Ann. Inst. Statist. Math.
29 (1977), Part A, 21-34

ESTIMATION PROCEDURES BASED ON PRELIMINARY TEST, SHRINKAGE TECHNIQUE AND INFORMATION CRITERION*

KATUOMI HIRANO

(Received Apr. 8, 1976)

1. Introduction and summary

In many practical problems, experimenters have some idea for the values of the parameters—due to their acquaintance with the properties of the problems. The estimation problems using some a priori information have been investigated. For example, Goodman [3], Searls [11], Khan [8] and Hirano [4], [5], [7] have discussed the estimation procedures for a mean under the assumption that the value of the coefficient of variation is available. Singh, Pandey and Hirano [12] and Hirano [6] have discussed the estimation procedures for the variance under the assumption that the value of the kurtosis is available. Further, Thompson [13], [14] and Mehta and Srinivasan [10] have proposed the shrinkage estimators for the mean μ under the following situation; We believe μ_0 is close to the true value of μ, or we fear that μ_0 may be the true value of μ. Various authors have discussed estimation procedures concerning with some preliminary test of significance (cf. References in Kitagawa [9] and Bock, Yancey and Judge [2]). In this paper we call these estimators the preliminary test estimators. Most of these are concerned with the two sample preliminary test. In this paper, we discuss an estimation procedure with a preliminary test estimator under the situation that only one sample is taken.

The preliminary test estimators are always depending on the level of significance. In this paper we show that it can be determined according to some utilization of the shrinkage technique. We see that it is about 0.15 in many cases. Nevertheless, it is shown that an preliminary test estimator derived from Akaike's information criterion [1] has a level of significance about 0.16. These given by two methods may approximately coincide in many cases.

* This research was partially supported by the Sakkokai Foundation.

2. Preliminary test estimator and shrinkage estimator for estimating the mean of the normal distribution

Let $\bar{X}$ be the sample mean based on a sample of size n from the normal distribution with the unknown mean μ and known variance σ^2. Also let α be the level of significance for testing a null hypothesis H_0 : $\mu=\mu_0$ against an alternative hypothesis H_1 : $\mu\neq\mu_0$ where the value of μ_0 is the a priori information. A preliminary test estimator PT_α for μ is μ_0 if the null hypothesis is accepted, and $\bar{X}$ if rejected;

$$(1)\qquad \mathrm{PT}_\alpha=\begin{cases} \mu_0 & \text{if } H_0 \text{ is accepted},\\ \bar{X} & \text{if } H_1 \text{ is accepted}.\end{cases}$$

It is well known that H_0 is accepted if and only if

$$(2)\qquad \mu_0-\frac{z_\alpha}{\sqrt{n}}\sigma<\bar{X}<\mu_0+\frac{z_\alpha}{\sqrt{n}}\sigma$$

where z_α is the upper $100\alpha/2$-percentile point of the standard normal distribution. Since the estimator PT_α always depends on the level of significance α, we are not able to decide the estimator PT_α uniquely. Hence we would like to decide the value of z_α under some conditions that PT_α is optimal.

Now, Thompson [13] has proposed the shrinkage estimator $\hat{c}\bar{X}$ for which the mean squared error (MSE) is less than that for the sample mean $\bar{X}$. We assume without loss of generality that $\mu_0=0$. A relative efficiency is defined by

$$(3)\qquad \mathrm{REF}(\hat{c}\bar{X}:\bar{X})=\frac{\mathrm{MSE}(\bar{X})}{\mathrm{MSE}(\hat{c}\bar{X})},$$

and it is a function of $\delta=\sqrt{n}\,|\mu/\sigma|$. Let δ_0 be a value of δ for which $\mathrm{REF}(\hat{c}\bar{X}:\bar{X})=1$. Then we can easily get $\delta_0=1.45\cdots$ (see [14] and [10]). When $\sqrt{n}\,|\mu/\sigma|<\delta_0$ we have $\mathrm{REF}(\hat{c}\bar{X}:\bar{X})>1$. If $\mathrm{REF}(\hat{c}\bar{X}:\bar{X})>1$, then $\hat{c}\bar{X}$ is more accurate than $\bar{X}$ in the sense of the mean squared error. Therefore it is reasonable to use an estimator

$$(4)\qquad \hat{c}\bar{X}=\begin{cases} \dfrac{\bar{X}^3}{\bar{X}^2+\sigma^2/n} & \text{if } -\dfrac{\delta_0}{\sqrt{n}}\sigma<\mu<\dfrac{\delta_0}{\sqrt{n}}\sigma\\ \bar{X} & \text{otherwise}\end{cases}$$

for μ. We note the following equivalent relation;

$$(5)\qquad -\frac{\delta_0}{\sqrt{n}}\sigma<\mu<\frac{\delta_0}{\sqrt{n}}\sigma \Longleftrightarrow v>\frac{\sqrt{n}}{\delta_0}$$

where the value of $v=|\sigma/\mu|$ is the coefficient of variation. If the value of v is suitably large, the mean μ is small in comparison with σ. If $\mu_0=0$, then when the value of μ_0 is close to the true value of μ, we shall estimate the mean μ by the value near zero. If (5) is satisfied, then it may be reasonable to use rather used the constant estimator μ_0 as an estimator for μ rather than the shrinkage estimator $\hat{c}\bar{X}$. Hence, for an arbitrary μ_0, we get the following relations;

$$(6)\qquad |\mu-\mu_0|<\frac{\delta_0}{\sqrt{n}}\sigma \Longleftrightarrow |\bar{X}-\mu_0|<\frac{\delta_0}{\sqrt{n}}\sigma$$

$$\Longleftrightarrow \mu_0-\frac{\delta_0}{\sqrt{n}}\sigma<\bar{X}<\mu_0+\frac{\delta_0}{\sqrt{n}}\sigma ,$$

if H_0 is true. Consequently we have an estimator

$$(7)\qquad \mathrm{PT}_{\delta_0}=\begin{cases} \mu_0 & \text{if } \mu_0-\frac{\delta_0}{\sqrt{n}}\sigma<\bar{X}<\mu_0+\frac{\delta_0}{\sqrt{n}}\sigma \\ \bar{X} & \text{otherwise} \end{cases}$$

for the mean μ, in this case the level of significance α in the preliminary test is about 0.15. It should be noted that PT_{δ_0} is a consistent estimator.

In Section 4, when σ^2 is known, we shall get the mean squared error of PT_{δ_0}.

Remarks. For simplicity we assume $\sigma=1$. If $-\delta_0/\sqrt{n}<\mu<\delta_0/\sqrt{n}$, then we may use the shrinkage estimator $\hat{c}\bar{X}$ for μ. However in this case the value of μ is approximately equal to zero for suitably large n. It is reasonable to estimate μ by a constant zero instead of the variable $\hat{c}\bar{X}$. Hence we propose the estimator with the decision rule (7) under some optimal conditions. Also the estimator PT_α is a super efficient estimator for some approximate values of μ_0 (see Fig. 1). The estimation problems for the mean have discussed in [4] when we have an approximately known coefficient of variation as the a priori information. From [4], it is reasonable to shrink $\bar{X}$ for the suitably large coefficient of variations.

In general, as we can see from the above discussions and Remarks, it is reasonable to propose the following estimator; the estimator is a constant (i.e. an approximate value as an a priori information) if the shrinkage estimator is more precise than the original estimator in the sense of the mean squared error, and the original, if otherwise. This estimator is called the one due to the shrinkage technique.

We call an estimator derived from the AIC statistic the estimator due to the information criterion. In Section 4 it will be shown that the estimators due to the shrinkage technique are approximately equal to those due to the information criterion.

3. Preliminary test estimator and AIC

We discuss the problem of estimating an r-tuple of parameters $\theta=(\theta_1,\cdots,\theta_r)$ of the probability distribution with a density function $f(x|\theta)$. Suppose that we have a priori information

$$(8) \qquad \theta_1\approx\theta_{10},\cdots,\theta_p\approx\theta_{p0} \qquad (p\leqq r)$$

where $\theta_{10},\cdots,\theta_{p0}$ are the known constants and $\theta_{p+1},\cdots,\theta_r$ are the unknown parameters. Let $\hat{\theta}$ and $\tilde{\theta}$ be the estimators (for example, the maximum likelihood estimators) for $(\theta_1,\cdots,\theta_r)$ and $(\theta_{10},\cdots,\theta_{p0},\theta_{p+1},\cdots,\theta_r)$, respectively.

Consider a testing statistical hypothesis

$$(9) \qquad \begin{aligned} &H_0\,;\quad \theta_1=\theta_{10},\cdots,\theta_p=\theta_{p0}\\ &H_1\,;\quad \text{at least one equality fails}\,. \end{aligned}$$

For the level of significance α, then we have the following consistent estimators, for $(\theta_1,\cdots,\theta_p)$

$$(10) \qquad \mathrm{PT}_\alpha(\theta_1,\cdots,\theta_p)=\begin{cases} \theta_{10},\cdots,\theta_{p0} & \text{if } H_0 \text{ is accepted}\,,\\ \hat{\theta} & \text{if } H_0 \text{ is rejected}\,, \end{cases}$$

and for $(\theta_1,\cdots,\theta_r)$

$$(11) \qquad \mathrm{PT}_\alpha(\theta_1,\cdots,\theta_r)=\begin{cases} \tilde{\theta} & \text{if } H_0 \text{ is accepted}\,,\\ \hat{\theta} & \text{if } H_0 \text{ is rejected}\,. \end{cases}$$

However since these estimators always depend on α, PT_α are not unique. Hence we would like to decide on the value of α under some conditions that these PT_α are optimal.

We consider these estimation problems using an information criterion given by Akaike [1]. According to this criterion, we choose the estimating model with the AIC, less than AIC's of the others, of the following form: AIC$=-2\log_e$(the maximum likelihood)$+2$(the number of the parameters). At first, $\mathrm{AIC}_{\mathrm{II}}$ of the estimating model for which this a priori information was used, is given by

$$(12) \qquad \mathrm{AIC}_{\mathrm{II}}=-2\log L(\theta_{10},\cdots,\theta_{p0},\tilde{\theta}_{p+1},\cdots,\tilde{\theta}_r)+2(r-p)$$

where $L(\theta_{10},\cdots,\theta_{p0},\tilde{\theta}_{p+1},\cdots,\tilde{\theta}_r)$ is a maximum likelihood estimator for a parameter $(\theta_{10},\cdots,\theta_{p0},\theta_{p+1},\cdots,\theta_r)$. On the other hand, when there is no a priori information, AIC_I of the estimating model in this case is given by

$$\text{AIC}_\text{I} = -2\log L(\hat{\theta}) + 2r \tag{13}$$

where $L(\hat{\theta})$ is a maximum likelihood estimator for $(\theta_1, \cdots, \theta_r)$. If a priori information is effective, then $\text{AIC}_\text{I} - \text{AIC}_\text{II} > 0$ from the information criterion. We have the following equivalent relation

$$\text{AIC}_\text{I} - \text{AIC}_\text{II} > 0 \Longleftrightarrow -2\log \frac{L(\theta_{10}, \cdots, \theta_{p0}, \tilde{\theta}_{p+1}, \cdots, \tilde{\theta}_r)}{L(\hat{\theta})} < 2p\,. \tag{14}$$

Therefore we obtain the following consistent estimators, for the parameter $(\theta_1, \cdots, \theta_p)$

$$\text{PT}\,(\theta_1, \cdots, \theta_p) = \begin{cases} (\theta_{10}, \cdots, \theta_{p0}) & \text{if (14) is satisfied}\,, \\ L(\hat{\theta}) & \text{otherwise}\,, \end{cases} \tag{15}$$

and for the parameter $(\theta_1, \cdots, \theta_r)$

$$\text{PT}\,(\theta_1, \cdots, \theta_r) = \begin{cases} L(\theta_{10}, \cdots, \theta_{p0}, \tilde{\theta}_{p+1}, \cdots, \tilde{\theta}_r) & \text{if (14) is satisfied}\,, \\ L(\hat{\theta}) & \text{otherwise}\,. \end{cases} \tag{16}$$

When the null hypothesis H_0 is true, the left-hand side of the second inequality in (14) converges in distribution to a chi-square distribution with $(r-p)$ degrees of freedom under some regularity conditions (for example, see Wilks [15]). Hence we obtain

$$\lim_{n\to\infty} \Pr\left(-2\log \frac{L(\theta_{10}, \cdots, \theta_{p0}, \tilde{\theta}_{p+1}, \cdots, \tilde{\theta}_r)}{L(\hat{\theta})} < 2p\right) = \Pr\,(\chi^2_{r-p} < 2p) \tag{17}$$

where χ^2_{r-p} is a random variable from the chi-square distribution with $(r-p)$ degrees of freedom. Consequently in the preliminary test the level of significance α is asymptotically $1-\Pr\,(\chi^2_{r-p} < 2p)$.

Here we consider the following three special cases.

(i) *Case* 1. (1-component parameter $\theta = \theta_1$; $r = p = 1$)

For the testing statistical hypothesis

$$H_0\,;\ \ \theta = \theta_{10} \qquad H_1\,;\ \ \theta \neq \theta_{10}\,, \tag{18}$$

we have from (15) an estimator for θ_1,

$$\text{PT}\,(\theta_1) = \begin{cases} \theta_{10} & \text{if } \left(\prod_{i=1}^n f(x_i \mid L(\hat{\theta})) \Big/ \prod_{i=1}^n f(x_i \mid \theta_{10})\right) < e \\ L(\hat{\theta}) & \text{otherwise}\,. \end{cases} \tag{19}$$

Therefore in this case, when H_0 is true,

$$(20)\qquad \lim_{n\to\infty} \Pr(\mathrm{PT}(\theta_1)=\theta_{10})$$
$$=\lim_{n\to\infty} \Pr\left(-2\log\left(\prod_{i=1}^{n} f(x_i\,|\,\theta_{10})\Big/\prod_{i=1}^{n} f(x_i\,|\,L(\hat{\theta}))\right)<2\right)$$
$$=0.8427\cdots.$$

The level of significance α in this test is asymptotically about 16%.

(ii) *Case* 2-1. (2-component parameter $\theta=(\theta_1, \theta_2)$; $r=2$, $p=1$)

For the testing statistical hypothesis

$$(21)\qquad H_0;\ \ \theta_1=\theta_{10} \qquad H_1;\ \ \theta_1\neq\theta_{10}$$

where θ_2 is the unknown parameter, we have an estimator for θ_1

$$(22)\qquad \mathrm{PT}(\theta_1)=\begin{cases} \theta_{10} & \text{if } \left(\prod_{i=1}^{n} f(x_i\,|\,L(\hat{\theta}))\Big/\prod_{i=1}^{n} f(x_i\,|\,L(\theta_{10}, \tilde{\theta}_2))\right)<e\,, \\ L(\hat{\theta}) & \text{otherwise}\,, \end{cases}$$

where $L(\theta_{10}, \tilde{\theta}_2)$ is the maximum likelihood estimator of (θ_{10}, θ_2). When H_0 is true,

$$(23)\qquad \lim_{n\to\infty} \Pr(\mathrm{PT}(\theta_1)=\theta_{10})$$
$$=\lim_{n\to\infty} \Pr\left(-2\log\left(\prod_{i=1}^{n} f(x_i\,|\,L(\theta_{10}, \tilde{\theta}_2))\Big/\prod_{i=1}^{n} f(x_i\,|\,L(\hat{\theta}))\right)<2\right)$$
$$=0.8427\cdots.$$

Also we obtain that the level of significance α is asymptotically about 16%.

(iii) *Case* 2-2. (2-component parameter $\theta=(\theta_1, \theta_2)$; $r=p=2$)

For the testing statistical hypothesis

$$(24)\qquad \begin{array}{l} H_0;\ \ \theta_1=\theta_{10},\ \ \theta_2=\theta_{20} \\ H_1;\ \ \text{at least one equality fails}\,, \end{array}$$

we have an estimator for (θ_1, θ_2)

$$(25)\qquad \mathrm{PT}(\theta_1, \theta_2)=\begin{cases} (\theta_{10}, \theta_{20}) & \text{if } \left(\prod_{i=1}^{n} f(x_i\,|\,L(\hat{\theta}))\Big/\prod_{i=1}^{n} f(x_i\,|\,(\theta_{10}, \theta_{20}))\right)<e^2 \\ L(\hat{\theta}) & \text{otherwise}\,. \end{cases}$$

When H_0 is true,

$$(26)\qquad \lim_{n\to\infty} \Pr(\mathrm{PT}(\theta_1, \theta_2)=(\theta_{10}, \theta_{20}))$$
$$=\lim_{n\to\infty} \Pr\left(-2\log\frac{L(\theta_{10}, \theta_{20})}{L(\hat{\theta})}<4\right)=\Pr(\chi_2^2<4)=0.8646\cdots.$$

We note that the level of significance α is asymptotically about 14%.

4. Estimator PT of the mean μ in the normal distribution with the known variance σ^2 and its properties

In estimation problems for the mean μ in the normal distribution with the known variance σ^2 when we have some approximated value μ_0 of the mean μ, from (12), we have

$$\text{AIC}_{\text{II}}=(-2)\log\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{(x_i-\mu_0)^2}{2\sigma^2}\right)+2\cdot 0\,. \tag{27}$$

In the usual model,

$$\text{AIC}_{\text{I}}=(-2)\log\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{(x_i-\bar{x})^2}{2\sigma^2}\right)+2\cdot 1\,. \tag{28}$$

When $\text{AIC}_{\text{I}}-\text{AIC}_{\text{II}}>0$, we choose the estimating model II with AIC_{II}. Since it is equivalent to

$$\mu_0-\sqrt{\frac{2}{n}}\sigma<\bar{X}<\mu_0+\sqrt{\frac{2}{n}}\sigma\,, \tag{29}$$

we obtain the estimator

$$\text{PT}(\mu)=\begin{cases}\mu_0 & \text{if } \mu_0-\sqrt{\frac{2}{n}}\sigma<\bar{X}<\mu_0+\sqrt{\frac{2}{n}}\sigma\\ \bar{X} & \text{otherwise}\,.\end{cases} \tag{30}$$

We note that in this case the level of significance α in the preliminary test is about 0.16.

We give the mean squared error of $\text{PT}(\mu)$. Since

$$\begin{aligned}\Pr(\text{PT}(\mu)=\mu_0)&=\Pr\left(\mu_0-\frac{z}{\sqrt{n}}\sigma<\bar{X}<\mu_0+\frac{z}{\sqrt{n}}\sigma\right)\\&=\Phi\left(\frac{\sqrt{n}}{\sigma}(\mu_0-\mu)+z\right)-\Phi\left(\frac{\sqrt{n}}{\sigma}(\mu_0-\mu)-z\right)\\&=q\,,\qquad(\text{say})\,,\end{aligned} \tag{31}$$

where $\Phi(x)$ is the standard normal distribution function, we get

$$\begin{aligned}\text{E}(\text{PT}(\mu)-\mu)^2=&\int_{-\infty}^{\mu_0-(z/\sqrt{n})\sigma}(x-\mu)^2\frac{\sqrt{n}}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{n(x-\mu)^2}{2\sigma^2}\right)dx\\&+(\mu_0-\mu)^2\Pr(\text{PT}(\mu)=\mu_0)\\&+\int_{\mu_0+(z/\sqrt{n})\sigma}^{\infty}(x-\mu)^2\frac{\sqrt{n}}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{n(x-\mu)^2}{2\sigma^2}\right)dx\,.\end{aligned} \tag{32}$$

Putting $(\mu_0-\mu)/\sigma=t$, we easily have

(33) $$\mathrm{E}(\mathrm{PT}(\mu)-\mu)^2/\sigma^2$$
$$=\begin{cases} qt^2+\dfrac{1}{2n}\left(2+G\left(\dfrac{(\sqrt{n}\,t-z)^2}{2}\right)-G\left(\dfrac{(\sqrt{n}\,t+z)^2}{2}\right)\right); & t\geqq z/\sqrt{n} \\ qt^2+\dfrac{1}{2n}\left(2-G\left(\dfrac{(\sqrt{n}\,t-z)^2}{2}\right)-G\left(\dfrac{(\sqrt{n}\,t+z)^2}{2}\right)\right); & z/\sqrt{n}>t>-z/\sqrt{n} \\ qt^2+\dfrac{1}{2n}\left(2-G\left(\dfrac{(\sqrt{n}\,t-z)^2}{2}\right)+G\left(\dfrac{(\sqrt{n}\,t+z)^2}{2}\right)\right); & -z/\sqrt{n}\geqq t \end{cases}$$

where $G(x)$ is a Gamma distribution function given by

(34) $$G(x)=\int_0^x \frac{1}{\Gamma(3/2)}y^{1/2}e^{-y}dy\ .$$

When $t=0$, for any n, we obtain

(35) $$\frac{\mathrm{E}(\mathrm{PT}(\mu)-\mu)^2}{\sigma^2}=\frac{1}{n}-\frac{1}{n}G\left(\frac{z^2}{2}\right)<\frac{\mathrm{Var}(\bar{X})}{\sigma^2}\ .$$

For $z=\sqrt{2}$ we get the relative efficiency REF (PT (μ): $\bar{X}$) defined by (3). When $t=0$ (i.e., $\mu=\mu_0$), REF (PT (μ): $\bar{X}$) has a maximum value 1/(1−

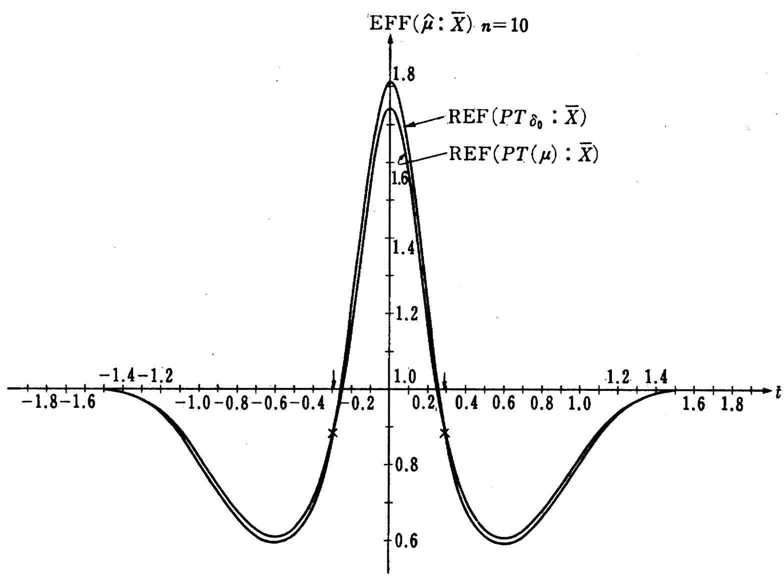

Fig. 1

$G(1))=1.747\cdots$ for all n.

Next, when $z=\delta_0$ in (33), we have $E(PT_{\delta_0}-\mu)^2/\sigma^2$. When $t=0$ (i.e., $\mu=\mu_0$), REF (PT_{δ_0}: $\bar{X}$) has a maximum value $1/(1-G(\delta_0^2/2))=1.813\cdots$ for all n. In Fig. 1, we exhibit REF (PT (μ): $\bar{X}$) and REF (PT_{δ_0}: $\bar{X}$) curves for each t, for $n=10$.

It should be noted that the estimators PT_{δ_0} and PT (μ) have approximately same definitions and properties, and that PT (μ) is a consistent estimator.

5. Estimator PT (μ) of the mean μ in the normal distribution with the unknown variance σ^2

In estimating problems of the mean μ in the normal distribution with the unknown variance σ^2, we consider, from the *Case* 2-1 in Section 3, the following test

$$H_0\,;\quad \mu=\mu_0 \qquad H_1\,;\quad \mu\neq\mu_0\,. \tag{36}$$

Then we easily have an estimator for the mean

$$\mathrm{PT}(\mu)=\begin{cases} \mu_0 & \text{if } \left(\sum_{i=1}^{n}(x_i-\mu_0)^2\Big/\sum_{i=1}^{n}(x_i-\bar{x})^2\right)<1+2/n\,, \\ \bar{X} & \text{otherwise}\,. \end{cases} \tag{37}$$

Since

$$\begin{aligned} &\left(\sum_{i=1}^{n}(x_i-\mu_0)^2\Big/\sum_{i=1}^{n}(x_i-\bar{x})^2\right)<1+2/n \\ &\iff \frac{|\bar{x}-\mu_0|}{\sigma/\sqrt{n}}\Big/\sqrt{\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)\Big/(n-1)}<\sqrt{\frac{2}{n}}\cdot\sqrt{n-1}\,, \end{aligned} \tag{38}$$

and also since the left-hand side of the second inequality in (38) has the t-distribution with $(n-1)$ degrees of freedom, we get

$$\Pr(\mathrm{PT}(\mu)=\mu_0)=\Pr(|t_{n-1}|<\sqrt{2-2/n}) \tag{39}$$

where t_{n-1} is distributed according to the t-distribution with $(n-1)$ degrees of freedom. Hence the level of significance in the preliminary test is exactly $1-\Pr(|t_{n-1}|<\sqrt{2-2/n})$. For sufficiently large n, this is asymptotically equal to the level of significance given by the previous section. The estimator PT (μ) is consistent.

6. Estimation procedure for the binomial distribution

Let us consider the case where X is a random variable having the binomial distribution with the unknown parameter p and the known parameter n. We regard the values of p as a priori information. When we have an approximate value p_0, the estimation procedures for p are discussed.

Let α be the level of significance for the test under a null hypothesis H_0; $p=p_0$ against an alternative hypothesis H_1; $p\neq p_0$. For p_1 and p_2 $(p_1\leqq p_2)$ so that $\Pr(X<p_1)=\alpha/2$ and $\Pr(X>p_2)=\alpha/2$, we define a preliminary test estimator as follows;

$$\text{(40)} \qquad \mathrm{PT}_\alpha=\begin{cases} p_0 & \text{if } np_1\leqq X\leqq np_2 \\ X/n & \text{otherwise}. \end{cases}$$

Since PT_α always depends on α, we are going to decide α by the shrinkage technique.

For p, the shrinkage estimator $\hat{p}$ is given by

$$\text{(41)} \qquad \hat{p}=\frac{(X-np_0)^3}{n(X-np_0)^2+X(n-X)}+p_0$$

where p_0 is the value towards which X/n is to be shrunken. The mean squared error of $\hat{p}$, $\mathrm{MSE}(\hat{p})$, is $\sum_{x=0}^{n}(\hat{p}-p)^2\,{}_nC_x\,p^x(1-p)^{n-x}$. Since

$$\text{(42)} \qquad \mathrm{REF}\left(\hat{p}:\frac{X}{n}\right)=\frac{\mathrm{MSE}(\hat{p})}{\mathrm{MSE}(X/n)}$$
$$=\frac{n}{p(1-p)}\sum_{x=0}^{n}\left[\frac{(x-np_0)^3}{n(x-np_0)^2+x(n-x)}+p_0-p\right]^2 \cdot {}_nC_x\,p^x(1-p)^{n-x},$$

we are able to get the values of p so that $\mathrm{REF}(\hat{p}:X/n)=1$ for each n and p_0. Denoting these values by p_1 and p_2 $(p_1\leqq p_2)$ (suppose $p_1=0$ when the solution has only one p_2), for $p_1\leqq p\leqq p_2$ the shrinkage estimator $\hat{p}$ is more precise than X/n in the sense of the mean squared error. Therefore we obtain the preliminary test estimator with the following decision rule;

$$\text{(43)} \qquad \mathrm{PT}=\begin{cases} p_0 & \text{if } np_1\leqq X\leqq np_2 \\ X/n & \text{otherwise}. \end{cases}$$

And we are able to decide the level of significance α so that the preliminary test estimator PT_α has the minimum mean squared error. Hence we obtain $\alpha=1-\Pr(\mathrm{PT}=p_0)$.

In Table 1, we show the values of p_1, p_2 and $\Pr(\mathrm{PT}=p_0)$ for each n and p_0.

It is reasonable to use the one-sided test in the preliminary test when p_0 is near 0 or 1. This is illustrated by the case that the equation $\mathrm{REF}(\hat{p}:X/n)=1$ has only one solution with respect to p. Hence we may use the one-sided preliminary test for some n and p_0. Though we cannot easily get some simple conditions between n and p_0 to use the one-sided preliminary test, these conditions may be approximately regarded as (47).

Next, for the binomial distribution we discuss this estimation prob-

Table 1

n	p_0	Shrinkage method			Information method		
		p_1	p_2	$\Pr(np_1 \leqq X \leqq np_2)$	p_1	p_2	$\Pr(np_1 \leqq X \leqq np_2)$
5	0.1	0.0000	0.3329	0.9185	0.0000	0.3325	0.9185
	0.2	0.0000	0.4637	0.6144	0.0046	0.4829	0.6144
	0.3	0.1338	0.5653	0.6689	0.0503	0.6063	0.8012
	0.4	0.1905	0.6560	0.8352	0.1154	0.7126	0.8352
	0.5	0.2674	0.7326	0.7813	0.1948	0.8052	0.9375
6	0.1	0.0000	0.3213	0.8857	0.0000	0.3095	0.8857
	0.2	0.0609	0.4562	0.6390	0.0148	0.4565	0.6390
	0.3	0.1346	0.5539	0.8119	0.0675	0.5790	0.8119
	0.4	0.2014	0.6430	0.7741	0.1372	0.6860	0.9124
	0.5	0.2744	0.7256	0.8750	0.2197	0.7803	0.8750
7	0.1	0.0000	0.3137	0.9743	0.0000	0.2919	0.9743
	0.2	0.0796	0.4431	0.7569	0.0241	0.4362	0.7569
	0.3	0.1377	0.5429	0.7916	0.0818	0.5578	0.7916
	0.4	0.2084	0.6328	0.8758	0.1546	0.6651	0.8758
	0.5	0.2847	0.7153	0.9297	0.2393	0.7607	0.9297
8	0.1	0.0000	0.3076	0.9619	0.0000	0.2779	0.9619
	0.2	0.0845	0.4307	0.7760	0.0324	0.4199	0.7760
	0.3	0.1443	0.5313	0.8844	0.0937	0.5407	0.8844
	0.4	0.2147	0.6230	0.8095	0.1690	0.6481	0.9334
	0.5	0.2925	0.7075	0.8203	0.2554	0.7446	0.8203
9	0.1	0.0000	0.3016	0.9470	0.0000	0.2664	0.9470
	0.2	0.0855	0.4203	0.7801	0.0397	0.4065	0.7801
	0.3	0.1494	0.5214	0.8608	0.1039	0.5266	0.8608
	0.4	0.2215	0.6139	0.8906	0.1812	0.6340	0.8906
	0.5	0.3005	0.6995	0.8906	0.2688	0.7312	0.8906
10	0.1	0.0000	0.2951	0.7361	0.0009	0.2569	0.5811
	0.2	0.0868	0.4111	0.8598	0.0462	0.3952	0.7718
	0.3	0.1531	0.5126	0.9244	0.1127	0.5146	0.9244
	0.4	0.2269	0.6061	0.8989	0.1916	0.6221	0.9392
	0.5	0.3072	0.6928	0.7734	0.2802	0.7198	0.9346
11	0.1	0.0167	0.2882	0.6677	0.0032	0.2487	0.5966
	0.2	0.0893	0.4025	0.8637	0.0520	0.3855	0.8637
	0.3	0.1571	0.5045	0.9020	0.1205	0.5044	0.9020
	0.4	0.2322	0.5987	0.8704	0.2006	0.6117	0.8704
	0.5	0.3137	0.6863	0.8540	0.2901	0.7099	0.8540
12	0.1	0.0269	0.2810	0.6919	0.0055	0.2416	0.6067
	0.2	0.0923	0.3946	0.8587	0.0571	0.3772	0.8587
	0.3	0.1610	0.4971	0.8683	0.1273	0.4954	0.8683
	0.4	0.2370	0.5921	0.9231	0.2086	0.6027	0.9231
	0.5	0.3194	0.6806	0.9077	0.2988	0.7012	0.9077
13	0.1	0.0331	0.2739	0.7117	0.0078	0.2355	0.7117
	0.2	0.0950	0.3874	0.9150	0.0618	0.3698	0.8459
	0.3	0.1643	0.4905	0.8740	0.1334	0.4875	0.9279
	0.4	0.2413	0.5861	0.8444	0.2156	0.5948	0.8897
	0.5	0.3247	0.6753	0.8204	0.3064	0.6936	0.9426

Table 1 (Continued)

n	p_0	Shrinkage method			Information method		
		p_1	p_2	$\Pr(np_1 \leqq X \leqq np_2)$	p_1	p_2	$\Pr(np_1 \leqq X \leqq np_2)$
	0.1	0.0368	0.2671	0.7271	0.0100	0.2300	0.7271
	0.2	0.0971	0.3810	0.9122	0.0660	0.3632	0.9122
14	0.3	0.1674	0.4844	0.8592	0.1389	0.4805	0.8999
	0.4	0.2454	0.5805	0.9019	0.2220	0.5877	0.9019
	0.5	0.3295	0.6705	0.8815	0.3133	0.6867	0.8815
	0.1	0.0388	0.2609	0.7386	0.0120	0.2251	0.7386
	0.2	0.0989	0.3751	0.9038	0.0699	0.3574	0.9038
15	0.3	0.1704	0.4788	0.9147	0.1439	0.4742	0.9147
	0.4	0.2492	0.5754	0.8778	0.2277	0.5813	0.8778
	0.5	0.3340	0.6660	0.7899	0.3195	0.6805	0.9232
	0.1	0.0400	0.2552	0.7977	0.0140	0.2207	0.7463
	0.2	0.1007	0.3696	0.8902	0.0734	0.3521	0.8902
16	0.3	0.1732	0.4737	0.8995	0.1484	0.4685	0.8995
	0.4	0.2527	0.5707	0.8765	0.2329	0.5755	0.9233
	0.5	0.3382	0.6618	0.8565	0.3251	0.6749	0.8565
	0.1	0.0406	0.2500	0.8111	0.0158	0.2167	0.7506
	0.2	0.1025	0.3646	0.9398	0.0767	0.3473	0.8718
17	0.3	0.1757	0.4690	0.8761	0.1525	0.4634	0.8761
	0.4	0.2559	0.5664	0.8617	0.2377	0.5703	0.8617
	0.5	0.3420	0.6580	0.9038	0.3302	0.6698	0.9038
	0.1	0.0409	0.2453	0.8217	0.0176	0.2131	0.7517
	0.2	0.1043	0.3600	0.9307	0.0797	0.3429	0.9307
18	0.3	0.1782	0.4646	0.8805	0.1564	0.4586	0.9262
	0.4	0.2590	0.5623	0.9096	0.2420	0.5655	0.9096
	0.5	0.3456	0.6544	0.8329	0.3349	0.6651	0.8329
	0.1	0.0413	0.2411	0.8297	0.0192	0.2098	0.7499
	0.2	0.1059	0.3557	0.8495	0.0825	0.3389	0.9180
19	0.3	0.1804	0.4605	0.8699	0.1599	0.4543	0.8699
	0.4	0.2619	0.5586	0.8886	0.2461	0.5610	0.8886
	0.5	0.3490	0.6510	0.8847	0.3392	0.6608	0.8847
	0.1	0.0417	0.2371	0.8353	0.0208	0.2067	0.8353
	0.2	0.1073	0.3517	0.8987	0.0852	0.3352	0.9018
20	0.3	0.1826	0.4567	0.9166	0.1632	0.4503	0.9166
	0.4	0.2646	0.5550	0.8925	0.2498	0.5569	0.9275
	0.5	0.3521	0.6479	0.8108	0.3432	0.6568	0.9217

lem using the information criterion. From Section 3 we easily obtain an estimator for p,

$$\hat{p}_I = \begin{cases} p_0 & \text{if } \left(\dfrac{p}{p_0}\right)^p \left(\dfrac{1-p}{1-p_0}\right)^{1-p} \leqq e^{1/n} \\ X/n & \text{otherwise} \end{cases} \tag{44}$$

when we have the approximate value p_0 regarding the values of p. For n and p_0, the equation with respect to p $(0<p<1)$

$$\left(\frac{p}{p_0}\right)^p \left(\frac{1-p}{1-p_0}\right)^{1-p} = e^{1/n} \tag{45}$$

is equivalent to the equation with respect to p

(46) $$p(\log p - \log p_0) + (1-p)[\log(1-p) - \log(1-p_0)] = 1/n$$

and has at most the two solutions p_1 and p_2 ($p_1 \leqq p_2$) (suppose $p_1=0$ when the solution has only one p_2), (see Fig. 2). The conditions that the solution of (46) is only one are

(47) $$\begin{aligned} -\log p_0 &< 1/n \quad &\text{if } 1 > p_0 \geqq 1/2 \\ -\log(1-p_0) &< 1/n \quad &\text{if } 1/2 > p_0 > 0\,, \end{aligned}$$

(see Fig. 2).

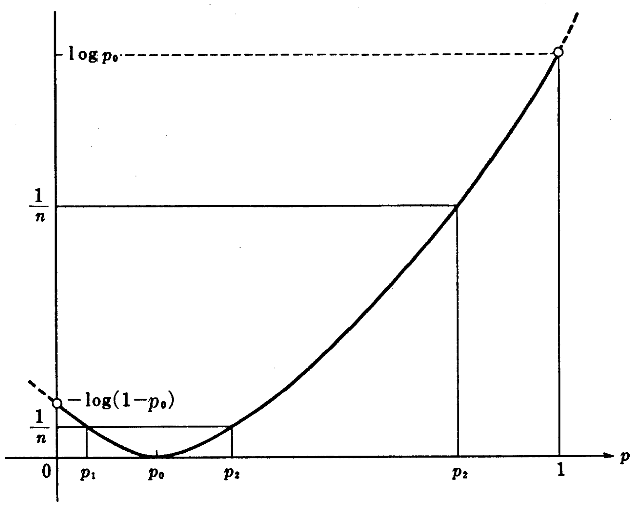

Fig. 2

The values of p_1 and p_2 are given at Table 1 for each n and p_0. Further for each n and p_0 we give the value of $\Pr(\hat{p}_{\mathrm{I}}=p_0)$. From Table 1 it is expected to have

(48) $$\lim_{n\to\infty} \Pr(\mathrm{PT}=p_0) = \lim_{n\to\infty} \Pr(\hat{p}_{\mathrm{I}}=p_0) = 0.8427\cdots$$

by the numerical calculations.

Acknowledgement

The author is indebted to Dr. K. Takahasi for his valuable advices and useful comments.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

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

[2] Bock, M. E., Yancey, T. A. and Judge, G. G. (1973). The statistical consequences of preliminary test estimators in regression, *J. Amer. Statist. Ass.*, **68**, 109-116.

[3] Goodman, L. A. (1953). A simple method for improving some estimators, *Ann. Math. Statist.*, **24**, 114-117.

[4] Hirano, K. (1972). Using some approximately known coefficient of variation in estimating mean, (in Japanese), *Proc. Inst. Statist. Math.*, **20**, 61-64.

[5] Hirano, K. (1973). Biased efficient estimator utilizing some a priori information, *J. Japan Statist. Soc.*, **4**, 11-13.

[6] Hirano, K. (1973). Some properties of an estimator for the variance of a normal distribution, *Ann. Inst. Statist. Math.*, **25**, 479-492.

[7] Hirano, K. (1974). The utilization of a known coefficient of variation matrix in the estimation procedure of mean vector, *Res. Memo.* No. 66, The Inst. Statist. Math.

[8] Khan, R. (1968). A note on estimating the mean of a normal distribution with known coefficient of variation, *J. Amer. Statist. Ass.*, **63**, 1039-1041.

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

[10] Mehta, J. S. and Srinivasan, R. (1971). Estimation of the mean by shrinkage to a point, *J. Amer. Statist. Ass.*, **66**, 86-90.

[11] Searls, D. T. (1964). The utilizing of a known coefficient of variation in estimation procedure, *J. Amer. Statist. Ass.*, **59**, 1225-1226.

[12] Singh, J., Pandey, B. N. and Hirano, K. (1973). On the utilizing of a known coefficient of kurtosis in estimation procedure of variance, *Ann. Inst. Statist. Math.*, **25**, 51-55.

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

[14] Thompson, J. R. (1968). Accuracy borrowing in the estimation of the mean by shrinkage to an interval, *J. Amer. Statist. Ass.*, **63**, 953-963.

[15] Wilks, S. S. (1962). *Mathematical Statistcs*, John Wiley & Sons, Inc., New York.

Ann. Inst. Statist. Math.
29 (1977), Part A, 317

CORRECTIONS TO

"ESTIMATION PROCEDURES BASED ON PRELIMINARY TEST, SHRINKAGE TECHNIQUE AND INFORMATION CRITERION"

KATUOMI HIRANO

In the above titled paper (this Annals 29 (1977), Part A, 21–34), the following corrections should be made:

1. (i) On page 29, in (37) and (38):

"$1+2/n$" should be "$\exp(2/n)$".

(ii) On page 29, in (38):

"$\sqrt{\frac{2}{n}}\cdot\sqrt{n-1}$" should be "$\sqrt{\exp(2/n)-1}\cdot\sqrt{n-1}$".

(iii) On page 29, in (39) and line 6 from the bottom:

"$\sqrt{2-2/n}$" should be "$\sqrt{\exp(2/n)-1}\cdot\sqrt{n-1}$".

If the above corrections are made, in Section 5 we should note that

$\lim_{n\to\infty}\sqrt{\exp(2/n)-1}\sqrt{n-1}=\sqrt{2}$.

2. (i) On page 32, line 5 from the bottom:

$$\hat{p}_I=\begin{cases} p_0 & \text{if } \left(\frac{p}{p_0}\right)^p\left(\frac{1-p}{1-p_0}\right)^{1-p}\leqq e^{1/n} \\ X/n & \text{otherwise}, \end{cases}$$

$$\Longrightarrow \qquad \hat{p}_I=\begin{cases} p_0 & \text{if } \left(\frac{\tilde{p}}{p_0}\right)^{\tilde{p}}\left(\frac{1-\tilde{p}}{1-p_0}\right)^{1-\tilde{p}}\leqq e^{1/n} \\ X/n & \text{otherwise}, \end{cases} \qquad \text{where } \tilde{p}=X/n,$$

(ii) On page 32, lines 3, 2 and 1 from the bottom and on page 33, line 1 and Fig. 2:

"p" should be "$\tilde{p}$".