Pertanika 6(1), 57–62 (1983)

An Efficient Maximum Likelihood Solution in Normal Model having Constant but Unknown Coefficients of Variation

MOHD. NAWI ABD-RAHMAN and THOMAS M. GERIG[1]
Department of Mathematics, Faculty of Science and Environmental Studies
Universiti Pertanian Malaysia, Serdang, Selangor, Malaysia.

Key words: **Coefficient of variation: Maximum likelihood procedure; Monte Carlo evaluations; Efficient estimators; Asymptotic relative efficiencies.**

RINGKASAN

Beberapa populasi normal merdeka yang mempunyai pekali ubahan pemalar tetapi tidak diketahui telah dipertimbangkan. Modelnya dalam bentuk lebih itlak. Bilangan kumpulan tidak terhad dan saiz sampel untuk setiap kumpulan mungkin berbeza-beza. Suatu kaedah penyelesaian berasaskan tatacara kebolejadian maksima telah dibentuk. Persamaan-persamaan kebolehjadian maksimum dirangkaikan kepada satu persamaan tunggal sahaja. Ini menghasilkan suatu penyelesaian berangka yang tepat. Penilaian Monte Carlo dikaji. Contoh dari makalah diambil untuk menunjukkan kaedah pemakaiannya. Penganggar bagi min telah ditunjukkan lebih cekap secara asimptot dari min biasa. Kecekapan sebanding asimptot meningkat apabila saiz sampel sebandingnya meningkat.

SUMMARY

A number of independent normal normal populations having constant but unknown coefficients of variation are considered. The model is of more general form. There is no restriction on the number of groups and the sample size in each group may differ from one another. An efficient method of solutions based on the maximum likelihood procedure is developed. The maximum likelihood equations are reduced to a single equation. This results in a numerically exact solutions. Monte Carlo evaluations are studied. Examples from the literature are taken to illustrate the method. The estimators for the means are shown to be asymptotically more efficient than the ordinary means. The asymptotic relative efficiency increases as the relative sample size increases.

INTRODUCTION

In many technical and biological applications researchers may not be willing to assume that their normal models have constant or homogeneous variances particularly when the observations are taken in groups at various points of time under different conditions. An alternative model that has recently been proposed is the less realistic assumption of constant and known coefficients of variation. We propose here a more general alternative model without the assumption that the coefficients of variation are known. The model is in its most general form in the sense of unrestricted number of groups of different sample sizes.

We consider k independent normal populations with $y_{ji} \sim NID(\mu_j, c^2 \mu_j^2)$, $i = 1, ..., n_j$, $j = 1, ..., k$. We assume the parameter space as the positive orthant

$$Q = \{ \mu_1, ..., \mu_k, c \mid \mu_j \in R^1_+, j = 1, ..., k, c \in R^1_+ \}.$$

Solutions for $k = 2$ and $n_j = n$, all j, are discussed in Lohrding (1969). When $n_j = n_j'$, $j \neq j'$, the solutions become fairly complicated even with $k = 2$. It is therefore essential to develop an efficient numerical method of estimation. We note that there is a possibility that real solutions do not exist in this case.

[1] Department of Statistics, N.C. State University, Raleigh, N.C. 27607 U.S.A.
Key to authors' names: M.N. Abd-Rahman, T.M. Gerig.

Basically, the method involves some simplications of the k + 1 equations into a single equation of the form $F(x) = x$. It is also established that both lower and upper bounds exist for $\hat{c}^2$. This enables us to apply the method of bisection (Kelly, 1967). Very accurate numerical solutions are possible.

It can be shown that the distribution function for this model satisfies the regularity conditions of Bradley and Gart (1962) (see Abd-Rahman 1978).

THEORY

Let L represent the likelihood function of k random samples $y_{j1}, \ldots, y_{jn_j}$ of size n_j, $j = 1, \ldots, k$. Denote log L by ℓ. The log-likelihood of $\theta' = (\mu_1, \ldots, \mu_k, c)$ given the random samples is

$$\ell = \text{const} - \sum_{j=1}^{k} [\, n_j \log \mu_j + n_j \log c + (2c^2 \mu_j^2)^{-1} \sum_{i=1}^{n_j} (y_{ji} - \mu_j)^2 \,].$$

Differentiating ℓ partially with respect to (w.r.t.) the parameters $\mu_1, \ldots, \mu_k$, and c, and on equating to zero, the likelihood equations may be adjusted to obtain the k + 1 equations as follows:

$$(\hat{c}^2 + 1)\hat{\mu}_j^2 - \hat{\mu}_j \bar{y}_j - \psi_j^2 = 0, \; j = 1, \ldots, k, \tag{2.1}$$

where $\bar{y}_j = n_j^{-1} \sum_i y_{ji}$, and $\psi_j^2 = n_j^{-1} \sum_i (y_{ji} - \hat{\mu}_j)^2$, and $\hat{c}^2 = n^{-1} \sum_j n_j \psi_j^2 / \hat{\mu}_j^2$, (2.2)

where $n = \sum_j n_j$. Substituting ψ_j^2 of (2.1) in (2.2) we have

$$\sum_j (n_j/\hat{\mu}_j)\, \bar{y}_j = n. \tag{2.3}$$

Also, if $s_j^2 = n_j^{-1} \sum_i (y_{ji} - \bar{y}_j)^2$, then (2.1) can be reduced to

$$\hat{c}^2 \hat{\mu}_j^2 + \hat{\mu}_j \bar{y}_j - (s_j^2 + \bar{y}_j^2) = 0, \; j=1, \ldots, k. \tag{2.4}$$

On dividing equation (2.4) by $\bar{y}_j^2$, we obtain

$$\hat{c}^2 (\hat{\mu}_j/\bar{y}_j)^2 + (\hat{\mu}_j/\bar{y}_j) - (t_j^2 + 1) = 0, \tag{2.5}$$

where $t_j^2 = s_j^2/\bar{y}_j^2$, $j=1, \ldots, k$. Considered as quadratic equations in $(\hat{\mu}_j/\bar{y}_j)$, we have

$$(\hat{\mu}_j) = [-1 + \{1 + 4\hat{c}^2 (t_j^2 + 1)\}^{\frac{1}{2}}]/2\hat{c}^2, \tag{2.6}$$

taking only the positive solutions. We note that $4\hat{c}^2 (t_j^2 + 1) > 0$. Substituting this quantity (2.6) in (2.3) and simplifying we get the required form of $F(\hat{c}^2) = \hat{c}^2$, namely

$$(n/2) / \sum_j n_j / [-1 + \{1 + 4\hat{c}^2 (t_j^2 + 1)\}^{\frac{1}{2}}] = \hat{c}^2. \tag{2.7}$$

The solution is a zero of $f(\hat{c}^2) = F(\hat{c}^2) - \hat{c}^2$, where F, and hence f, is differentiable to any order. To determine whether the method of bisection can be applied to solve equation (2.7) it is sufficient to show that $f'x) < 0$ at its zero.

Denote $\hat{c}^2$ by x, and let

$$a_j(x) = -1 + \{1 + 4x(t_j^2 + 1)\}^{\frac{1}{2}}, \tag{2.8}$$

which is greater than zero. Then

$$a_j'(x) = \frac{d}{dx} a_j(x) = (2t_j^2 + 1)/\{1 + 4x(t_j^2 + 1)\}^{\frac{1}{2}}. \tag{2.9}$$

Writing a_j for $a_j(x)$, the function F under consideration is $F(x) = \frac{1}{2} (\sum_j P_j a_j^{-1})^{-1}$, where $P_j = n_j/n$. Thus

$F'(x) = \frac{1}{2} (\sum_j P_j a_j^{-1})^{-2} (\sum_j P_j a_j^{-2} a_j')$. By equations (2.8) and (2.9) we have that $a_j' = 2a_j (a_j + 2) / 4x (a_j + 1)$, and hence $F'(x) = (\sum_j P_j a_j^{-1})^{-2} \{\sum_j P_j a_j^{-2} a_j (a_j + 2)/4x (a_j + 1)\}$. When $F(x) = a$, then $F'(x) = x \{\sum_j P_j (a_j + 2)/aj (a_j + 1)\} = \sum_j q_j \cdot \frac{1}{2} (a_j + 2) / (a_j + 1)$, where $q_j = P_j a_j^{-1} / \sum_j P_j a_j^{-1}$. That is, $F'(x)$ is the weighted average of $\frac{1}{2} (a_j + 2) / (a_j + 1)$. Since $\frac{1}{2} (a_j + 2)/(a_j + 1) < 1$ then we have $F'(x) < 1$, or $f'(x) < 0$. This holds for all x or $\hat{c}^2$ on the entire positive half-line.

Upper bound. For the model under consideration the quantity

$$c_u^2 = n^{-1} \sum_j n_j (s_j/\bar{y}_j)^2 \qquad (2.10)$$

is an upper bound for $\hat{c}^2$.

For a proof of this claim we visualize the expressions $\sum (P_j/a_j)^{-1}$ as a harmonic mean of a_j, and $\sum_j P_j a_j$ as the expectation of a_j. By the well known relationship between them and a Cauchy-Schwarz inequality we have $\hat{c}^2 = \frac{1}{2} (\sum_j P_j a_j^{-1})^{-1} \leq \frac{1}{2} \sum_j P_j a_j = \frac{1}{2} \sum_j P_j [-1 + \{1 + 4\hat{c}^2 (t_j^2 + 1)\}^{\frac{1}{2}}]$

$= \frac{1}{2} [-1 + \sum_j P_j^{\frac{1}{2}} P_j^{\frac{1}{2}}{}_2 \{1 + 4\hat{c}^2 (t_j^2 + 1)\}^{\frac{1}{2}}] \leq \frac{1}{2} [-1 + [\sum_j P_j \sum_i P_j \{1 + 4\hat{c}^2 (t_j^2 + 1)\}^{\frac{1}{2}}$. This implies that $(2\hat{c}^2 + 1)^2 \leq 1 + 4\hat{c}^2 (\sum_j P_j t_j^2 + 1)$. But $\sum_j P_j t_j^2 = n^{-1} \sum_j n_j (s_j/\bar{y}_j)^2$. Hence we finally have $\hat{c}^2 \leq \hat{c}_u^2$. Thus c_u^2 is an upper bound for $\hat{c}^2$.

Lower bound. For the same model, let $t_1 \leq t_2 \leq ... \leq t_k$ where t's are defined as before Then $c^2 \leq t_1$.

We prove it as follows. We first note that the assumption that $t_1 \leq t_2 \leq ... \leq t_k$ is made for convenience, with no loss in generality. By the inequalities $\min_j (x_j) \leq \sum_j w_j x_j \leq \max (x_j)$, where w_j is some weight, $j=1, ..., k$, we have $\hat{c}^2 = \frac{1}{2} (\sum_j P_j a_j^{-1})^{-1} \geq \frac{1}{2} a_1$, since $a_j = -1 + \{1 + 4\hat{c}^2 (t_j^2 + 1)\}^{\frac{1}{2}}$ is also such that $a_1 \leq a_2 \leq ... \leq a_k$. Then $\hat{c}^2 \geq t_1^2$. That is, t_1^2 is a lower bound for $\hat{c}^2$. In practice, we find that $c_L^2 = (\sum_j n_j t_j / \sum_j n_j)^2$ provides a more efficient lower bound for $\hat{c}^2$. It may be of some interest to compare this finding with table 1 of Zeigler (1973).

Using c_u^2 and c_L^2 as the two bounds in the bisection method we obtain the solution for $\hat{c}^2$ in equation (2.7) based on the following criterion of convergence: compute $p = F(c_i^2)/c_i^2$ at the ith iteration, and if $p = 1.0 \pm 10^{-16}$ we say that the convergence is attained and put $\hat{c}^2 = c_i^2$. Then $\hat{\mu}_1, ..., \hat{\mu}_k$ are obtained using equation (2.6).

ASYMPTOTIC RELATIVE EFFICIENCY (A.R.E.)

Suppose $\theta = (\mu_1, ..., \mu_k, c)'$. From $E_\theta (-\partial^2 \theta / \partial\theta_r \partial\theta_s)$ we obtain (r, s) element of the information matrix $I(\theta)$. This is given by

$$I_{rs}(\theta) = \begin{cases} (2c^2 + 1)n_r/c^2\mu_r & \text{if } r = s \\ 0 & \text{if } r \neq s,\ r,s = 1, ..., k, \end{cases}$$

$$I_{k+1,s}(\theta) = I_{s,k+1}(\theta) = 2n_s/c^2\mu_s\ , s = 1, \ldots k, \text{ and}$$

$$I_{k+1,k+1}(\theta) = 2n/c^2.$$

This symmetric matrix is of a familiar form, and is easily invertable. The asymptotic variance-covariance matrix of the estimators is given by

$$V_{rs}(\theta) = \begin{cases} c^2\mu_r^2(1 + 2n_rc^2/n)/n_r(2c^2 + 1) & \text{if } r = s \\ 2\mu_r\mu_s\, c^4/n(2c^2 + 1) & \text{if } r \neq s,\ r,s = 1, \ldots, k, \end{cases}$$

$$V_{k+1,s}(\theta) = V_{s,k+1}(\theta) = -\,c^3\mu_s/n\ ,\ s = 1, \ldots, k, \text{ and}$$

$$V_{k+1,k+1}(\theta) = c^2(2c^2 + 1)/2n.$$

The A.R.E. of $\hat{\mu}_1, \ldots, \hat{\mu}_k$ w.r.t. $\bar{y}_1, \ldots, \bar{y}_k$ is given by ARE $(\hat{\mu},\bar{y}) = |\text{var-cov}(\bar{y})| / |V(\hat{\mu})|$, where $|.|$ denotes the determinant. Since $|\text{var-cov}(\bar{y})| = c^{2k}\prod_j \mu_j^2/n_j$, and it can be shown that

$$|V(\hat{\mu})| = (c^{2k}/(2c^2 + 1)^{k-1}]\prod_j \mu_j^2/n_j,$$

then ARE $(\hat{\mu},\bar{y}) = (2c^2 + 1)^{k-1}$. Thus, $\hat{\mu}_1, \ldots, \hat{\mu}_k$ is asymptotically more efficient than $\bar{y}_1, \ldots, \bar{y}_k$. The efficiency increases sharply with c. The A.R.E. of a single estimate $\hat{\mu}_j$ w.r.t. $\bar{y}_j$ is ARE$(\hat{\mu}_j,\bar{y}_j) = \text{var}(\bar{y}_j)/\text{var}(\hat{\mu}_j) = (2c^2 + 1)/\{(2c^2n_j/n) + 1\}$.

This is also dependent on the relative sample size of each group. The larger the relative size the more efficient $\hat{\mu}_j$ will be.

MONTE CARLO EVALUATIONS AND EXAMPLE

Sufficiently large ranges of value for $\mu_1, \ldots, \mu_k$, and c are simulated: .20–10,000 for μ_s and .01–1.0 for c. We present in Table 1 three of such simulations. 'True' denotes the value of the parameter used in generating the data set, 'sd', the estimate of sample standard deviation, 'asd' the estimate of the asymptotic standard deviation, and # iter, the number of iterations required for convergence. For example, for a random sample of size 75 taken from a population having a mean $\mu_1 = 0.20$, the maximum likelihood solution gives an estimate of 0.202 for μ_1, with an asymptotic standard deviation of 0.010. The ordinary mean estimate is 0.197 wiht a standard deviation of 0.010. In general, it is observed that 'asd' is smaller than 'sd'.

For illustration of this method we work out an example using data on "Absorbance values of three substances in a chemical assay for leucine amino peptidase" in Azen and Reed (1973), with 19 runs of test each. The following result is obtained (no. of groups = 3) :

$\hat{c}$ = .0351 , (# iter = 1)

$\hat{\mu}_1$ = 120.2101 , $\hat{\mu}_2$ = 70.4245 , $\hat{\mu}_3$ = 69.7337.

The asymptotic variance-covariance matrix is

.93518			
.00045	.32097		
.00045	.00026	.31470	
−.00009	−.00005	−.00005	.00001

That is, the 'asd' of the estimates of μ_1, μ_2, and μ_3 are (with the corresponding 'sd' in brackets) 0.967(0.993), 0.567(0.594), and 0.561 (0.566) respectively.

Table 1
Some results from Monte Carlo evaluations

Parameters		μ_1	μ_2	μ_3	μ_4	μ_5	μ_5	μ_6	c
1	true	.20	.30	.35	.40	.60	.80	.85	.40
n_j=75	$\bar{y}_j$±(sd)	.197±.010	.305±.013	.348±.016	.364±.018	.616±.028	.789±.035	.830±.04	–
j=1, ..., 7	$\hat{\theta}_j$±(asd)	.202±.010	.298±.012	.347±.014	.369±.015	.610±.025	.779±.032	.840±.035	.400±.014
# iter=9									
2	true	12.00	15.00	17.00	18.00	22.00	21.00	25.00	.10
n_j=130	$\bar{y}$±(sd)	11.92±.11	15.05±.13	16.86±.14	17.91±.16	22.38±.19	20.77±.18	24.83±.21	–
j=1, ..., 7	$\hat{\theta}_j$±(asd)	11.93±.10	15.04±.13	16.86±.14	17.92±.15	22.38±.19	20.76±.18	24.83±.21	.099±.002
# iter=3									
3	true	1122	1155	1270	1890	2222	2190	2590	.20
n_j=130	$\bar{y}$±(sd)	1107±20	1162±20	1249±22	1872±33	2300±39	2142±37	2555±45	–
j=1, ..., 7	$\hat{\theta}_j$±(asd)	1112±19	1160±19	1149±21	1874±31	2296±39	2139±36	2555±43	.198±.005
# iter=4									

CONCLUSION

A numerically exact method of solution to a realistic but seemingly complicated model is made possible. The alternative model that assumes a constant coefficient of variation without further assumption that the coefficient is known can now be handled efficiently. The estimators for the means in this model are shown to be asymptotically more efficient than the ordinary means, jointly as well as individually. In both cases the A.R.E. increases with the increase in the coefficient of variation. For the individual estimator the A.R.E. also increases with the relative sample size of the corresponding group.

REFERENCES

ABD-RAHMAN, M.N. (1978): Maximum likelihood estimation in normal models whose error variances are some functions of their means Ph.D. Thesis. Institute of Statistics. Mimeograph Series. N.C. State University.

ABRAMOWITZ, MILTON and STEGUN, IRENE A. (1965): *Handbook of Mathematical Functions.* New York Dover.

AMEMIYA, TAKESHI (1973): Regression analysis when the variance of the dependent variable is proportional to the square of its expectation. *J. Am. Statist. Assoc.* **68**: 928-934.

AZEN, S.P. and REED, A.H. (1973): Maximum likelihood estimation of correlation between variates having equal coefficients of variation. *Technometrics.* **15(3)**: 457-462.

BRADLEY, R.A. and GART, J.J. (1962): The asymptotic properties of ML estimators when sampling from associated populations. *Biometrika.* **49**: 205-214.

KELLY, LOUIS G. (1967): *Handbook of Numerical Methods and Applications.* California. Addison – Wesley.

LOHRDING, R.K. (1969): A test of equality of two normal population means assuming homogeneous coefficient of variation. *Ann. Math. statist.* **40(4)**: 1374-1385.

NOBLE, BEN (1969): *Applied Linear Algebra.* New Jersey: Prentice-Hall.

SEN, A.R. and GERIG, T.M. (1975): Estimation of population means having equal coefficient of variation on successive occasions. Contributed paper. *40th session of the Int. Statist. Inst.* Warsaw.

WILKS, S.S. (1963): *Mathematical Statistics.* New York. John. Wiley.

ZEIGLER, R.K. (1973): Estimators of coefficient of variation using k-samples. *Technometrics,* **15(2)**: 409-413.

(Received 17 September 1982)