ON THE WEIGHTED LEAST SQUARES ESTIMATION AND THE EXISTENCE THEOREM OF ITS OPTIMAL SOLUTION

by
Tadashi NAKAMURA and Taichi KARIYA

Department of Mathematics, Kawasaki Medical School
Kurashiki 701—01, *Japan*
(*Received on Oct.* 31, 1975)

Abstract

In practice, samples from continuous distributions are often grouped. This means that we are not given the individual observations, but only the number of observations falling into certain specified intervals. In such a case the estimation of parameters is usually performed by the method of maximum likelihood based on grouped data. However, there are other estimation procedures, say, the chi-squares minimum method, the modified chi-squares method and the method of least squares. These methods are based on comparing the specified grouped distribution with its observed grouped distribution, or comparing the specified cumulative distribution function with its sample analogue, the empirical distribution function.

The method of weighted least squares which we shall propose in this paper is of the same kind as those ones. However, this method is sensitive to discrepancies at the tails of the distibution rather than near the median, because we would equalize the sampling error over the entire range by weighting the deviation by the reciprocal of the standard deviation. Therefore it will be expected that this method is effective for the parameter estimation problems in the mixed distribution composed of heterogeneous distributions to which the method of maximum likelihood is not applicable.

At first we shall formulate the problem and we shall give the existence theorem of an optimal solution for the weighted least squares problem. And then we shall compare the weighted least squares estimates with the estimates by the method of maximum likelihood based on individual observations.

§1. The Formulation of the Problem

Let $F(x\,;\,\theta)$ denote a distribution function of known mathematical form, containing unknown parameter θ, and suppose we have a random sample of size n from the corresponding population. We assume that we are not given the individual observations, but only the empirical distribution function

$$y_i = \frac{\text{no. of sample value} \leqq x_i}{n},$$

where $\{x_i\}$ is a sequence of m real numbers such as $-\infty < x_1 < x_2 < \ldots < x_m < +\infty$. Then

$y_1, \ldots, y_m$ satisfy the relation: $0 \leqq y_1 \leqq y_2 \leqq \ldots \leqq y_m \leqq 1$. For a given value of x, y is a binomial variable; it is distributed in the same way as the proportion of successes in n trials, where the probability of success is $F(x\,;\,\theta)$. Thus, $E[y]=F(x\,;\,\theta)$ and $V[y]=(1/n)F(x\,;\,\theta)(1-F(x\,;\,\theta))$. In order to equalize the sampling error over the entire range of x, we will weight the deviation by the reciprocal of the standard deviation; that is, we use

$$w(x)=n/(F(x\,;\,\theta)(1-F(x\,;\,\theta)))$$

as a weight function.

We define the function $Q(\theta)$ as follows;

$$Q(\theta)\equiv \sum_{i=1}^{m} w(x_i)(F(x_i\,;\,\theta)-y_i)^2=\sum_{i=1}^{m}\frac{n(F(x_i\,;\,\theta)-y_i)^2}{F(x_i\,;\,\theta)(1-F(x_i\,;\,\theta))}.$$

It seems natural to attempt to determine a best value of parameter θ so as to render $Q(\theta)$ as small as possible. This is called a weighted least squares estimation method.

Then we consider the following extremum problem;

(1) Choose an estimate $\theta \in \Theta$ so that θ minimizes $Q(\theta)$ for the given values of $(y_1, \ldots, y_m)$. To simplify the exposition, let Θ denote the upper-half plane in the 2-dimensional space R^2, that is, $\Theta=\{\theta=(\mu, \sigma) \in R^2\,;\, -\infty<\mu<+\infty \text{ and } \sigma>0\}$ and let $F(x\,;\,\theta)$ denote the normal distribution function;

$$F(x\,;\,\theta)=\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{x}\exp(-(s-\mu)^2/(2\sigma^2))\,ds.$$

Now we shall prove the existence theorem of an optimal solution of the problem (1).

§2. Preliminary Inequalities

We shall give three inequalities which are elementary but very important.

We have

Lemma 1. Let $\{a_i\}$ and $\{b_i\}$ be sets of m real numbers and let $b_1 \leqq b_2 \leqq \ldots \leqq b_m$ and put $r=\max\{i\,;\,b_i=b_1\}$. If $\sum_{i=1}^{m} a_i=0$ and $\sum_{i=k}^{m} a_i \geqq 0$ for each $k\,(r+1 \leqq k \leqq m)$, then

(2) $$\sum_{i=1}^{m} a_i b_i \geqq 0.$$

In particular, the strict inequality holds in (2) if $\sum_{i=r+1}^{m} a_i>0$ and $b_m-b_1>0$.

Proof. Let us put $A=\sum_{i=1}^{m} a_i b_i$. It follows from $\sum_{i=1}^{m} a_i=0$ that

$$A=\sum_{i=2}^{m} a_i b_i - \sum_{i=2}^{m} a_i b_1 = \sum_{i=2}^{m} a_i \ (b_i - b_1)$$

$$=\sum_{i=r+1}^{m-2} a_i(b_i-b_1)+a_{m-1}(b_{m-1}-b_1)+a_m(b_m-b_1).$$

Since $a_m \geqq 0$ and $b_m - b_1 \geqq b_{m-1} - b_1 \geqq 0$ we obtain

$$A \geqq \sum_{i=r+1}^{m-2} a_i(b_i-b_1)+(a_{m-1}+a_m)(b_{m-1}-b_1).$$

It is easily seen by the same arguement as above that

$$A \geqq \sum_{i=r+1}^{m-3} a_i(b_i-b_1)+(a_{m-2}+a_{m-1}+a_m)(b_{m-2}-b_1).$$

By repeating this process we have

(3) $$A \geqq (\sum_{i=r+1}^{m} a_i)(b_{r+1}-b_1) \geqq 0.$$

Our assertion follows from (3).

Corollary 1. Let $\{a_i\}$ and $\{b_i\}$ be increasing sequences of m real numbers. Then

(4) $$m \sum_{i=1}^{m} a_i b_i \geqq (\sum_{i=1}^{m} a_i)(\sum_{i=1}^{m} b_i).$$

In particular, the strict inequality holds in (4) if $a_m - a_1 > 0$ and $b_m - b_1 > 0$.

Proof. Let r be the same as in Lemma 1 and put $c_k = (\sum_{i=1}^{k} a_i)/k$ $(1 \leqq k \leqq m)$ and $a_i' = a_i - c_m$. Then $a_1' \leqq a_2' \leqq \dots \leqq a_m'$ and $\sum_{i=1}^{m} a_i' = 0$. It can be easily shown that $c_k \leqq c_{k'}$ $(1 \leqq k \leqq k' \leqq m)$ and that $c_k < c_m$ if $k < m$ and $a_m - a_1 > 0$. For each $k (r+1 \leqq k \leqq m)$ we have $\sum_{i=k}^{m} a_i' = (k-1)(c_m - c_{k-1}) \geqq 0$.

We have from Lemma 1

(5) $$\sum_{i=1}^{m} a_i' b_i \geqq 0,$$

which leads to (4). If $a_m - a_1 > 0$ and $b_m - b_1 > 0$, then $\sum_{i=r+1}^{m} a_i' > 0$, so that the strict inequality holds in (5) (or equivalently(4)) by Lemma 1.

Corollary 2. Let $\{a_i\}$, $\{b_i\}$ and $\{c_i\}$ be increasing sequences of m real numbers and let $\sum_{i=1}^{m} b_i > 0$. Assume that $a_i = a_j$ if and only if $b_i = b_j$. If $(a_i - a_j)/(b_i - b_j) \geqq (\sum_{i=1}^{m} a_i)/(\sum_{i=1}^{m} b_i)$ for every i and j such as $b_i \neq b_j$, then

(6) $$(\sum_{i=1}^{m} a_i)(\sum_{i=1}^{m} b_i c_i) \leqq (\sum_{i=1}^{m} a_i c_i)(\sum_{i=1}^{m} b_i).$$

In particular, the strict inequality holds in (6) if $c_m - c_1 > 0$ and there exist i and

j such that $b_i \neq b_j$ and $(a_i - a_j)/(b_i - b_j) > (\sum_{i=1}^{m} a_i)/(\sum_{i=1}^{m} b_i)$.

Proof. We put $a = \sum_{i=1}^{m} a_i$, $b = \sum_{i=1}^{m} b_i$ and $K = b\sum_{i=1}^{m} a_i c_i - a\sum_{i=1}^{m} b_i c_i$. Setting $d_i = a_i b - ab_i$, we obtain

$$K = \sum_{i=1}^{m} c_i d_i \text{ and } \sum_{i=1}^{m} d_i = 0.$$

For $1 \leqq i \leqq j \leqq m$, we have

$$d_j - d_i = \mathrm{b}(a_j - a_i) - a(b_j - b_i),$$

so that $d_i \leqq d_j$ by our assumption. By Corollary 1 we get $mK \geqq (\sum_{i=1}^{m} c_i)(\sum_{i=1}^{m} d_i) = 0$, hence $K \geqq 0$. If there exist i and $j (i < j)$ such that $b_i \neq b_j$ and $(a_i - a_j)/(b_i - b_j) > a/b$, then $d_i < d_j$, so that $K > 0$ by Corollary 1.

§ 3. Existence Theorem

Let $F(\theta)$ denote the row vector with components $F(x_i ; \theta)$. Note that the image $F(\Theta)$ of Θ under F is a bounded subset of m-dimensional Euclidean space R^m.

We define a subset $\partial F(\Theta)$ of R^m as follows: point $z \in R^m$ belongs to $\partial F(\Theta)$ if and if only $z \bar{\in} F(\Theta)$ and there exists a sequence $\{\theta_n\}$ in θ such that $F(\theta_n) \to z$ as $n \to \infty$. Let $L_i (1 \leqq i \leqq m)$ denote the line segment $\{z = (z_1, z_2, \dots, z_m) \in R^m ; z_j = 0 \ (j < i),\ 0 \leqq z_i \leqq 1$ and $z_j = 1 (i < j)\}$ and let L_o denote the line segment $\{z = (z_1, z_2, \dots, z_m) \in R^m ; z_1 = z_2 = \dots = z_m$ and $0 \leqq z_1 \leqq 1\}$.

We prove

Thorem 1. $\partial F(\Theta) = \bigcup_{i=o}^{m} L_i$.

Proof. We put $L = \bigcup_{i=o}^{m} L_i$. We shall show at first that $L \subset \partial F(\Theta)$. We put $t(x, \theta) = (x - \mu)/\sigma$ and $t_i(\theta) = t(x_i, \theta)$ for $\theta = (\mu, \sigma)$ and $x \in R$. And for a fixed $\theta \in \Theta$ we put $W_x(\theta) = \{\theta' \in \Theta ; F(x ; \theta') = F(x ; \theta)\} = \{\theta' = (\mu', \sigma') \in \Theta ; \mu' = x - t(x, \theta)\sigma'\}$. Note that

$$(7) \qquad t_i(\theta) = t(x, \theta) + (x_i - x)/\sigma \text{ for all } i \text{ and } F(x ; \theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t(x,\theta)} \exp(-s^2/2)\, ds.$$

Let $z = (z_1, z_2, \dots, z_m) \in L_k (1 \leqq k \leqq m)$. Then $z_i = 0 (i < k)$, $0 \leqq z_k \leqq 1$ and $z_i = 1 (k < i)$. In case that $z_k = 0$, taking $x \in R$ such as $x_k < x < x_{k+1}$, we have from (7) that $t_i(\theta_n) \to -\infty$ as $n \to \infty$ for each $i (i \leqq k)$ and $t_i(\theta_n) \to \infty$ as $n \to \infty$ for each $i (i > k)$, hence $F(\theta_n) \to z$ as $n \to \infty$ for any sequence $\{\theta_n\}$ in $W_x(\theta)$ such that $\theta_n \to (x, 0)$ as $n \to \infty$. Since $F(\Theta) \cap L = \phi$, we have $z \in \partial F(\Theta)$. In case that $z_k = 1$, taking $x \in R$ such that $x_{k-1} < x < x_k$, we have from (7) that

$t_i(\theta_n) \to -\infty$ as $n \to \infty$ for each $i (i < k)$ and $t_i(\theta_n) \to \infty$ as $n \to \infty$ for each $i (i \geqq k)$, hence $F(\theta_n) \to z$ as $n \to \infty$ for any sequence $\{\theta_n\}$ in $W_x(\theta)$ such that $\theta_n \to (x, 0)$ as

$n \to \infty$. Thus $z \in \partial F(\Theta)$. In case that $0<z_k<1$, there exists $\theta_o \in \Theta$ such that $F(x_k; \theta_o)=z_k$. It follows from (7) that

$t_i(\theta_n) \to -\infty$ as $n \to \infty$ for each $i(i<k)$ and $t_i(\theta_n) \to \infty$ as $n \to \infty$ for each $i(i>k)$,

hence $F(\theta_n) \to z$ as $n \to \infty$ for any sequence $\{\theta_n\}$ in $W_{x_k}(\theta_o)$ such that $\theta_n \to (x_k, 0)$ as $n \to \infty$. Thus $z \in \partial F(\Theta)$. Finally we consider the case that $z=(z_1, z_2, \ldots, z_m) \in L_o$. In case that $z_1=0$ or 1, we have shown previously that $z \in \partial(\Theta)$, so that we may as sume that $0<z_1<1$. Then there exists $\theta_o \in \Theta$ such that $F(x_1; \theta_o)=z_1$. It follows from (7) that

$$t_i(\theta_n) \to t_1(\theta_o) \text{ as } n \to \infty \text{ for all } i,$$

hence $F(\theta_n) \to z$ as $n \to \infty$ for any sequence $\{\theta_n\}$ in $W_{x_1}(\theta_o)$ such that $\sigma_n \to \infty$ as $n \to \infty$, where $\theta_n=(\mu_n, \sigma_n)$. Thus $z \in \partial F(\Theta)$. Next we shall establish the converse inclusion. Let $z \in \partial F(\Theta)$. Then there exists a sequence $\{\theta_n\}$ in Θ such that $F(\theta_n) \to z$ as $n \to \infty$ and $z \bar{\in} F(\Theta)$. This implies that the sequence $\{\theta_n\}$ has no cluster point in Θ. Consider m sequences $\{t_i(\theta_n)\}$. Suppose first that one of these sequences has a convergent subsequence with a finite limit. Without loss of generality we may assume that $\{t_j(\theta_n)\}$ has a finite limit. We put $\theta_n=(\mu_n, \sigma_n)$. It can be easily seen that $\underline{\lim}_{n \to \infty} \sigma_n=0$ or ∞. In case that $\underline{\lim}_{n \to \infty} \sigma_n=\infty$, we can choose a subsequence $\{\theta_{n'}\}$ by (7) such that $\lim_{n' \to \infty} t_i(\theta_{n'})=\lim_{n' \to \infty} t_j(\theta_{n'})$ for all i, so that $z \in L_o$. In case that $\underline{\lim}_{n \to \infty} \sigma_n=0$, we can find a subsequence $\{\theta_{n'}\}$ by (7) such that $\lim_{n' \to \infty} t_i((\theta_{n'})=-\infty (i<j)$ and $\lim_{n' \to \infty} t_i(\theta_{n'})=\infty (j<i)$, so that $z \in L_j$. In both cases $z \in L$. Secondly we assume that each sequence $\{t_i(\theta_n)\}$ has no convergent subsequence with a finite limit. The each sequence has a subsequence which converges to $-\infty$ or ∞. If $\{t_1(\theta_n)\}$ has a subsequence $\{\theta_{n'}\}$ such that $\lim_{n' \to \infty} t_1(\theta_{n'})=\infty$, $z=(1, 1, \ldots, 1) \in L_1$. Thus we may assume that $\{t_1(\theta_n)\}$ converges to $-\infty$. If $\{t_2(\theta_n)\}$ has a subsequence which converges to ∞, then we can find a subsequence $\{\theta_{n'}\}$ by (7) such that $\lim_{n' \to \infty} t_i(\theta_{n'})=\infty$ for all $i(i \geqq 2)$, so that $z=(0, 1, 1, \ldots, 1) \in L_2$. Hence we may assume that $\{t_2(\theta_n)\}$ converges to $-\infty$. By repeating this argument, finaly we may assume that all $\{t_i(\theta_n)\}$ converge to $-\infty$. In this case we have $z=(0, 0, \ldots, 0) \in L_o$. This completes the proof.

By this theorem we can determine the boundary value of $Q(\theta)$. Hereafter we regard $Q(\theta)/n$ as $Q(\theta)$.

Lemma 2. Assume that there exist i and j such that $0<y_i<y_j<1$. Then

$$\lim_{n \to \infty} Q(\theta_n)=\infty$$

for any sequence $\{\theta_n\}$ in Θ such that $\lim_{n \to \infty} F(\theta_n) \in \partial F(\Theta)-L_o$.

Proof. We have by definition

$$Q(\theta_n) \geqq [F(x_i; \theta_n)(1-F(x_i; \theta_n))]^{-1}(F(x_i; \theta_n)-y_i)^2 + [F(x_j; \theta_n)(1-F(x_j; \theta_n))]^{-1}(F(x_j; \theta_n)-y_j)^2,$$

where $\{\theta_n\}$ is the sequence in Θ such that $\lim_{n\to\infty} F(\theta_n) \in \partial F(\Theta) - L_0$. There exists L_k $(1 \leqq k \leqq m)$ such that $\lim_{n\to\infty} F(\theta_n) = (a_1, a_2, \dots a_m) \in L_k$ by Theorem 1, so that either $F(x_i ; \theta_n)(1-F(x_i ; \theta_n))$ or $F(x_j ; \theta_n)(1-F(x_j ; \theta_n))$ converges to 0 as $n\to\infty$. Since $0<y_i<y_j<1$, we have

$$\text{either } (a_i-y_i)^2>0 \text{ or } (a_j-y_j)^2>0,$$

so that $Q(\theta_n)\to\infty$ as $n\to\infty$.

Let us define functions $Q(\lambda)$ and $f(\lambda)$ on $(0,1)$ as follows:

$$Q(\lambda)\equiv[\lambda(1-\lambda)]^{-1}\sum_{i=1}^{m}(\lambda-y_i)^2 \text{ and } f(\lambda)\equiv(m-2\sum_{i=1}^{m}y_i)\lambda^2+2(\sum_{i=1}^{m}y_i^2)\lambda-\sum_{i=1}^{m}y_i^2.$$

We obtain

Proposition 1. Assume that $0<y_i<1$ for some i. Then there exists a unique value λ^* of λ minimizing $Q(\lambda)$ over $(0, 1)$. Moreover λ^* is a solution of the equation $f(\lambda)=0$.

Proof. Note that $Q(\lambda)\geqq[\lambda(1-\lambda)]^{-1}(\lambda-y_i)^2\to\infty$ as $\lambda\to0$ or $\lambda\to1$. By the continuity of $Q(\lambda)$ we can find λ^* in $(0, 1)$ which minimizes $Q(\lambda)$ over $(0, 1)$. Therefore the derivative $Q'(\lambda^*)$ of $Q(\lambda)$ at λ^* vanishes. It is easily seen that $0=Q'(\lambda^*)=[\lambda^*(1-\lambda^*)]^{-2}f(\lambda^*)$, so that $f(\lambda^*)=0$. Assume that $\lambda' \in (0, 1)$ mininizes $Q(\lambda)$ over $(0, 1)$. Then $f(\lambda')=0$. Since the equation $f(\lambda)=0$ has only one solution in $(0, 1)$, we have $\lambda'=\lambda^*$.

Lemma 3. Let λ^* be a solution of $f(\lambda)=0$ and let $\bar{y}$ be the mean value of $\{y_i\}$. Then

(i) $\lambda^*<1/2$ if and only if $\bar{y}>1/2$.

(ii) $\lambda^*=1/2$ if and only if $\bar{y}=1/2$.

(iii) $\lambda^*>1/2$ if and only if $\bar{y}>1/2$.

Proof. Setting $Y=\sum_{i=1}^{m}y_i^2$, we see that

$f(\lambda)=m(1-2y)\lambda^2+2Y\lambda-Y$, so that $f(1/2)=m(1-2\bar{y})/4$. Thus we can easily prove (i), (ii) and (iii).

Let $\lambda \in (0, 1)$ and let Φ^{-1} be the inverse function of the standard normal distribution function. For any real number x, denote by $D_\lambda(x)$ the half-line in Θ with inclination $-(\Phi^{-1}(\lambda))^{-1}$ through the point $(x, 0)$, i.e.,

$$D_\lambda(x)=\{\theta=(\mu, \sigma) \in \Theta ; \mu=x-\Phi^{-1}(\lambda)\sigma\}.$$

We set

$A_i(x)=(x-x_i)(1-2y_i)$, $B_i(x)=(x-x_i)y_i^2$, $g(\lambda)=[\lambda(1-\lambda)]^{-2}\exp(-(\Phi^{-1}(\lambda))^2/2)$ and

$$G(\lambda, x)=g(\lambda)\sum_{i=1}^{m}[A_i(x)\lambda^2+2B_i(x)\lambda-B_i(x)]. \quad (8)$$

Let $P(\theta)$ be an arbitrary real-valued function defined on Θ. For $\theta=(\mu, \sigma) \in D_\lambda(x)$, denote $P_\lambda'(\theta)$ the directional derivative of $P(\theta)$ at θ along the half-line $D_\lambda(x)$, i.e.,

$$P_\lambda'(\theta)=\lim_{\sigma'\to\sigma}(P(\theta')-P(\theta))/(\sigma'-\sigma)$$
$$\theta'=(\mu',\ \sigma')\ \varepsilon\ D_\lambda(x)$$

if the right side of the above exists.

We obtain

Lemma 4. $\lim_{\sigma\to\infty}\sigma^2 Q_\lambda'(\theta)=(2\pi)^{-1/2}G(\lambda,\ x)$.

Proof. We set $h(\theta)=F(x_i\,;\theta)$ and $q(\theta)=[h(\theta)(1-h(\theta)]^{-1}(h(\theta)-y_i)^2$. By the relation

$$h_\lambda'(\theta)=[(x-x_i)/(\sqrt{2\pi}\,\sigma^2)]\exp(-t_i(\theta)^2/2),$$

we have

(9) $$q_\lambda'(\theta)=[(x-x_i)/(\sqrt{2\pi}\,\sigma^2)][\exp(-t_i(\theta)^2/2)][h(\theta)(1-h(\theta))]^{-2}$$
$$(h(\theta)-y_i)(h(\theta)+y_i-2y_ih(\theta)).$$

Noting that

$$t_i(\theta)^2=[(x_i-x)/\sigma+\Phi^{-1}(\lambda)]^2\to(\Phi^{-1}(\lambda))^2 \text{ as } \sigma\to\infty,$$

we have

(10) $$\sigma^2 h_\lambda'(\theta)\to[(x-x_i)/\sqrt{2\pi}]\exp[-(\Phi^{-1}(\lambda))^2/2] \text{ as } \sigma\to\infty$$

and

(11) $$h(\theta)\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\Phi^{-1}(\lambda)}\exp(-s^2/2)ds=\lambda \text{ as } \sigma\to\infty.$$

It follows from (9), (10) and (11) that
$\lim_{\sigma\to\infty}\sigma^2 Q_\lambda'(\theta)=(2\pi)^{-1/2}G(\lambda,\ x)$. This completes the proof.

Lemma 5. Let λ^* be a solution of $f(\lambda)=0$ and assume that there exist i and j such that $0<y_i<y_j<1$. Then $G(\lambda^*,\ x)$ is positive, i.e., $G(\lambda^*,\ x)>0$.

Proof. Setting $a=\sum_{i=1}^m(1-2y_i)$, $b=\sum_{i=1}^m y_i^2$, $c=\sum_{i=1}^m x_i(1-2y_i)$ and $d=\sum_{i=1}^m x_iy_i^2$, we have $f(\lambda^*)=a(\lambda^*)^2+2b\lambda^*-b=0$ and $G(\lambda^*,\ x)=xg(\lambda^*)f(\lambda^*)-g(\lambda^*)[c(\lambda^*)^2+2d\lambda^*-d]$, so that

$$-aG(\lambda^*,\ x)=g(\lambda^*)(ad-bc)(2\lambda^*-1)$$

or

(12) $$m(2\bar{y}-1)G(\lambda^*,x)=g(\lambda^*)(ad-bc)(2\lambda^*-1)$$

where $\bar{y}$ is the mean value of $\{y_i\}$. We show first that $K=ad-bc$ is strictly positive. Putting $a_i=2y_i-1$ and $b_i=y_i^2$, we have

$$K=(\sum_{i=1}^m a_ix_i)(\sum_{i=1}^m b_i)-(\sum_{i=1}^m a_i)(\sum_{i=1}^m b_ix_i).$$

We have only to verify that $\{a_i\}$, $\{b_i\}$ and $\{x_i\}$ satisfy the conditions of Corollary

2 of Lemma 1. It is clear by our assumption that $a_1 \leqq a_2 \leqq \ldots \leqq a_m$, $b_1 \leqq b_2 \leqq \ldots \leqq b_m$, $\sum_{i=1}^{m} b_i > 0$ and $a_i = a_j$ if and only if $b_i = b_j$. In case that $b_i \neq b_j$, we put $U = (a_i - a_j)/(b_i - b_j) = 2/(y_i + y_j) (>0)$ and $V = (\sum_{i=1}^{m} a_i)/(\sum_{i=1}^{m} b_i) = m(2\bar{y}-1)/b$. It is clear that $V > U$ if $\bar{y} \leqq 1/2$. If $\bar{y} > 1/2$, then it can be shown that $\bar{y}^2/(2\bar{y}-1) > 1$. We have by Schwarz's inequality $\bar{y}^2 \leqq b/m$, so that $\bar{y}^2/(2\bar{y}-1) \leqq b/[m(2\bar{y}-1)]$. Therefore $m(2\bar{y}-1)b < 1$. Since $y_i \neq y_j$, it follows that $y_i + y_j < 2$, so that $V > U$. Thus $K > 0$. Next we shall prove that $G(\lambda^*, x) > 0$. Since $g(\lambda^*) > 0$, by the aid of Lemma 3 and (12) we conclude that $G(\lambda^*, x) > 0$ if $y \neq 1/2$. We consider the case that $\bar{y} = 1/2$. Then $\lambda^* = 1/2$, so that $G(1/2, x) = -g(1/2)c/4$. We can show that $c < 0$ by Corollary 1 of Lemma 1. This completes the proof.

Now we are ready for the main theorem.

Theorem 2. Assume that there exist i and j such as $0 < y_i < y_j < 1$. Then the problem (1) has an optimal solution.

Proof. Let λ^* be a solution of $f(\lambda) = 0$. It follows from Lemma 4 and Lemma 5 that

$$Q'_{\lambda^*}(\theta) > 0 \text{ for sufficiently large } \sigma.$$

This implies that there exists $\theta_o \in D_{\lambda^*}(x)$ such as $Q(\theta_o) < Q(\lambda^*)$, since $Q(\theta) \to Q(\lambda^*)$ as $\sigma \to \infty$ whenever $\theta = (\mu, \sigma) \in D_{\lambda^*}(x)$ We put $S = \{\theta \in \Theta ; Q(\theta) \leqq Q(\theta_o)\}$. Let $\{\theta_n\}$ be any sequence in S and suppose that $\{\theta_n\}$ has no cluster point in S, that is, in Θ. We can find a subsequence $\{\theta_{n'}\}$ of $\{\theta_n\}$ such that $F(\theta_{n'}) \to z \in \partial F(\Theta)$ as $n' \to \infty$ (see the proof of Theorem 1). Thus $\lim_{n' \to \infty} Q(\theta_{n'}) \geqq Q(\lambda^*)$, so that $Q(\theta_o) \geqq Q(\lambda^*)$. This is a contradiction. Therefore all cluster points of $\{\theta_n\}$ belong to S, so that S is a compact set in Θ. By the continuity of $Q(\theta)$ we conclude that there exists $\theta^* \in \Theta$ such that $Q(\theta^*) = \min\{Q(\theta); \theta \in \Theta$.

§ 4. Practical Estimation Procedure

It has been proved that there exists an optimal solution for the problem (1). In order to find an optimal solution θ, we have to solve the equations $\partial Q/\partial\mu = 0$, $\partial Q/\partial\sigma^2 = 0$. But these equations are so complicated that an iterative method must be used to find an optimal solution, starting from some initial value. To simplify the calculations we shall regard the denominator in $Q(\theta)$ as constant when differentiating, and shall assume in addition that the $F(x_i ; \mu_o + \Delta\mu, \sigma_o^2 + \Delta\sigma^2)$ are linear functions of $\Delta\mu$ and $\Delta\sigma^2$. Under these assutions, substituting $\mu = \mu_o + \Delta\mu$, $\sigma^2 = \sigma_o^2 + \Delta\sigma^2$ for $\partial Q/\partial\mu = 0$ and $\partial Q/\partial\sigma^2 = 0$ gives

$$(13) \quad \begin{cases} (\sum_{i=1}^{m} w_{io}\varphi_{io}^2)\Delta\mu + \dfrac{1}{2\sigma_o^2}(\sum_{i=1}^{m} w_{io}t_{io}\varphi_{io}^2)\Delta\sigma^2 = \sigma_o \sum_{i=1}^{m} w_{io}\varphi_{io}(F(x_i; \mu_o, \sigma_o^2) - y_i), \\ (\sum_{i=1}^{m} w_{io}t_{io}\varphi_{io}^2)\Delta\mu + \dfrac{1}{2\sigma_o^2}(\sum_{i=1}^{m} w_{io}t_{io}^2\varphi_{io}^2)\Delta\sigma^2 = \sigma_o \sum_{i=1}^{m} w_{io}t_{io}\varphi_{io}(F(x_i; \mu_o, \sigma_o^2) - y_i), \end{cases}$$

where $t_{io}=(x_i-\mu_o)/\sigma_o$, $\varphi_{io}=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t_{io}^2}{2}\right)$, $w_{io}=(\Phi_{io}(1-\Phi_{io}))^{-1}$, $\Phi_{io}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t_{io}}\exp\left(-\frac{s^2}{2}\right)ds=\frac{1}{\sqrt{2\pi}\,\sigma_o}\int_{-\infty}^{x_i}\exp(-(s-\mu_o)^2/(2\sigma_o^2))ds=F(x_i;\,\mu_o,\,\sigma_o^2)$. The iterative weighted least squares estimating procedure is, then, as follows.

(1°) Set initial values μ_o and σ_o^2.

(2°) Solve the system (13) for $\Delta\mu$ and $\Delta\sigma^2$ and calculate μ_1 and σ_1^2 by the following equations respectively.

$$\mu_1=\mu_o+\Delta\mu,$$
$$\sigma_1^2=\sigma_o^2+\Delta\sigma^2.$$

(3°) Set μ_1, σ_1^2 as initial values in place of μ_o, σ_o^2 respectively.

Continue steps (1°)~(3°) until the latest set of adjustments is negligible. We shall denote the latest set of estimates by μ, σ^2. These will be the optimal solution of the weighted least squares problem (1).

The sequences of $\{\mu\}$ and $\{\sigma^2\}$ may fail to converge in particular cases. Even when they do converge, if the weighted least squares equation has multiple solutions there is no guarantee that they will converge to the solutions corresponding to the absolute minimum of $Q(\theta)$. In most cases it seems to converge to the optimum solution.

§ 5. Comparison between WLS Estimates and ML Estimates

We now illustrate the computational procedure of our method using the argorithm as described above and examine the goodness of an WLS estimates. For this purpose we generated 1000 samples of size 100 from the normal population with mean 100 and variance 100, and computed WLS estimates $\tilde{\mu}$, $\tilde{\sigma}^2$ based on $(y_1, \ldots, y_7)$ and ML estimates $\hat{\mu}$, $\hat{\sigma}^2$ based on 100 individual observations for each sample. Here we set $m=7$ and $x_1=85$, $x_2=90$, $x_3=95$, $x_4=100$, $x_5=105$, $x_6=110$, $x_7=115$, and $y_1, \ldots, y_7$ were calculated by the formula as mentioned for each sample. We generated uniform random numbers by the Lehmer's formula;

$$v_{i+1}=630360016v_i \text{ (Mod } 2^{31}-1),$$

where $v_o=418369880$, and tranformed them into standard normal numbers by the method of Box-Muller, and then converted them into normal random numbers with mean 100 and variance 100.

Table 1 presents the results of simulation. Figure 1 presents the correlation diagram between $\tilde{\mu}$ and $\hat{\mu}$ and figure 2 presents the correlation diagram between $\tilde{\sigma}^2$ and $\hat{\sigma}^2$. From these results it looks that WLS estimates based on $(y_1, \ldots y_7)$ are as good as ML estimates based on 100 individual observations. As to the precision of estimation it seems that there was no difference between WLS estimates and ML estimates. If there is no loss of information caused by grouping date—this means simplification of measuring—as far as these kind of parameter estimation can be performed, it is very important for practical statisticians.

Table 1. Comparison between WLS Estimates and ML Estimates-Results of Simulation (on 1000 samples of size 100 from N(100, 10))-

	WLS E on $(y_1, \ldots, y_7)$	ML E on observations
MEAN	$E(\tilde{\mu})=99.9783$ $V(\tilde{\mu})=1.0370$	$E(\hat{\mu})=99.9850$ $V(\hat{\mu})=0.9710$
VARIANCE	$E(\tilde{\sigma}^2)=99.6091$ $V(\tilde{\sigma}^2)=274.3744$	$E(\hat{\sigma}^2)=99.8026$ $V(\hat{\sigma}^2)=210.3664$

Figure 1. Correlation diagram between $\tilde{\mu}$ and $\hat{\mu}$.

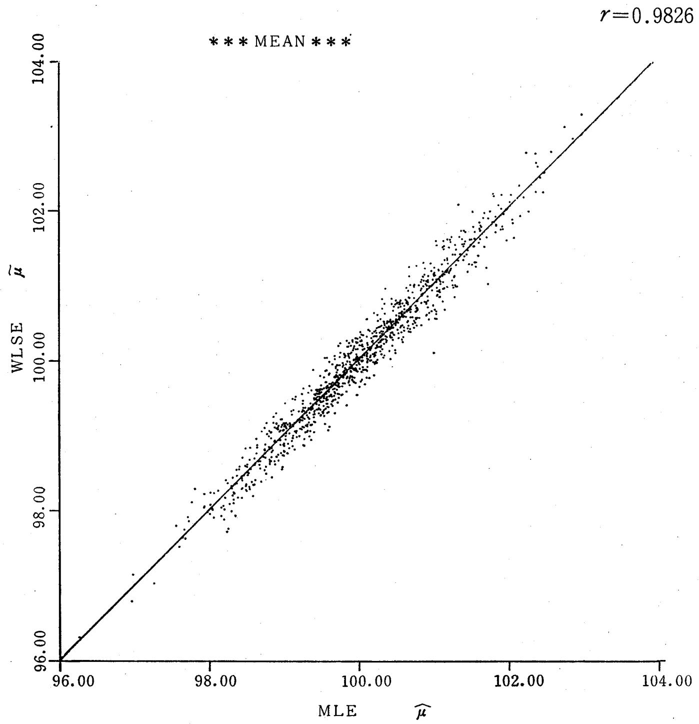
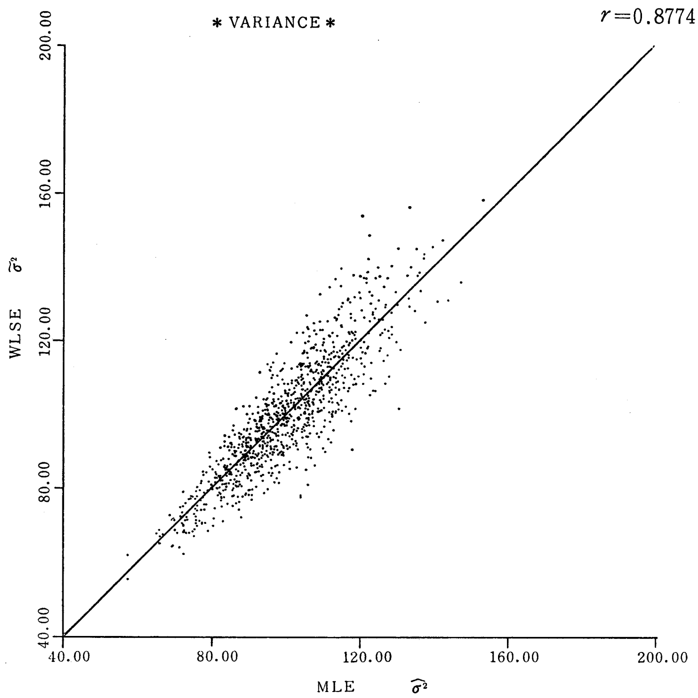


Figure 2. Correlation diagram between $\tilde{\sigma}^2$ and $\hat{\sigma}^2$.



References

[1] Hartley, H. O. (1961); The Modified Gauss-Newton Method for the Fitting of Non-Linear Regression Functions by Least Squares. Technometrics **3**, 269-280.

[2] Anderson, T. W. and D. A. Darling (1954); A Test of Goodness of Fit. J. Amer. Statist. Ass., **49**, 765-769.

[3] Yanagawa, A. and S. Shirohata (1974); Unbiased estimates of the population mean on ranked set (Japanese). Ohyo Tokeigaku., Vol. **4**, No. 2, 55-63.