Communications in Statistics - Theory and Methods

Further analysts of the data by akaike' s information criterion and the finite corrections

Nariaki Sugiura [a]

[a] Department of Mathematics, University of Tsukuba, Sakuramura, Ibaraki, JAPAN

PLEASE SCROLL DOWN FOR ARTICLE

FURTHER ANALYSIS OF THE DATA BY AKAIKE'S

INFORMATION CRITERION AND THE FINITE CORRECTIONS

Nariaki Sugiura

Department of Mathematics, University of Tsukuba
Sakuramura Ibaraki 300-31, JAPAN

ABSTRACT

Using Akaike's information criterion, three examples of
statistical data are reanalyzed and show reasonably definite
conclusions. One is concerned with the multiple comparison
problem for the means in normal populations. The second is concerned
with the grouping of the categories in a contingency table. The
third is concerned with the multiple comparison problem for the
analysis of variance by the logit model in contingency tables.
Finite correction of Akaike's information criterion is also proposed.

1. INTRODUCTION

Let $L(\theta)$ be the likelihood based on a random sample of size
n having probability density function $f(x|\theta)$, where the unknown
parameter θ lies in the union of open sets $\Theta = \theta^{(1)} \cup \ldots \cup \theta^{(k)}$
in Euclidean spaces of different dimensions, that is, $\theta^{(i)}$ is an
open set in p_i dimensional Euclidean space $R^{p}i$. We can say that the

13

model $\Theta^{(i)}$ is true if $\theta \varepsilon \Theta^{(i)}$. Let $\hat{\theta}_n^{(i)}$ be the maximum likelihood estimate of θ under the model $\Theta^{(i)}$ and assume that $\hat{\theta}_n^{(i)}$ is the best asymptotically normal estimate of θ under $\Theta^{(i)}$.

Akaike (1973, 1974, 1975, 1976) proposes to use the following information criterion,

$$AIC = -2 \log L(\hat{\theta}_n^{(i)}) + 2p_i , \qquad (1.1)$$

as a measure of goodness of fit for the model $\Theta^{(i)}$. The smaller is the AIC the better is the goodness of fit. This is similar to the usual goodness of fit test by χ^2-distribution. If $\theta \varepsilon \Theta^{(i)}$, then $-2 \log L(\hat{\theta}_n^{(i)})+2 \log L(\hat{\theta}_n)$ is the likelihood ratio statistic having asymptotically χ^2-distribution with $p-p_i$ degrees of freedom, where $\hat{\theta}_n$ is the maximum likelihood estimates of θ under $\theta \varepsilon \Theta \subseteq R^p$. The term of $2p_i$ in AIC is obtained by Akaike, as a correction for the asymptotic bias of the estimate $n^{-1}\log L(\hat{\theta}_n^{(i)})$ for

$$I_i = E_{\hat{\theta}_n^{(i)}}[\int f(y|\theta_0)\log f(y|\hat{\theta}_n^{(i)}) dy] \qquad (1.2)$$

when $\theta_0 \varepsilon \Theta^{(i)}$. Here $f(y|\hat{\theta}_n^{(i)})$ is the predicted probability density for future observation y when $\theta \varepsilon \Theta^{(i)}$ and y is assumed to be independent of $\hat{\theta}_n^{(i)}$. Akaike's idea is to select the model $\Theta^{(i)}$ when $I_i = \max\{I_1,\ldots,I_k\}$, which is equivalent to selecting the model such that the expected value of Kullback's discrimination information

$$E_{\hat{\theta}_n^{(i)}} [\int f(y|\theta_0)\log\{f(y|\theta_0)/f(y|\hat{\theta}_n^{(i)})\}dy] \qquad (1.3)$$

is minimized. The quantity I_i in (1.2) cannot be computed from the data so that the maximization of the asymptotically unbiased estimate $n^{-1}\log L(\hat{\theta}_n^{(i)})-p_i/n$ is recommended.

Three examples are given to show how the model is selected by AIC.

2. ANALYSIS OF THE DATA

2.1. The Data with Different Means

The following Table I is a reproduction from Snedecor and Cochran (1967, p.259). For this, equality of four means, namely

TABLE I

Mean Grams of Fat (minus 100) Absorbed per Batch of
Doughnuts. Each sample size = 6.

Fat	1	2	3	4
$\bar{X}$±s.d.	72±13.3	85±7.8	76±9.9	62±8.2

$\mu_1 = \mu_2 = \mu_3 = \mu_4$, with common unknown variance in normal populations
is rejected with 5% level of significance. Then several methods
are mentioned to distinguish the true differences between four
means, namely, LSD (Least Significant Difference) method implies
$\mu_2 \neq \mu_1$, $\mu_2 \neq \mu_4$ and $\mu_3 \neq \mu_4$, Q method implies only $\mu_2 \neq \mu_4$, and so on
(Senedecor and Cochran, 1967, p.272).

However it is concluded on page 275 of this book that no method
is uniformly best. We cannot make any conclusions with exact
confidence. Graphical presentation of Table I is given in FIG. 1
from which we can easily imagine equalities and inequalities between
four means.

As Akaike notes, the problem of detecting the differences
between four means is not a problem of hypothesis testing but a
multiple decision problem. Akaike's information criterion for this
problem is given by

$$AIC = 2n \log(\hat{\sigma}\sqrt{2}\pi) + n + 2(\text{number of estimated parameters}),$$

where n=total sample sizes=24 and $\hat{\sigma}^2$ is the maximum likelihood esti-
mate of common variance under the model considered. By this AIC we

FIG.1
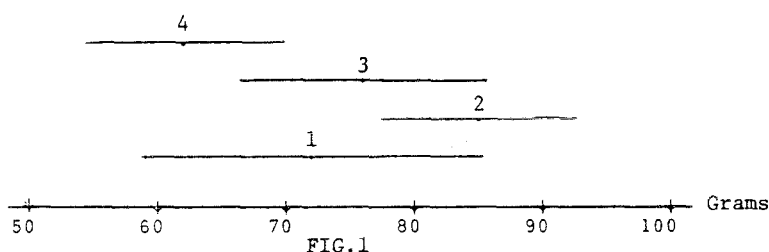
can reach definite conclusions, as shown in Table II. The best
model giving minimum AIC is $\mu_1=\mu_3\neq\mu_2\neq\mu_4$ (AIC=183.0). The second
best model is $\mu_1\neq\mu_2\neq\mu_3\neq\mu_4$ (AIC=184.5). The worst model is $\mu_2=\mu_4\neq\mu_1$
$\neq\mu_3$ (AIC=196.4). The last columns in Table II give the corrected
AIC discussed in the next section, from which we can see that the
second best model is not one but two; namely, the models $\mu_1=\mu_2=\mu_3$
$\neq\mu_4$ and $\mu_2=\mu_3\neq\mu_1\neq\mu_4$ are second best. From FIG. 1 these statements
seem to be very much reasonable.
The p columns in Table II stand for the number of estimated
parameters in each model.

2.2 Grouping of the Categories in a Contingency Table.

Table III is a reproduction from Sugiura and Otake (1973),
from which we wish to decide an appropriate regrouping of the seven
categories. One method proposed originally by Otake is based on
the χ^2 statistics, and approximate critical points are computed in
Sugiura and Otake(1973), which implies that for this data the first
two groups from the low dose should be combined, the third through
the fifth groups should be combined (dose 20∿199), and, finally,
the sixth and seventh groups should be combined so that a 2x3
table is obtained.

FIG.2 is a graphical presentation of Table III to facilitate
the appropriate regrouping of the categories.

We can formulate this problem in the following way. Since
each observed number X_i (i=1∿7) of leukemia deaths has independent

TABLE II

p	Model	AIC	c-AIC	p	Model	AIC	c-AIC
2	$\mu_1=\mu_2=\mu_3=\mu_4$	192.7	193.3	4	$\mu_1=\mu_2\neq\mu_3\neq\mu_4$	187.8	190.0
3	$\mu_1=\mu_2=\mu_3\neq\mu_4$	186.1	187.3	4	$\mu_1=\mu_3\neq\mu_2\neq\mu_4$	183.0	185.1
3	$\mu_1=\mu_2=\mu_4\neq\mu_3$	194.5	195.7	4	$\mu_1=\mu_4\neq\mu_2\neq\mu_3$	185.8	187.9
3	$\mu_1=\mu_3=\mu_4\neq\mu_2$	186.9	188.1	4	$\mu_2=\mu_3\neq\mu_1\neq\mu_4$	185.2	187.3
3	$\mu_2=\mu_3=\mu_4\neq\mu_1$	194.6	195.8	4	$\mu_2=\mu_4\neq\mu_1\neq\mu_3$	196.4	198.5
3	$\mu_1=\mu_2\neq\mu_3=\mu_4$	190.9	192.1	4	$\mu_3=\mu_4\neq\mu_1\neq\mu_2$	188.6	190.7
3	$\mu_1=\mu_3\neq\mu_2=\mu_4$	194.7	195.9	4	$\mu_1\neq\mu_2\neq\mu_3\neq\mu_4$	184.5	187.8
3	$\mu_1=\mu_4\neq\mu_2=\mu_3$	186.2	187.4				

TABLE III

Dose(Rad)	<5	5~	20~	50~	100~	200~	300+	Total
Rate Of	2	0	3	2	2	2	5	16
Leukemia	4603	1161	480	273	245	100	154	7016
		1	2	3	4	5	6	

binomial distribution $b(n_i, p_i)$, regrouping of the categories means
the homogeneity of p's (population rates) within those categories.
So that we should decide which p's are equal and which p's are un-
equal. This is again a multiple decision problem. Note that cate-
gories are ordered. Only the equality of the adjoining categories
should be examined. Table IV gives the AIC and the corrected AIC
for all possible combinations of the p equalities.

The first columns in Table IV indicate the models according to
the positions of the separation of the categories, corresponding to
the small arabic numerals$(_1, _2, _3 \ldots)$ shown in Table III and FIG. 2.
For instance, the symbol (1,2,5) in Table IV means that Dose <5 is
the first group; Dose 5~19 is the second group; Dose 20~199 is the
third group; and, finally, Dose 200+ is the fourth group; so that
we get a 2x4 contingency table. According to AIC or corrected AIC
in Table IV, this regrouping of the categories is the best among all
possible 2x4 tables. The minimum AIC or the minimum corrected AIC
is attained by regrouping (2,5) (underlined in Table IV) which is
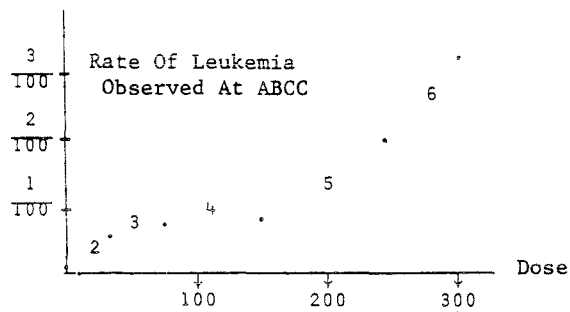the same conclusion as obtained in Sugiura and Otake (1973) by

FIG. 2

TABLE IV

	AIC	c-AIC		AIC	c-AIC		AIC	c-AIC
2x1			2x4			2x5		
	63.5	63.6	(1,2,3)	29.6	30.6	(1,2,3,4)	30.0	31.5
2x2			(1,2,4)	28.1	28.9	(1,2,3,5)	27.2	28.5
(1)	45.9	46.5	(1,2,5)	25.3	26.1	(1,2,3,6)	27.9	29.1
(2)	28.4	29.0	(1,2,6)	26.3	27.1	(1,2,4,5)	27.3	28.6
(3)	35.5	35.8	(1,3,4)	35.4	36.9	(1,2,4,6)	27.7	28.8
(4)	38.7	33.9	(1,3,5)	32.6	33.8	(1,2,5,6)	26.9	28.3
(5)	40.1	40.4	(1,3,6)	33.3	34.5	(1,3,4,5)	34.6	36.6
(6)	46.6	46.8	(1,4,5)	34.6	35.9	(1,3,4,6)	35.0	36.8
2x3			(1,4,6)	35.0	36.2	(1,3,5,6)	34.3	36.0
(1,2)	29.5	30.1	(1,5,6)	35.8	37.1	(1,4,5,6)	36.2	38.1
(1,3)	35.0	35.9	(2,3,4)	28.9	30.4	(2,3,4,5)	28.1	30.1
(1,4)	35.4	36.2	(2,3,5)	26.1	27.4	(2,3,4,6)	28.5	30.3
(1,5)	34.1	34.9	(2,3,6)	26.8	28.0	(2,3,5,6)	27.8	29.5
(1,6)	37.4	38.2	(2,4,5)	26.2	27.5	(2,4,5,6)	27.8	29.7
(2,3)	28.5	29.5	(2,4,6)	26.6	27.7	(3,4,5,6)	36.8	38.6
(2,4)	27.0	27.8	(2,5,6)	25.8	27.2	2x6		
(2,5)	24.2	25.0	(3,4,5)	35.1	36.5	(1,2,3,4,5)	29.2	31.2
(2,6)	25.2	26.0	(3,4,6)	35.5	36.7	(1,2,3,4,6)	29.6	31.4
(3,4)	35.9	36.7	(3,5,6)	34.8	35.9	(1,2,3,5,6)	28.9	30.6
(3,5)	33.1	33.7	(4,5,6)	39.5	40.8	(1,2,4,5,6)	28.9	30.8
(3.6)	33.8	34.4				(1,3,4,5,6)	36.2	38.8
(4,5)	37.9	38.7				(2,3,4,5,6)	29.8	32.3
(4,6)	38.3	38.9				2x7		
(5,6)	41.8	42.6				(1,2,3,4,5,6)	30.9	33.4

another method based on χ^2 statistics. From FIG. 2 we can see that this regrouping seems to be reasonable. The formula for the corrected AIC is given in the next section. Tables II and IV tell us that corrections always increase AIC, but less than 10%.

2.3. Analysis of Variance by Logit Model

Table V is a reproduction from Sugiura and Otake (1974), originally reported by Jablon and Kato (19__), showing the number of deaths from leukemia observed at Atomic Bomb Casuality Commission (now Radiation Effect Research Foundation).

TABLE V

Age	Not in City	0~	10~	Dose 50~	100~	200+
0~	$\dfrac{0}{3015}$	$\dfrac{7}{10739}$	$\dfrac{3}{2992}$	$\dfrac{1}{695}$	$\dfrac{4}{422}$	$\dfrac{11}{398}$
10~	$\dfrac{5}{5973}$	$\dfrac{4}{11815}$	$\dfrac{6}{2626}$	$\dfrac{1}{772}$	$\dfrac{3}{795}$	$\dfrac{6}{826}$
20~	$\dfrac{2}{5671}$	$\dfrac{8}{10836}$	$\dfrac{3}{2801}$	$\dfrac{1}{798}$	$\dfrac{3}{599}$	$\dfrac{9}{633}$
35~	$\dfrac{3}{6161}$	$\dfrac{19}{12664}$	$\dfrac{4}{3570}$	$\dfrac{2}{974}$	$\dfrac{1}{695}$	$\dfrac{10}{618}$
50+	$\dfrac{3}{3698}$	$\dfrac{7}{9060}$	$\dfrac{3}{2418}$	$\dfrac{2}{657}$	$\dfrac{2}{395}$	$\dfrac{6}{295}$

Let d_{ij} be the number of leukemia deaths observed at i-th age group and j-th dose group. Then d_{ij} has independent binomial distribution $b(n_{ij}, p_{ij})$. An analysis of this data was given by Sugiura and Otake (1974), assuming the logit model for p_{ij}, namely

$$\log \frac{p_{ij}}{1 - p_{ij}} = \mu + \alpha_i + \beta_j , \qquad (2.1)$$

where $\Sigma \, \alpha_i = 0$ and $\beta_1 = 0$.

The α parameters are called the age factor and the β's are called the dose factor. We have computed the maximum likelihood estimates for these parameters by Newton-Raphson iteration or the iterative scaling method (Sugiura, 1974), giving the asymptotic 95% simultaneous confidence intervals for dose factors as

If $\beta_1 = \beta_2 = \ldots = \beta_6 = 0$, we can say that, eliminating the age factor, radiation dose exposed has no effect for leukemia. However, this is not the case and with high significance, the above hypothesis was rejected. Then the problem is to decide which β's are positive. Simultaneous confidence intervals give some information about this, but it is not sufficient since we cannot get definite conclusions from Table VI. The parameters β_5 and β_6 are certainly positive but other β's are only possibly positive.

TABLE VI

Simultaneous Confidence Intervals for β

$-0.55 < \beta_2 < 1.55$ $-0.23 < \beta_3 < 2.17$ $-0.28 < \beta_4 < 2.85$

$0.92 < \beta_5 < 3.54$ $2.42 < \beta_6 < 4.54$

Graphical presentation of the maximum likelihood estimates for β are given in FIG. 3 to facilitate obtaining tentative solutions to this problem.

For each model in Table VII, we have computed the maximum likelihood estimates for β and then Akaike's information criterion by (1.1), where the likelihood $L(\theta)$ is given by

$$\pi_{ij} \binom{n_{ij}}{d_{ij}} p_{ij}^{d_{ij}} (1 - p_{ij})^{n_{ij} - d_{ij}} \qquad (2.2)$$

and some of the p's are equal, corresponding to the model considered. From Table VII we can conclude that only β_1 is equal to zero and the other β's are positive; namely, six radiation dose groups have different effects and should not be combined in any way. This implies that even the first group (Not in City) and the second have different effects, though they received almost the same radiation dose. It was said earlier that the first group seems to have some problem as a control.
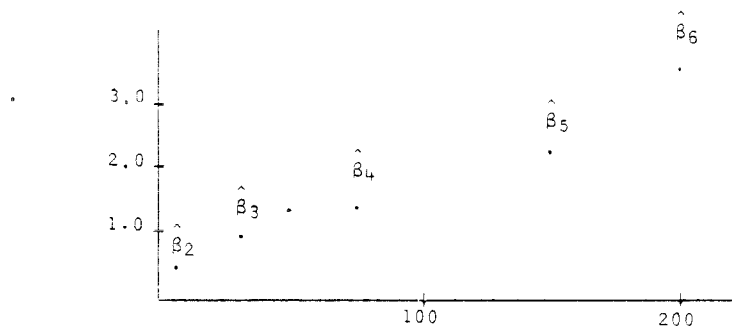

FIG. 3

TABLE VII

Models	AIC	AIC
$\hat{\beta}_1=0$	140.2	without
$\hat{\beta}_1=\hat{\beta}_2=0$	141.1	assuming logit
$\hat{\beta}_1=\hat{\beta}_2=\hat{\beta}_3=0$	143.6	$= 152.4$
$\hat{\beta}_1=\hat{\beta}_2=\hat{\beta}_3=\hat{\beta}_4=0$	144.9	

From FIG. 3 we can also see that our conclusions are natural. The advantage of assuming a logit model is also supported by Table VII, since we have smaller AIC than without assuming logit. However there remains the possibility of getting models with further reduced AIC.

3. FINITE CORRECTION OF AIC

Since Akaike's information criterion is based on eliminating the asymptotic bias of the maximum likelihood, namely $-2 \log L(\hat{\theta})$ for the ML estimate $\hat{\theta}$ of θ, we can refine it by considering the exact bias for the following problems of practical importance. In dealing with the multisample cases, the quantity I in (1.2) is slightly modified so that we have multisample future observations of exactly the same sample sizes as the original observations.

3.1. Normal Populations with Different Means.

Let $X_{i1},\ldots,X_{in_i}$ be a random sample from normal distribution $N(\mu_i,\sigma^2)$ for $i=1,\ldots,k$. The hypotheses of interest are $H_{j_1\cdots j_c} : \mu_{j_1} =\ldots=\mu_{j_c}$ with unknown σ^2 for all possible choices of c and $j_1,\ldots,j_c$ such that $1\leq c\leq k$ and $1\leq j_1<\ldots<j_c\leq k$. The parameters are denoted by $\theta=(\mu_1,\ldots,\mu_k,\sigma^2)$. Put $n=\Sigma_{i=1}^k n_i$. Then the likelihood is given by

$$\log L(\theta) = -n \log(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\Sigma_{i=1}^k\Sigma_{j=1}^{n_i} (x_{ij}-\mu_i)^2. \quad (3.1)$$

By Akaike (1975) or from Section 1, the maximum likelihood $\log L(\hat{\theta})$ may be regarded as an estimate of

$$J = E_{\hat{\theta}_n} [\int f(y|\theta_0)\log f(y|\hat{\theta}_n) \, dy], \quad (3.2)$$

where $f(y|\theta_0)$ is the probability density of the future observation
$Y=(Y_{11},\ldots,Y_{1n_1},\ldots,Y_{kn_k})$ of the same size as the X's, and it is
assumed to be independent of the X's. The expectation is taken
under the distribution of X when $H_{j_1\cdots j_c}$ is true, namely the true
value $\theta_0 \varepsilon H_{j_1\cdots j_c}$. Note that under $H_{j_1\cdots j_c}$, $n\hat{\sigma}^2/\sigma^2$ has χ^2-distribu-
tion with n-k+c-1 degrees of freedom, and we get

$$J = -n\ E[\log(\hat{\sigma}\sqrt{2\pi})] - \frac{\sigma^2}{2}E[\frac{1}{\sigma^2}](n+k-c+1)$$

$$= -n\ E[\log(\hat{\sigma}\sqrt{2\pi})] - \frac{n(n+k-c+1)}{2(n-k+c-3)} = E[\log\ L(\hat{\theta})] - \frac{n(k-c+2)}{n-k+c-3}\ . \tag{3.3}$$

The second term in the last equation gives the exact bias of
$\log L(\hat{\theta})$. Multiplying by -2, we get the corrected AIC as

$$c\text{-AIC} = -2\ \log\ L(\hat{\theta}) + \frac{2n(k-c+2)}{n-k+c-3}\ , \tag{3.4}$$

which coincides with AIC as n tends to infinity. The corrected
AIC in Table II was computed by this formula.

3.2. Regression Models.

Suppose we have a nx1 vector X having normal distribution
$N(A\Theta,\sigma^2 I)$, where A(nxk) is a known nxk matrix of rank k and
Θ(kx1) is the unknown parameter. The hypothesis of interest
is H: $B\Theta=0$ for possible known matrices B(bxk) of rank b.
Considering the canonical form we may assume that $X=(X_1,\ldots,X_n)$ has
$N(*,\sigma^2 I)$ where $E[X_i]=\theta_i$ for i=1,...,k and $E[X_i]=0$ for i=k+1,...,n.
The hypothesis H is equivalent to saying that $\theta_1=\ldots=\theta_b=0$. By the
same argument as in Section 3.1 we get

$$c\text{-AIC} = -2\ \log\ L(\hat{\theta}) + \frac{2n(k-b+1)}{n-k+b-2}\ . \tag{3.5}$$

This is an extension of Section 3.1. The problems for the linear
hypothesis models are included in this formula. For example, if
we consider two-way classification models without interaction such
that X_{ijk} has $N(\mu+\alpha_i+\beta_j,\sigma^2)$ for $1\leq i\leq A$, $1\leq j\leq B$, $1\leq k\leq C$ and the hypothesis

of interest is $H: \sigma_{j_1} = \ldots = \sigma_{j_a}$ (σ^2:unknown), then n=ABC, k=A+B-1, b=a-1 and the formula (3.5) works. This is an extension of (3.4). We can further generalize (3.5) to the following multivariate case.

Let each row of X(nxp) have independent p-variate normal distribution $N(\mu,\Sigma)$, where E(X)=Aθ for a known matrix A(nxk) of rank k, and the hypothesis of interest is Bθ=0 where a known matrix B(bxk) is of rank b. Again this can be transformed to the canonical form similar to the univariate case, giving

$$c\text{-AIC} = -2 \log L(\hat{\theta}) + \frac{2n(k-b+(p+1)/2)p}{n-k+b-p-1} . \qquad (3.6)$$

Let n tend to infinity in (3.6). We get the bias of $-2 \log L(\hat{\theta})$ as twice the number of estimated parameters. When p=1 the formula (3.6) is equal to (3.5). We can see that correction of AIC always increases the value.

3.3 Normal Populations with Different Variances.

Suppose that $X_{i1},\ldots,X_{in_i}$ is a random sample from $N(\mu_i,\sigma_i^2)$ for i=1,...,k and that the hypothesis of interest is $H: \sigma_{j_1}^2 = \ldots \sigma_{j_c}^2$ with unknown μ_i. From (3.2) we get

$$J = -\Sigma_{i=1}^{k} E[\log(\hat{\sigma}_{j_i} \sqrt{2\pi}) + \{(n_{j_i}+1)\sigma_{j_i}^2 / (2\hat{\sigma}_{j_i}^2)\}]$$

$$= E[\log L(\hat{\theta})] + \frac{n}{2} - \frac{1}{2} \sum_{i=c+1}^{k} \{n_{j_i}(n_{j_i}+1)/(n_{j_i}-3)\}$$

$$- \frac{1}{2}(\Sigma_{i=1}^{c} n_{j_i})(\sum_{i=1}^{c} n_{j_i} +c)/(\sum_{i=1}^{c} n_{j_i} -c-2), \quad (3.7)$$

where we have put $n_{j_{c+1}},\ldots,n_{j_k}$ such that $\{n_{j_1},\ldots,n_{j_c},\ldots,n_{j_k}\}$ = $\{n_1,\ldots,n_k\}$. From (3.7) we get the corrected AIC as

$$c\text{-AIC} = -2 \log L(\hat{\theta}) + 2[(c+1) \sum_{i=1}^{c} n_{j_i} /(\sum_{i=1}^{c} n_{j_i} -c-2)$$

$$+\{2 \sum_{i=c+1}^{k} n_{j_i} /(n_{j_i} -3)\}]. \qquad (3.8)$$

This can be generalized to a multivariate case where we have a random sample from the p-variate normal distribution $N(\mu_i,\Sigma_i)$, and

the equality of covariance matrices $H: \Sigma_{j_1} = \ldots = \Sigma_{j_c}$ with unknown μ_i is of interest. The result is given by

$$c\text{-AIC} = -2 \log L(\hat{\theta}) + 2p[(c + \tfrac{p+1}{2}) \sum_{i=1}^{c} n_{j_i} / (\sum_{i=1}^{c} n_{j_i} - c - p - 1)$$

$$+ \tfrac{p+3}{2} \sum_{i=c+1}^{k} n_{j_i} / (n_{j_i} - p - 2)]. \qquad (3.9)$$

If all n_i tend to infinity in (3.9), the bias of $-2 \log L(\hat{\theta})$ tends to $2p[\{c + (p+1)/2\} + \{(p+3)(k-c)/2\}]$, which is equal to twice the number of estimated parameters.

3.4. Heterogeneity of Binomial Populations

Let X_i have binomial distribution $b(n_i, p_i)$ independently for $i = 1, \ldots, k$. The hypothesis of interest is $H: p_1 = \ldots = p_{a_1} \equiv \theta_1, p_{a_1+1} = \ldots = p_{a_2} \equiv \theta_2, \ldots, p_{a_{r-1}+1} = \ldots = p_{a_r} \equiv \theta_r$, where all θ's are unknown. For convenience we shall put $a_0 = 0$. Then we get

$$E[\log L(\hat{\theta})] - J = \sum_{i=1}^{k} E[(X_i - n_i p_i) \log(\hat{p}_i / (1 - \hat{p}_i))]$$

$$= \sum_{u=0}^{r-1} E[(\sum_{i=a_u+1}^{a_{u+1}} X_i - \theta_u \sum_{i=a_u+1}^{a_{u+1}} n_i) \log(\frac{\hat{\theta}_{u+1}}{1 - \hat{\theta}_{u+1}})] \qquad (3.10)$$

where $\hat{\theta}_{u+1} = \sum_{i=a_u+1}^{a_{u+1}} X_i / \sum_{i=a_u+1}^{a_{u+1}} n_i$. If X has binomial distribution $b(n, p)$, then it is well known that

$$E[\{(X - np)/\sqrt{npq}\}^3] = (1 - 2p)/\sqrt{npq},$$

$$E[\{(X - np)/\sqrt{npq}\}^4] = 3 - 6/n + 1/(npq) \qquad (3.11)$$

for $q = 1 - p$, which yields

$$E[(X - np) \log \frac{X/n}{1 - X/n}] = 1 - 1/n + (1/2npq) + O(n^{-3/2}). \qquad (3.12)$$

The above equality is not true in the strict sense. However by suitably modifying $\log X/n$ when $X = 0$ and $\log(1 - \frac{X}{n})$ when $X = n$, this

is true. Hence we can eliminate the asymptotic bias of the maximum likelihood up to the order n^{-1} as

$$c\text{-AIC} = -2 \log L(\hat{\theta}) + 2r + \sum_{u=0}^{r-1} (1/ \sum_{i=a_u+1}^{a_{u+1}} n_i)$$

$$\times [-2+\{1/\theta_{u+1}(1-\theta_{u+1})\}]. \qquad (3.13)$$

The corrected AIC in Table IV was computed by (3.13), in which the maximum likelihood estimate $\hat{\theta}$ was used instead of θ. If $\hat{\theta}_u=0$ for some u, the last term in (3.13), namely $[-2+\{1/\theta_u(1-\theta_u)\}]$, was understood to be zero. Again, in this case the correction of AIC increases the value.

BIBLIOGRAPHY

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *2nd Int. Symp. Inf. Theory* (B.N.Petrov and F.Csaki, Eds.). Budapest: Akademiai Kiado, 267-81.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automatic Control AC-19*, 716-23.

Akaike, H. (1975). Concept of information and the theory of statistics. Invited paper presented to meeting of Math. Soc. Japan, October.

Akaike, H. (1976). What is an information criterion AIC? *Math. Sci. 14* (3), in Japanese, Diamond Co.(Tokyo), 5-11.

Jablon, S. and Kato, H. (1971). Mortality among A-bomb survivors, 1950-1970. Japanese National Institute of Health-Atomic Bomb Casuality Commission Life Span Study Report 6 (TR 10-71), ABCC, Hiroshima, Japan.

Snedecor, G.W. and Cochran, W.G. (1967). *Statistical Methods*. Ames, IA: Iowa State University Press.

Sugiura, N. and Otake, M. (1973). Approximate distribution of the maximum of c-1 χ^2-statistics (2x2) derived from 2xc contingency table. *Commun. Statist. 1* (1), 9-16.

Sugiura, N. and Otake, M. (1974). An extension of the Mantel-
    Haenszel procedure to k 2xc contingency tables and the relation
    to the logit model. *Commun. Statist.* 3 (9), 829-42.

Sugiura, N. (1974). Maximum likelihood estimates for the logit
    model and the iterative scaling method. *Commun. Statist.* 3 (10),
    985-93.