)dz/[(n+r)\int_{0}^{(n+r)an\log k/r} z^{n-3}\exp(-z)dz], & \text{if } \log k<0,\ r<0, \\ 0, & \text{otherwise} \end{cases}$$

$$\begin{cases} \geqq (n-2)/(n+r), & \text{if } \log k>0,\ r\geqq 0, \\ = (n-2)/(n+r), & \text{if } \log k=0,\ r\geqq 0 \text{ or } \log k<0,\ r\geqq 0, \\ \leqq (n-2)/(n+r), & \text{if } \log k<0,\ r<0, \\ = 0, & \text{otherwise.} \end{cases}$$

For r such that $r\geqq 0$ and $(n-3)/n<(n-2)/(n+r)$, it holds that $(n-3)/n<(n-2)/(n+r)\leqq u(r)$, so that we obtain

(A.13) $E[([(n-2)/(n+r)]d_n/a - 1)^2|r] < E[([(n-3)/n]d_n/a - 1)^2|r]$,

noting that (A.11) is a quadratic function of u. Examining (A.9), clearly

(A.14) $\Pr\{r \geqq 0,\ (n-3)/n < (n-2)/(n+r)\} > 0.$

Taking expectations of the both sides of (A.13) with respect to r and noting the form of (A.6), we have

$$\text{(A.15)} \quad E[d(d_n,k')/a - 1]^2 < E[((n-3)/n)d_n/a - 1]^2.$$

Thus it has been proved that d_{n-3} is inadmissible.

Next, adopting $I=100[(R(d_{n-3},(a,k))-R(d,(a,k)))/R(d_{n-3},(a,k))]$ (%) as a measure of the improvement of d over d_{n-3}, we obtain the following table by numerical computation. We note that the value of I depends on (a,k) only through k^a.

Table 4

Values of I, the improvement of d over d_{n-3} (%)

n	k^a							
	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6
4	0.03	0.43	2.16	6.83	16.67	19.36	11.53	7.02
5	0 +	0.10	0.79	3.34	10.19	9.98	3.41	1.18
6	0 +	0.03	0.34	1.93	7.34	5.21	0.83	0.13
7	0 +	0.01	0.16	1.20	5.74	2.53	0.15	0.01
8	0 +	0 +	0.08	0.79	4.72	1.10	0.02	0 +
9	0 +	0 +	0.04	0.54	4.00	0.41	0 +	0 +
10	0 +	0 +	0.02	0.37	3.48	0.13	0 +	0 +

Table 4 shows that the advantage of using d instead of d_{n-3} is almost negligible for large n. Therefore using the simple best equivariant estimator d_{n-3} in such case may be appropriate although it is inadmissible. But the improvement is considerable for small n and for (a,k) with k^a near 1, so that if we had some possibly vague prior information that k^a is near 1, we would hesitate to use d_{n-3}. Arnold[1] showed that the improvement of his estimator over the best location and scale equivariant estimator is at most 3%, far less effective than in the present case.

REFERENCES

[1] Arnold,B.C.(1970). Inadmissibility of the usual scale estimate for a shifted exponential distribution. *J.Amer.Statist.Assoc.*,65,1260-1264.

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

[3] Bancroft,T.A. and Han,C.(1977). Inference based on conditional specification: a note and a bibliography. *International Statistical Review*, 45,117-127.

[4] Baxter,M.A.(1980). Minimum variance unbiased estimation of the parameters of the Pareto distribution. *Metrika*,27,133-138.

[5] Blyth,C.R.(1951). On minimax statistical decision procedures and their admissibility. *Ann.Math.Statist.*,22,22-42.

[6] Bozivich,H., Bancroft,T.A. and Hartley,H.O.(1956). Power of analysis of variance test procedures for certain incompletely specified models I. *Ann.Math.Statist.*,27,1017-1043.

[7] Cohen,A.(1965). A hybrid problem on the exponential family. *Ann.Math. Statist.*,36,1185-1206.

[8] Ghosh,J.K. and Singh,R.(1970). Estimation of the reciprocal of scale parameter of a gamma density. *Ann.Inst.Statist.Math.*,22,51-55.

[9] Gupta,V.P. and Srivastava,S.R.(1968). Inference for a linear hypothesis model using two preliminary tests of significance. *Trabajos de Edstadistica*,19,75-105.

[10] Gupta,V.P. and Srivastava,S.R.(1969). Bias and mean square of an estimation procedure after two preliminary tests of significance in ANOVA model I. *Sankhyā A*,31,319-332.

[11] Gurland,J. and McCullough,R.S.(1962). Testing equality of means after a preliminary test of equality of variances. *Biometrika*,49,403-417.

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

[13] Hirano,K.(1978). A note on the level of significance of the preliminary test in pooling variances. *J.Japan Statist.Soc.*,8,71-75.

[14] Inaba,T. and Nagata,Y.(1985). Minimaxity and nonminimaxity of preliminary test estimators for the multivariate normal mean. *Submitted.*

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

[16] James,W. and Stein,C.(1961). Estimation with quadratic loss. *Proc.4th.*

Berkeley Symp.Math.Statist.Prob.,1,361-379.

[17] Karlin,S.(1958). Admissibility for estimation with quadratic loss. *Ann.Math.Statist.*,29,406-436.

[18] Kitagawa,T.(1963). Estimation after preliminary tests of significance. *Univ.California Pub.Statist.*,3,147-186.

[19] Kitagawa,T.(1984). Successive processes of statistical inference.(in Japanese) *Lecture Series on Information Science,Kyoritsu Shuppan.*

[20] Lehmann,E.L.(1951). A general concept of unbiasedness. *Ann.Math.Statist.*, 22,587-592.

[21] Lehmann,E.L.(1983). Theory of point estimation. *New York,Wiley.*

[22] Mead,R., Bancroft,T.A. and Han,C.(1975). Power of analysis of variance test procedures for incompletely specified fixed models. *Ann.Statist.*, 3,797-808.

[23] Meeden,G. and Arnold,B.C.(1979). The admissibility of a preliminary test estimator when the loss incorporates a complexity cost. *J.Amer. Statist.Assoc.*,74,872-874.

[24] Nagata,Y.(1983). Estimation of the Pareto parameter and its admissibility. *Mathematica Japonica*,28,149-155.

[25] Nagata,Y.(1983). An admissible estimator in the one-parameter exponential family with ambiguous information. *Ann.Inst.Statist.Math.*,35, 193-199.

[26] Nagata,Y.(1983). Admissibility of some preliminary test estimators for the mean of normal distribution. *Ann.Inst.Statist.Math.*,35,365-373.

[27] Nagata,Y. and Inaba,T.(1984). Minimaxity of a preliminary test estimator for the mean of normal distribution. *To appear in Ann.Inst. Statist.Math.*

[28] Nagata,Y.(1984). Minimaxity and inadmissibility of the model selection and estimation procedure for the mean of multivariate normal distribution. *To appear in J.Japan Statist.Soc.*

[29] Nagata,Y. and Araki,T.(1984). Interval estimation of a cell mean in a fixed ANOVA model under a preliminary test. *To appear in Reports of Statistical Application Research, JUSE.*

[30] Nagata,Y. and Araki,T.(1984). Practical pooling method in a fixed ANOVA model and size and power trade-off. *Submitted.*

[31] Nagata,Y.(1985). Data-compatible model selection and a minimax family of model selection and estimation procedures for multivariate normal mean. *Submitted.*

[32] Ohtani,K. and Toyoda,T.(1978). Minimax regret critical values for a preliminary test in pooling variances. *J.Japan Statist.Soc.*,8,15-20.
[33] Patnaik,P.B.(1949). The non-central χ^2- and F-distributions and their applications. *Biometrika*,36,202-232.
[34] Paull,A.E.(1950). On a preliminary test for pooling mean squares in the analysis of variance. *Ann.Math.Statist.*,21,539-556.
[35] Ralescu,D. and Ralescu,S.(1981). A class of nonlinear admissible estimators in the one-parameter exponential family. *Ann.Statist.*,9,177-183.
[36] Sawa,T. and Hiromatsu,T.(1973). Minimax regret significance points for a preliminary test in regression analysis. *Econometrica*,41,1093-1101.
[37] Sawa,T. and Takeuchi,K.(1977). Unbiased decision rules for the choice of regression models. *Working Paper No.400, College of Commerce and Business Administration, Univ. Illinois at Urbana-Champaign.*
[38] Sawa,T.(1978). Information criteria for discriminating among alternative regression models. *Econometrica*,46,1273-1291.
[39] Saxena,K.P. and Srivastava,S.R.(1970). Inference for a linear hypothesis model using two preliminary tests of significance. *Bull.Math. Statist.*,14,83-102.
[40] Sclove,S.L. and Morris,C. and Radhakrishnan,R.(1972). Non-optimality of preliminary test estimators for the mean of a multivariate normal distribution. *Ann.Math.Statist.*,43,1481-1490.
[41] Sen,P.K.(1979). Asymptotic properties of maximum likelihood estimators based on conditional specification. *Ann.Statist.*,7,1019-1033.
[42] Shibata,R.(1984). A survey of the theory of the variable selection. (in Japanese) *Suugaku*,36,344-352, *Iwanami Shoten.*
[43] Singh,J.(1971). Pooling mean squares. *J.Amer.Statist.Assoc.*,66,82-85.
[44] Srivastava,S.R. and Bozivich,H.(1962). Power of certain analysis of variance test procedures involving preliminary tests. *Bull.International Statistical Institute, Proc.33rd Session*,39,133-143.
[45] Srivastava,S.R.(1964). Power of analysis of variance test procedure on the basis of preliminary tests of significance. *J.Indian.Statist.Assoc.*, 2,137-153.
[46] Srivastava,S.R. and Gupta,V.P.(1965). Estimation after preliminary testing in ANOVA model I. *Biometrics*,21,752-758.
[47] Srivastava,S.R.(1972). Pooling mean squares in ANOVA model II. *J.Amer. Statist.Assoc.*,67,676-679.

[48] Stein,C.(1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann.Inst.Statist. Math.*,16,155-160.

[49] Stone,C.(1982). Admissibility and local asymptotic admissibility of procedures which combine estimation and model selection. *Statistical decision theory and related topics III, vol.2,ed.Gupta,S.S. and Berger, J.O.*,317-333.

[50] Toyoda,T. and Wallace,T.D.(1975). Estimation of variance after a preliminary test of homogeneity and optimal levels of significance for the pre-test. *J.Econometrics*,3,395-404.

[51] Toyoda,T. and Wallace,T.D.(1976). Optimal critical values for pretesting in regression. *Econometrica*,44,365-375.