Ann. Inst. Statist. Math.
38 (1986), Part A, 385-398

A PROJECTION METHOD OF ESTIMATION FOR A SUBFAMILY OF EXPONENTIAL FAMILIES

SHINTO EGUCHI

(Received Mar. 11, 1985; revised Dec. 6, 1985)

Summary

This paper is concerned with estimation for a subfamily of exponential-type, which is a parametric model with sufficient statistics. The family is associated with a surface in the domain of a sufficient statistic. A new estimator, termed a projection estimator, is introduced. The key idea of its derivation is to look for a one-to-one transformation of the sufficient statistic so that the subfamily can be associated with a flat subset in the transformed domain. The estimator is defined by the orthogonal projection of the transformed statistic onto the flat surface. Here the orthogonality is introduced by the inverse of the estimated variance matrix of the statistic on the analogy of Mahalanobis's notion (1936, *Proc. Nat. Inst. Sci. Ind.*, 2, 49-55). Thus the projection estimator has an explicit representation with no iterations. On the other hand, the MLE and classical estimators have to be sought as numerical solutions by some algorithm with a choice of an initial value and a stopping rule. It is shown that the projection estimator is first-order efficient. The second-order property is also discussed. Some examples are presented to show the utility of the estimator.

1. Introduction and results

First let us look at a linear regression model

$$y=X\beta+e$$

with a design matrix X of size $n\times m$ and a parameter vector β of m-component. Here the error term e is supposed to have the n-variate normal distribution with mean 0 and covariance Σ_0. The statistical estimation for the model enjoys a simple and intuitive interpretation. The model locus

Key words and phrases: ABO blood group model, exponential family, inbreeding coefficient, maximum likelihood estimator, projection estimator.

$$\mathcal{M}X=\{X\boldsymbol{\beta}:\boldsymbol{\beta}\in\boldsymbol{R}^m\}$$

is flat in the data space $\boldsymbol{R}^n$. The MLE, or the Gauss-Markov estimator, of $\boldsymbol{\beta}$ has a closed form

$$\hat{\boldsymbol{\beta}}=(X'\Sigma_0^{-1}X)^{-1}X'\Sigma_0^{-1}\boldsymbol{y}\ ,$$

which is the orthogonal projection of $\boldsymbol{y}$ onto $\mathcal{M}X$ with respect to $\langle\ ,\ \rangle$. Here $\langle\ ,\ \rangle$ is an inner product on $\boldsymbol{R}^n$ defined by $\langle\boldsymbol{y}_1,\boldsymbol{y}_2\rangle=\boldsymbol{y}_1'\Sigma_0^{-1}\boldsymbol{y}_2$. At the same time, the MLE is a minimizer of the squared Mahalanobis distance

$$D^2=\|\boldsymbol{y}-X\boldsymbol{\beta}\|^2$$

with the norm $\|\ \|$ defined by $\langle\ ,\ \rangle$. The Pythagorean identity

$$\|\boldsymbol{y}-X\boldsymbol{\beta}\|^2=\|\boldsymbol{y}-X\hat{\boldsymbol{\beta}}\|^2+\|X\hat{\boldsymbol{\beta}}-X\boldsymbol{\beta}\|^2$$

holds. Thus the problem is reduced to the Euclidean geometry associated with the metric $\langle\ ,\ \rangle$. There are a number of books from this geometric viewpoint, see e.g. Draper and Smith [5] and Takeuchi, Yanai and Mukherjee [12].

Secondly we consider the ABO blood group model. Let a, b and o be the frequency parameters of genes A, B and O respectively, so that $a+b+o=1$. The expected probabilities of the four phenotypes in random mating are given as follows:

(1.1)

Phenotype	Expected probability
A	$\pi_1=\pi_1(a,b)=a^2+2ao$
B	$\pi_2=\pi_2(a,b)=b^2+2bo$
O	$\pi_3=\pi_3(a,b)=o^2$
A B	$\pi_4=\pi_4(a,b)=2ab$

with $o=1-a-b$, see § 5g in Rao [11]. For observed probabilities $\bar{p}_1$, $\bar{p}_2$, $\bar{p}_3$ and $\bar{p}_4$ with sample size n, the likelihood is given by

$$\frac{n!}{(n\bar{p}_1)!(n\bar{p}_2)!(n\bar{p}_3)!(n\bar{p}_4)!}\pi_1^{n\bar{p}_1}\pi_2^{n\bar{p}_2}\pi_3^{n\bar{p}_3}\pi_4^{n\bar{p}_4}\ .$$

This lead to the likelihood equation system

$$(1.2)\qquad \frac{\bar{p}_1}{\pi_1}o-\frac{\bar{p}_2}{\pi_2}b-\frac{\bar{p}_3}{\pi_3}o+\frac{\bar{p}_4}{\pi_4}b=0\ ,\qquad -\frac{\bar{p}_1}{\pi_1}a+\frac{\bar{p}_2}{\pi_2}o-\frac{\bar{p}_3}{\pi_3}o+\frac{\bar{p}_4}{\pi_4}a=0\ ,$$

which forms an algebraic surface of degree 5. The explicit representation for the MLE is not known. So we have to use some algorithm so as to obtain the value of the MLE. It is troublesome to look for an initial value and build a stopping rule of iteration. Thus it is seen

that the simple structure of the MLE under the linear regression model is spoiled for the ABO blood group model. This aspect may be caused by the reason why the model locus

$$\mathcal{M}=\{(\pi_i(a, b))_{i=1,\cdots,4}: (a, b) \in S_2\}$$

is non-flat in the space of four cell probabilities, where S_2 denotes the open simplex of dimension 2, see Fig. 1.

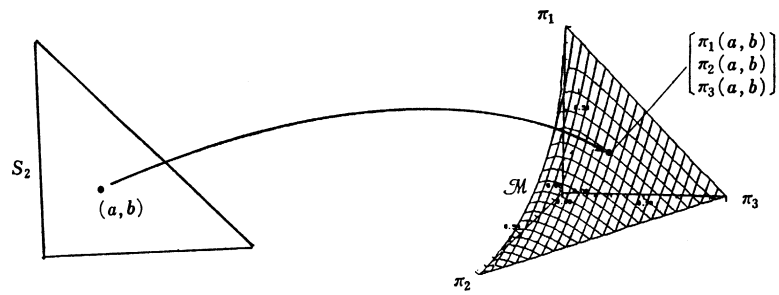


Fig. 1. The model locus of ABO blood group model from the space of gene frequency parameters. Here we look at the locus in the 3-dimensional simplex in place of the set $\{(\pi_1, \pi_2, \pi_3, \pi_4): \pi_1+\pi_2+\pi_3+\pi_4=1,\ \pi_i \geqq 0,\ i=1,\cdots,4\}$.

The likelihood equations (1.2) are expressed as

$$X_{a,b}V_{a,b}\{\boldsymbol{\pi}(a, b)-\bar{\boldsymbol{p}}\}=0\ ,$$

where $V_{a,b}=\mathrm{diag}\,(1/\pi_1(a, b), 1/\pi_2(a, b), 1/\pi_3(a, b), 1/\pi_4(a, b))$ and

$$X_{a,b}=\begin{pmatrix}\dfrac{\partial}{\partial a}\boldsymbol{\pi}(a, b)\\ \dfrac{\partial}{\partial b}\boldsymbol{\pi}(a, b)\end{pmatrix}=2\begin{pmatrix}o, & -b, & -o, & b\\ -a, & o, & -o, & a\end{pmatrix}$$

with $\boldsymbol{\pi}(a, b)=(\pi_i(a, b))_{i=1,\cdots,4}$ and $\boldsymbol{p}=(\bar{p}_i)_{i=1,\cdots,4}$. Since the tangent vectors $(\partial/\partial a)\boldsymbol{\pi}(a, b)$ and $(\partial/\partial b)\boldsymbol{\pi}(a, b)$ are parameter-dependent together with the weight matrix $V_{a,b}$, the MLE cannot be obtained by the projection onto the known flat subset.

In order to reform the difficulty with the MLE, we introduce a new estimator, termed a projection estimator. We state the definition of the estimator in more general situation. Let $\mathcal{F}$ be an n-dimensional exponential family of densities with respect to a sigma-finite measure μ on a data space $\boldsymbol{R}^n$, which is

$$\mathcal{F}=\{f_{\boldsymbol{\theta}}(\boldsymbol{x})=\exp\,(\boldsymbol{x}'\boldsymbol{\theta}-\psi(\boldsymbol{\theta})):\ \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$$

with a natural parameter vector $\boldsymbol{\theta}$, where

$$\Theta=\left\{\boldsymbol{\theta}\in\boldsymbol{R}^n:\ \int\exp(\boldsymbol{x}'\boldsymbol{\theta})d\mu(\boldsymbol{x})<\infty\right\}\ .$$

The family $\mathcal{F}$ is often expressed by the expectation parameter vector $\boldsymbol{\pi}$ with the transformation

$$\boldsymbol{\pi}=\int\boldsymbol{x}f_{\boldsymbol{\theta}}(\boldsymbol{x})d\mu(\boldsymbol{x})\ .$$

It is known that the relation $\boldsymbol{\pi}=(\partial/\partial\boldsymbol{\theta})\psi(\boldsymbol{\theta})$ holds if Θ is an open subset of $\boldsymbol{R}^n$, which is assumed hereafter. Throughout this paper, we focus on a situation with more knowledge of experiments, so that the family of our interest is reduced to a subfamily of lower dimension m:

$$\tilde{\mathcal{F}}=\{f_{\boldsymbol{\theta}(\boldsymbol{u})}(\boldsymbol{x})\in\mathcal{F}:\ \boldsymbol{u}\in U\}$$

with an open subset U of $\boldsymbol{R}^m$. Here the mapping $\boldsymbol{\theta}(\cdot)$ of U into Θ is supposed to be non-singular, i.e., the Jacobi matrix of $\boldsymbol{\theta}(\cdot)$ is of rank m on U. From this assumption the family $\tilde{\mathcal{F}}$ is associated with a smooth surface $\{\boldsymbol{\theta}(\boldsymbol{u}):\ \boldsymbol{u}\in U\}$ of dimension m in Θ. Similarly $\tilde{\mathcal{F}}$ has a locus $\{\boldsymbol{\pi}(\boldsymbol{u}):\ \boldsymbol{u}\in U\}$ in the expectation parameter space Π, where $\boldsymbol{\pi}(\boldsymbol{u})=(\partial/\partial\boldsymbol{\theta})\psi[\boldsymbol{\theta}(\boldsymbol{u})]$ and $\Pi=\{(\partial/\partial\boldsymbol{\theta})\psi(\boldsymbol{\theta}):\ \boldsymbol{\theta}\in\Theta\}$.

Let $(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_N)$ be a random sample from a population with density $f_{\boldsymbol{\theta}(\boldsymbol{u})}$ with a parameter vector $\boldsymbol{u}$ to be estimated. Note that the sample mean vector $\bar{\boldsymbol{x}}=(\boldsymbol{x}_1+\boldsymbol{x}_2+\cdots+\boldsymbol{x}_N)/N$ is sufficient. Further $\bar{\boldsymbol{x}}$ is the MLE of $\boldsymbol{\pi}$ if the underlying density is supposed only to belong to the full family $\mathcal{F}$. Let $\boldsymbol{\phi}$ be a one-to-one transformation of $\boldsymbol{\pi}$ into $\boldsymbol{\tau}$. We say the family $\tilde{\mathcal{F}}$ to be flat with respect to $\boldsymbol{\phi}$, or $\boldsymbol{\tau}$-flat if there exists a parameter vector $\boldsymbol{v}$ of $\tilde{\mathcal{F}}$ with the transformation $\boldsymbol{\xi}$ of $\boldsymbol{u}$ into $\boldsymbol{v}$ such that $\boldsymbol{\phi}$ is an affine mapping of $\boldsymbol{v}$, i.e.,

$$\boldsymbol{\phi}[\boldsymbol{\pi}(\boldsymbol{u})]=X\boldsymbol{\xi}(\boldsymbol{u})+\boldsymbol{c} \tag{1.3}$$

with a known matrix X of size $n\times m$ and a vector $\boldsymbol{c}$ of dimension n. The relation (1.3) is equivalent that the image $\{\boldsymbol{\phi}[\pi(\boldsymbol{u})]:\ \boldsymbol{u}\in U\}$ is the intersection of an m-dimensional plane and the image $\{\boldsymbol{\phi}(\boldsymbol{\pi}):\ \boldsymbol{\pi}\in\Pi\}$.

The implicit function theorem leads to a local existence of $\boldsymbol{\phi}$ such that $\tilde{\mathcal{F}}$ is flat with respect to $\boldsymbol{\phi}$. Generally we cannot give a construction of $\boldsymbol{\phi}$. However we will show in a subsequent discussion that such a transformation for some genetic models is possible to be naturally constructed by using the genetical relations.

Example 1. We consider a genetic model that two factors are linked with a recombination fraction u, see § 5g in Rao [11]. In the case of coupling, the cell parameters are described as

$$\pi_1(u)=\frac{1}{4}(3-2u+u^2)\ ,$$

$$\pi_2(u)=\pi_3(u)=\frac{1}{4}(2u-u^2)\ ,$$

$$\pi_4(u)=\frac{1}{4}(1-2u+u^2)\ ,$$

Then in terms of a transformation $\xi(u)=(1-u)^2$, this model is flat with respect to the identity mapping, or

$$\boldsymbol{\pi}(u)=\left(\frac{1}{4},-\frac{1}{4},-\frac{1}{4},\frac{1}{4}\right)'\xi(u)+\left(\frac{1}{2},\frac{1}{4},\frac{1}{4},0\right)'.$$

Mahalanobis [10] proposed a notion of a generalized distance. We apply this notion to our situation with respect to the transformation $\boldsymbol{\phi}$ of $\boldsymbol{\pi}$ into $\boldsymbol{\tau}$. A quasi-distance between densities $f_{\boldsymbol{\pi}_1}$ and $f_{\boldsymbol{\pi}_2}$ in $\mathcal{F}$ is introduced as

$$(\boldsymbol{\tau}_1-\boldsymbol{\tau}_2)'(\mathrm{Var}_{\boldsymbol{\tau}_1}\boldsymbol{\phi}(\bar{\boldsymbol{x}}))^{-1}(\boldsymbol{\tau}_1-\boldsymbol{\tau}_2)\ ,$$

or an approximated version

$$(\boldsymbol{\tau}_1-\boldsymbol{\tau}_2)'I_{\boldsymbol{\tau}_1}(\boldsymbol{\tau}_1-\boldsymbol{\tau}_2)\qquad(=D^2(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2),\ \text{say})$$

with the Fisher information matrix $I_{\boldsymbol{\tau}}$ of $\boldsymbol{\tau}$, where $\boldsymbol{\tau}_1=\boldsymbol{\phi}(\boldsymbol{\pi}_1)$ and $\boldsymbol{\tau}_2=\boldsymbol{\phi}(\boldsymbol{\pi}_2)$. For notational covariance we write $\boldsymbol{\phi}[\boldsymbol{\pi}(\boldsymbol{u})]=X\boldsymbol{u}+\boldsymbol{c}$ by reparametrization of $\boldsymbol{u}$ into $\boldsymbol{v}$. Further we write $\boldsymbol{\tau}(\boldsymbol{u})=\boldsymbol{\phi}[\boldsymbol{\pi}(\boldsymbol{u})]$ and $\hat{\boldsymbol{\tau}}=\boldsymbol{\phi}(\bar{\boldsymbol{x}})$. Then a minimizer of $D^2[\hat{\boldsymbol{\tau}},\boldsymbol{\tau}(\boldsymbol{u})]$ with respect to $\boldsymbol{u}$ is given as

$$\hat{\boldsymbol{u}}=(X'I_{\hat{\boldsymbol{\tau}}}X)^{-1}X'I_{\hat{\boldsymbol{\tau}}}(\hat{\boldsymbol{\tau}}-\boldsymbol{c})\ .$$

We call $\hat{\boldsymbol{u}}$ a projection estimator corresponding to $\boldsymbol{\phi}$. The statistic $\boldsymbol{\tau}(\hat{\boldsymbol{u}})$ is the point projected $\hat{\boldsymbol{\tau}}$ onto the flat surface $\{\boldsymbol{\tau}(\boldsymbol{u})\colon \boldsymbol{u}\in U\}$. At the same time it holds that

$$D^2[\hat{\boldsymbol{\tau}},\boldsymbol{\tau}(\hat{\boldsymbol{u}})]=D^2[\hat{\boldsymbol{\tau}},\boldsymbol{\tau}(\boldsymbol{u})]+\hat{D}^2[\boldsymbol{\tau}(\boldsymbol{u}),\boldsymbol{\tau}(\hat{\boldsymbol{u}})]$$

for any $\boldsymbol{u}$ in U, where

$$\hat{D}^2(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=(\boldsymbol{\tau}_1-\boldsymbol{\tau}_2)'I_{\hat{\boldsymbol{\tau}}}(\boldsymbol{\tau}_1-\boldsymbol{\tau}_2)\ .$$

Note that $\hat{\boldsymbol{\tau}}$ is a sufficient statistic and the MLE for the full family $\mathcal{F}$ without structure. The projection estimator $\hat{\boldsymbol{u}}$ has the following properties:

(1) *Let $\boldsymbol{\xi}$ be a parameter transformation of $\boldsymbol{u}$ into $\boldsymbol{v}$. Then the projective estimator $\hat{\boldsymbol{v}}$ of $\boldsymbol{v}$ has a property $\hat{\boldsymbol{v}}=\boldsymbol{\xi}(\hat{\boldsymbol{u}})$.*

(2) *The projection estimator corresponding to $\boldsymbol{\phi}$ is invariant under*

one-to-one affine transformations of ϕ.

(3) *The limiting distribution* $\sqrt{N}(\hat{u}-u)$ *is the normal law* $\mathcal{N}[0, (X' \cdot I_{\tau(u)}X)^{-1}]$. *That is, the projection estimator is first-order efficient.*

(4) *The statistic* $ND^2(\hat{\tau}, \tau(\hat{u}))$ *follows asymptotically the* χ^2 *distribution with* $(n-m)$ *degrees of freedom.*

Note that the property (1) is also satisfied with all the contrast estimators, see Eguchi [7]. The proof of (2) follows from the invariance of D^2 under one-to-one affine transformations. The proofs of (3) and (4) are immediate on account of the usual delta method, since $\sqrt{N}(\hat{\tau}-\tau(u))$ converges to $\mathcal{N}[0, I^{-1}_{\tau(u)}]$ in distribution.

When we consider as if the sufficient statistic $\hat{\tau}$ followed the linear regression

$$\hat{\tau}=Xu+e$$

with the normal error e of mean 0 and covariance $I^{-1}_{\hat{\tau}}$, the projection estimator can be regarded as the Gauss-Markov estimator under the hypothetical model. It is natural for us to think that the approximation to the model should lead to some difference between the projection estimator and the MLE.

We investigate the second-order property of the projection estimator by applying the formula of Eguchi [6]. Let $\mathcal{P}$ be a regular parametric family of densities with the carrier measure μ:

$$\mathcal{P}=\{p_\theta(x):\ \theta \in \Theta\}$$

with the parameter vector (coordinates) θ. We call ρ a contrast function on $\mathcal{P}$ if $\rho(\theta_1, \theta_2) \geqq 0$ with equality if and only if $\theta_1=\theta_2$. Note that the function D^2 is a contrast function on $\mathcal{F}$. The contrast function ρ generates a metric $I^{(\rho)}$ and an affine connection $\Gamma^{(\rho)}$ on $\mathcal{P}$ which are defined as

$$I^{(\rho)}_{ij}(\theta)=-\frac{\partial^2}{\partial\theta_1^i\partial\theta_2^j}\rho(\theta_1, \theta_2)\Big|_{\theta_1=\theta_2=\theta}$$

$$\Gamma^{(\rho)}_{ij,k}(\theta)=-\frac{\partial^3}{\partial\theta_1^i\partial\theta_1^j\partial\theta_2^k}\rho(\theta_1, \theta_2)\Big|_{\theta_1=\theta_2=\theta}$$

with the coordinates $\theta=(\theta^i)$ respectively. On the other hand, Amari introduced the information metric I and the mixture connection $\Gamma^{(m)}$ to be the components

$$I_{ij}(\theta)=E\left[-\frac{\partial^2}{\partial\theta^i\partial\theta^j}\log p_\theta\right]$$

$$\Gamma^{(m)}_{ij,k}(\theta)=E\left[\frac{\partial^2}{\partial\theta^i\partial\theta^j}\log p_\theta \frac{\partial}{\partial\theta^k}\log p_\theta\right]$$

$$+E\left[\frac{\partial}{\partial\theta^i}\log p_\theta \frac{\partial}{\partial\theta^j}\log p_\theta \frac{\partial}{\partial\theta^k}\log p_\theta\right]$$

with respect to θ-coordinates, see Amari [1] for the historical backgrounds. Applying this formulas to our situation, it holds on the full exponential family $\mathcal{F}$ that

$$\begin{aligned} &I_{ij}^{(D^2)}(\tau)=I_{ij}(\tau)\ , \\ &\Gamma_{ij,k}^{(m)}(\tau)=\frac{\partial\pi^r}{\partial\tau^i\partial\tau^j}\frac{\partial\tau^l}{\partial\pi^r}I_{lk}(\tau)\ , \\ &\Gamma_{ij,k}^{(D^2)}(\tau)=\frac{\partial}{\partial\tau^i}I_{jk}(\tau)+\frac{\partial}{\partial\tau^j}I_{ki}(\tau) \end{aligned} \tag{1.4}$$

with respect to τ-coordinates, where $\tau(\pi)=\phi(\pi)$ and $\pi(\tau)=\phi^{-1}(\tau)$ with the transformation $\phi: \pi\to\tau$. Here we use the summation convension as for the indices r and l in the second equation. Note that the first relation of (1.4) gives another proof of the first-order efficiency of the projection estimator $\hat{u}$.

We adopt information loss as measure of optimality for estimation. Let $\check{u}$ be an estimator of u. The information loss due to $\check{u}$ is defined by the difference between the Fisher information matrix with the sample of size N and that with the estimator $\check{u}$, which we denote by $\Delta_N(\check{u})$. By applying the relations (1.4) to the formula of Eguchi [6], we have that for the projection estimator $\hat{u}$,

$$\lim_{N\to\infty}\Delta_N(\hat{u})=\lim_{N\to\infty}\Delta_N(\check{u})+\langle\Gamma^{(m)}-\Gamma^{(D^2)}\rangle^2$$

where $\check{u}$ denotes a second-order efficient estimator, e.g. the MLE. Here the (a, b)-entry of $\langle\Gamma^{(m)}-\Gamma^{(D^2)}\rangle^2$ is

$$\{\Gamma_{ij,k}^{(m)}-\Gamma_{ij,k}^{(D^2)}\}\{\Gamma_{rs,t}^{(m)}-\Gamma_{rs,t}^{(D^2)}\}X_a^iX_b^rE^{js}E^{kt}\ ,$$

evaluated at the true parameter point $\tau(u)$, where $X=(X_a^i)_{i,a}$ and

$$E^{ij}=I^{ij}-X_a^i\tilde{I}^{ab}X_b^j$$

with the inverse element I^{ij} of I_{ji} and the inverse element $\tilde{I}^{ab}$ of $X_b^iI_{ij}X_a^j$. Therefore the projection estimator $\hat{u}$ is not generally second-order efficient. However the second-order inefficiency of $\hat{u}$ is not so serious for us. Because the one-step MLE from the projection estimator $\hat{u}$,

$$\hat{u}_1=\hat{u}+\tilde{I}_{\hat{u}}^{-1}S(\hat{u})$$

is seen to be second-order efficient by using Theorem 1 in [6], where $\tilde{I}_u$ is the Fisher information matrix of u and $S(u)$ is the score function of u.

2. Some examples

We present some examples applicable for the projection estimator. First we back to the ABO blood group model in Section 1. The following relations are easily seen from (1.1):

$$\pi_1(a, b)+\pi_3(a, b)=(a+o)^2\,, \qquad \pi_2(a, b)+\pi_3(a, b)=(b+o)^2\,.$$

Based on this relations, we introduce a transformation ϕ of $\tilde{\boldsymbol{\pi}}=(\pi_1, \pi_2, \pi_3)'$ into $\boldsymbol{\tau}=(\tau_1, \tau_2, \tau_3)'$ by

$$\boldsymbol{\tau}=\phi(\tilde{\boldsymbol{\pi}})=(\sqrt{\pi_1+\pi_3}-\sqrt{\pi_3}\,, \sqrt{\pi_2+\pi_3}-\sqrt{\pi_3}\,, \sqrt{\pi_3})'\,.$$

The ABO blood group model is represented as a linear form:

$$\phi[\tilde{\boldsymbol{\pi}}(a, b)]=(a, b, o)'=X(a, b)'+(0, 0, 1)'\,,$$

where

$$X=\begin{pmatrix}1 & 0 & -1\\ 0 & 1 & -1\end{pmatrix}'\,,$$

see Fig. 2.

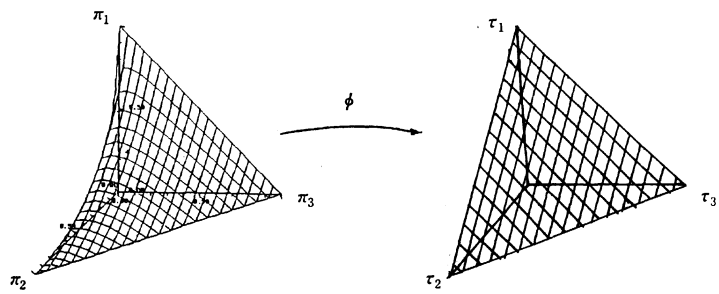


Fig. 2. The deformation of $\{\tilde{\boldsymbol{\pi}}(a, b)\colon (a, b)\in S_2\}$ into $\{\boldsymbol{\tau}(a, b)\colon (a, b)\in S_2\}$ where S_2 denotes the 2-dimensional simplex.

By analogy with squared Mahalanobis distance D^2, we define a function as

$$D^2(\hat{\boldsymbol{\tau}}, \boldsymbol{\tau}(a, b))=\{\hat{\boldsymbol{\tau}}-\boldsymbol{\tau}(a, b)\}'V(\hat{\boldsymbol{\tau}})^{-1}\{\hat{\boldsymbol{\tau}}-\boldsymbol{\tau}(a, b)\}\,,$$

where $V(\hat{\boldsymbol{\tau}})$ is an estimate of the variance of $\hat{\boldsymbol{\tau}}$ with $\hat{\boldsymbol{\tau}}=\phi((\bar{p}_1, \bar{p}_2, \bar{p}_3)')$ and $\boldsymbol{\tau}(a, b)=\phi[\tilde{\boldsymbol{\pi}}(a, b)]$. The projection estimator corresponding to ϕ is expressed as

$$\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix}=(X'V(\hat{\boldsymbol{\tau}})^{-1}X)^{-1}X'V(\hat{\boldsymbol{\tau}})^{-1}\left\{\hat{\boldsymbol{\tau}}-\begin{pmatrix}0\\ 0\\ 1\end{pmatrix}\right\}$$

by minimization of the function with respect to (a, b). We adopt the inverse of the Fisher information matrix of τ as $V(\hat{\tau})$, i.e.,

$$V(\hat{\tau})^{-1}=J(\pi)'^{-1}G(\pi)J(\pi)^{-1}$$

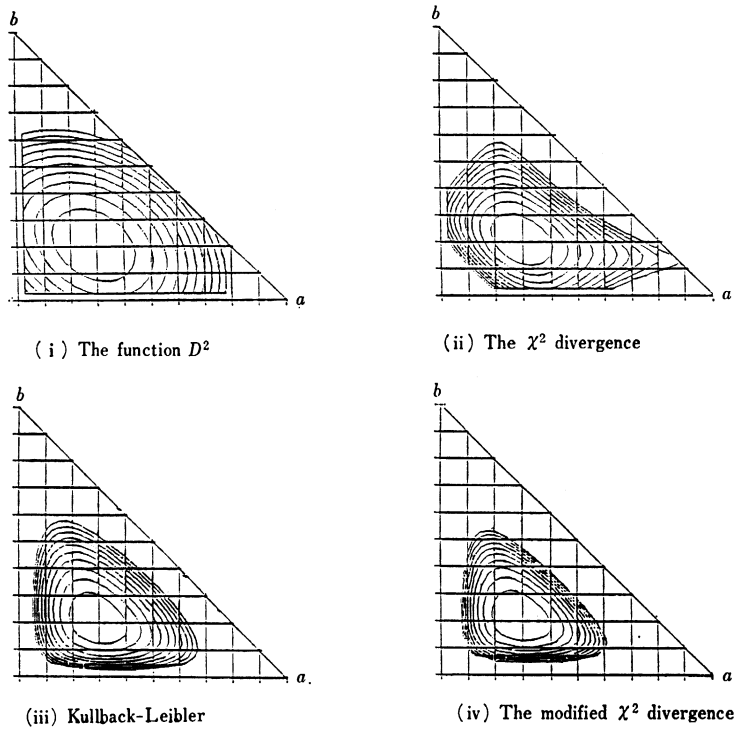
with $\tilde{\pi}=\phi^{-1}(\tau)$, where $G(\tilde{\pi})=(\delta_{ij}/\pi_i+1/(1-\pi_1-\pi_2-\pi_3))_{i,j=1,2,3}$ and

$$J(\tilde{\pi})=\frac{1}{2}\begin{pmatrix}1/(\sqrt{\pi_1+\pi_3}) & 0 & 1/(\sqrt{\pi_1+\pi_3})-1/\sqrt{\pi_3}\\ 0 & 1/(\sqrt{\pi_2+\pi_3}) & 1/(\sqrt{\pi_2+\pi_3})-1/\sqrt{\pi_3}\\ 0 & 0 & 1/\sqrt{\pi_3}\end{pmatrix}.$$

On the other hand, the classical estimators including the MLE are defined by minimization of the respective functions as follows:
The MLE; the Kullback-Leibler

$$\rho_{KL}(\bar{p}, \pi)=\sum_{i=1}^{4}\bar{p}_i \log(\bar{p}_i/\pi_i),$$

the minimum χ^2 estimator;



(i) The function D^2 (ii) The χ^2 divergence

(iii) Kullback-Leibler (iv) The modified χ^2 divergence

Fig. 3. Contours of contrast functions in the space of genes frequencies a, b. For the corresponding contrast functions (ρ's, say), we define a contour

$$C_\kappa=\{(a, b): \rho(\bar{x}, \pi(a, b))=\kappa\}$$

centered at $\bar{x}=(.38, .26, .11)$ with $\kappa=.1, .2, \cdots, 1.0$.

$$\chi^2(\bar{\boldsymbol{p}}, \boldsymbol{\pi})=\sum_{i=1}^{4}\frac{(\bar{p}_i-\pi_i)^2}{\pi_i},$$

the minimum modified χ^2 estimator; mod $\chi^2(\bar{\boldsymbol{p}}, \boldsymbol{\pi})=\chi^2(\boldsymbol{\pi}, \bar{\boldsymbol{p}})$ and the minimum Haldane k-discrepancy estimator;

$$\rho_k(\bar{\boldsymbol{p}}, \boldsymbol{\pi})=\frac{1}{k(1-k)}\left(1-\sum_{i=1}^{4}\pi_i^{1-k}\bar{p}_i^k\right).$$

As well as the MLE, other estimators have no explicit form. We note that all the contours of these functions are sinuous in the parameter space S_2 of gene frequencies a and b. On the other hand, the contours of $D^2(\hat{\boldsymbol{\tau}}, \boldsymbol{\tau}(a, b))$ form concentric ellipses in S_2, see Fig. 3. A numerical comparison among these estimators is given by the following table. We observe from the table that the projection estimator has reasonable values together with the classical estimators.

Table. A numerical comparison among the projection estimator and classical estimators based on the simulated values generated at $(a, b)=(.3, .2)$ with replications 100.

observed frequencies	(Case 1) $\bar{x}=(.38, .26, .25, .11)$	(Case 2) $\bar{x}=(.37, .13, .31, .19)$	(Case 3) $\bar{x}=(.44, .22, .18, .16)$
(i) The projection estimator:			
	$\hat{a}=.28664$	$\hat{a}=.30485$	$\hat{a}=.36668$
	$\hat{b}=.20670$	$\hat{b}=.13950$	$\hat{b}=.21215$
(ii) The minimum χ square estimator:	$\hat{a}=.28704$	$\hat{a}=.32254$	$\hat{a}=.36673$
	$\hat{b}=.20722$	$\hat{b}=.18154$	$\hat{b}=.21216$
(iii) The maximum likelihood estimator:	$\hat{a}=.28098$	$\hat{a}=.32499$	$\hat{a}=.36673$
	$\hat{b}=.20709$	$\hat{b}=.16932$	$\hat{b}=.21213$
(vi) The modified minimum χ square estimator:	$\hat{a}=.28682$	$\hat{a}=.30527$	$\hat{a}=.36674$
	$\hat{b}=.20669$	$\hat{b}=.13423$	$\hat{b}=.21205$

Example 2. (MNS system of blood groups). We treat another genetic model that consists of four allels MS, Ms, NS and Ns with gene frequencies q_1, q_2, q_3 and q_4, respectively, see Cepellini, Siniscalco and Smith [4]. A population with random mating under the assumption of the Hardy-Weinberg law follows from the phenotype frequencies:

phenotype	expected frequencies
MS	$\pi_1=q_1^2+2q_1q_2$
Ms	$\pi_2=q_2^2$
MNS	$\pi_3=2q_1q_3+2q_1q_4+2q_2q_3$
MNs	$\pi_4=2q_2q_4$
NS	$\pi_5=q_3^2+2q_3q_4$
Ns	$\pi_6=q_4^2$

because of the dominance of S to s. By the similar way to the ABO blood group system, we define a transformation $\boldsymbol{\phi}$ of $\boldsymbol{\pi}=(\pi_1, \pi_2, \pi_3, \pi_4, \pi_5)'$ into $\boldsymbol{\tau}=(\tau_1, \tau_2, \tau_3, \tau_4, \tau_5)'$ by

$$\boldsymbol{\phi}(\boldsymbol{\pi})=(\sqrt{\pi_1+\pi_2}, \sqrt{\pi_2}, \sqrt{\pi_5+\pi_6}, \sqrt{\pi_6}, \pi_4/2\sqrt{\pi_2})'$$

with $\pi_6=1-\pi_1-\pi_2-\pi_3-\pi_4-\pi_5$. Thus the model has a linear form

$$\boldsymbol{\phi}(\pi)=X\boldsymbol{q}+\boldsymbol{b}$$

with $\pi=(\pi_1, \pi_2, \pi_3, \pi_4, \pi_5)'$, $\boldsymbol{q}=(q_1, q_2, q_3)'$,

$$X'=\begin{pmatrix}1 & 0 & -1 & -1 & -1\\ 1 & 1 & -1 & -1 & -1\\ 0 & 0 & 0 & -1 & -1\end{pmatrix}$$

and $\boldsymbol{b}=(0, 0, 1, 1, 1)'$. Based on the observation quated by Cepellini et al. [4]:

Phenotype	MS	Ms	MNS	MNs	NS	Ns	
Data	44	35	62	47	21	21	(Total 230),

we have the projection estimate corresponding to ϕ is $\hat{\boldsymbol{q}}=(.37419, .19649, .2319)'$ and the MLE is $\check{\boldsymbol{q}}=(.3798, .2004, .2892)$.

Example 3 (linear covariance structure). Anderson [2] considered the following structure

$$\Sigma(\boldsymbol{u})=u_1V_1+u_2V_2+\cdots+u_mV_m$$

with a parameter vector $\boldsymbol{u}=(u_1, u_2, \cdots, u_m)'$, where the symmetric matrices $V_1, V_2, \cdots, V_m$ are linearly independent. He proposed a first-order efficient estimator $\hat{\boldsymbol{u}}=G(S)^{-1}\boldsymbol{e}(S)$ for a sample covariance matrix S from $\mathcal{N}(0, \Sigma(\boldsymbol{u}))$, where

$$G(S)=(\text{trace}\,(V_aS^{-1}V_bS^{-1}))_{a,b=1,2,\cdots,m}$$

and

$$\boldsymbol{e}(S)=(\text{trace}\,(V_aS^{-1}))_{a=1,2,\cdots,m}\,.$$

By our terminology, the model is Σ-flat and $\hat{\boldsymbol{u}}$ is the projection estimator.

Example 4. Guerrero and Johnson [8] apply a linear regression for binary data adapted the Box-Cox transformation. Let $\{y_i\}$ be independently distributed binomial random variables $Bi(N_i, \pi_i)$ for $i=1, 2, \cdots, n$. The Box-Cox transformation for the odds ratio is adapted as follows:

$$\phi_\lambda\left(\frac{\pi_i}{1-\pi_i}\right)=\boldsymbol{Z}_i'\boldsymbol{\beta}$$

with a parameter $\boldsymbol{\beta}$ and a regressor $\boldsymbol{Z}_i$, where

$$\phi_\lambda(t)=\begin{cases}\log t\,, & \lambda=0\,,\\ \lambda^{-1}(1-t^\lambda)\,, & \text{otherwise}\,.\end{cases}$$

The random vector $(y_1, y_2, \cdots, y_n)'$ associates with a flat model with respect to $\boldsymbol{\phi}=(\phi_\lambda\circ\psi, \cdots, \phi_\lambda\circ\psi)'$ where $\psi(\pi)=\pi/(1-\pi)$. Thus the projection estimator corresponding to ϕ has a closed form

$$\hat{\boldsymbol{\beta}}=(ZVZ')^{-1}ZV\hat{\boldsymbol{\tau}}\,,$$

where $\hat{\boldsymbol{\tau}}=\boldsymbol{\phi}((y_1/N_1, \cdots, y_n/N_n)')$, $Z=(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \cdots, \boldsymbol{Z}_n)$ and V is a diagonal matrix of $(\{\pi_i/(1-\pi_i)\}^{1-2\lambda}(1-\pi_i)^2)_{i=1,2,\cdots,n}$.

Example 5. We treat a tri-allelic locus model with inbreeding, see Yasuda [13]. Let p, q and r be the frequency parameters of alleles A, B and C, so that $r=1-p-q$ and let α be the inbreeding coefficient. The phenotypes are given as follows:

Phenotype	Expected frequency
A A	$\pi_1=\beta p^2+\alpha p$
B B	$\pi_2=\beta q^2+\alpha q$
C C	$\pi_3=\beta r^2+\alpha r$
A B	$\pi_4=2\beta pq$
B C	$\pi_5=2\beta qr$
C A	$\pi_6=2\beta rp$

with $\beta=1-\alpha$. Noting the equations

$$\frac{q}{p}=\frac{2\pi_2+\pi_4}{2\pi_1+\pi_4}\qquad\text{and}\qquad\frac{r}{q}=\frac{2\pi_3+\pi_5}{2\pi_2+\pi_5}\,,$$

we can construct a transformation $\boldsymbol{\phi}$ of $\boldsymbol{\pi}=(\pi_1, \pi_2, \pi_3, \pi_4, \pi_5)'$ into $\boldsymbol{\tau}=(\tau_1, \tau_2, \tau_3, \tau_4, \tau_5)'$ by

$$\tau_1=(2\pi_1+\pi_4)(2\pi_2+\pi_5)/\lambda(\boldsymbol{\pi})\,,$$

$$\tau_2=(2\pi_2+\pi_4)(2\pi_2+\pi_5)/\lambda(\boldsymbol{\pi})\,,$$

$$\tau_3=(2\pi_1+\pi_4)(2\pi_2+\pi_5)\pi_5/[\lambda(\boldsymbol{\pi})\pi_4]\,,$$

$$\tau_4=1-\frac{1}{2}\pi_4\lambda(\boldsymbol{\pi})^2/\{(2\pi_1+\pi_4)(2\pi_2+\pi_5)(2\pi_2+\pi_4)(2\pi_2+\pi_5)\}\,,$$

$$\tau_5=\frac{1}{2}(1-\pi_1-\pi_2-\pi_3-\pi_4-\pi_5)\lambda(\boldsymbol{\pi})^2\pi_4/\{(2\pi_1+\pi_4)^2(2\pi_2+\pi_5)^2\pi_5\}\ ,$$

where

$$\lambda(\boldsymbol{\pi})=(2\pi_1+\pi_4)(2\pi_2+\pi_5)+(2\pi_2+\pi_5)(2\pi_2+\pi_4)+(2\pi_2+\pi_4)(2\pi_3+\pi_5)\ .$$

Accordingly we have a linear transformation $\phi(\boldsymbol{\pi})=(p, q, r, \alpha, \beta)'$ for $\boldsymbol{\pi}=(\pi_1, \pi_2, \pi_3, \pi_4, \pi_5)$ with relation to $r=1-p-q$ and $\beta=1-\alpha$. That is to say, the genetic model is flat with respect to $\boldsymbol{\phi}$. Hence we can obtain the projection estimator corresponding to $\boldsymbol{\phi}$ in terms of the Jacobi matrix of $\boldsymbol{\phi}$.

3. Discussion

The method of estimation by projection has been found usefull in both the computational and asymptotical aspects. We have to mention as its demerit that this method is necessary to construct a transformation of the sufficient statistic so that the model may be flat with respect to the transformation. But we observe through some genetic models that such transformations can be naturally sought by using the relations among the phenotopic frequencies. Haberman [9] presented a simple iterative method for obtaining the MLE in a model with indirect observation including our treating models. Burn [3] recommended the use of the program package GLIM by representing models as loglinear form with composit link functions. It is our assertion that we can obtain a first-order efficient estimate without such algorithms.

Furthermore we present a simple modification of the projection estimator as follows. Return to the situations that the model is flat with respect to a transformation $\boldsymbol{\phi}: \boldsymbol{\pi}\to\boldsymbol{\tau}$, i.e.,

$$\boldsymbol{\phi}(\boldsymbol{\pi}(\boldsymbol{u}))=X\boldsymbol{u}+\boldsymbol{c}\ .$$

The simplified projection estimator is defined as

$$\hat{\boldsymbol{u}}_s=(X'X)^{-1}X'\{\boldsymbol{\phi}(\bar{\boldsymbol{x}})-\boldsymbol{c}\}$$

with the sufficient statistic $\bar{\boldsymbol{x}}$. It is easily seen that the estimator $\hat{\boldsymbol{u}}_s$ is Fisher-consistent, which implies to be strongly consistent. In this way, we can introduce a strongly consistent estimator without the evaluation of the Fisher information matrix of $\boldsymbol{\tau}=\boldsymbol{\phi}(\boldsymbol{\pi})$ at $\boldsymbol{\pi}=\bar{\boldsymbol{x}}$.

Acknowledgments

I am grateful to a referee for careful reading the manuscript. I wish also to thank Professor M. Okamoto and Dr. M. Goto for valuable

suggestions and Professor Y. Fujikoshi for inspiring encouragement.

HIROSHIMA UNIVERSITY

REFERENCES

[1] Amari, S. (1985). *Differential geometrical methods in statistics, Lecture Note in Statistics*, Springer, New York.
[2] Anderson, T. W. (1973). Asymptotic efficient estimator of covariance matrices with linear structure, *Ann. Statist.*, **1**, 135-141.
[3] Burn, R. (1982). Loglinear models with composite link functions in genetics, GLIM 82: *Proc. Internat. Conf. General. Linear Models* (ed. R. Girchrist), 144-154, Springer, New York.
[4] Cepellini, R., Siniscalo, M. and Smith, C. A. B. (1955). The estimation of gene frequencies in random mating population, *Ann. Hum. Genet.*, **20**, 97-115.
[5] Draper, N. R. and Smith, H. (1968). *Applied Regression Analysis*, Wiley, New York.
[6] Eguchi, S. (1983). Second order efficiency of minimum contrast estimators in a curved exponential family, *Ann. Statist.*, **11**, 793-803.
[7] Eguchi, S. (1984). A characterization of second order efficiency in a curved exponential family, *Ann. Inst. Statist. Math.*, **36**, 199-206.
[8] Gurrero, V. M. and Johnson, R. A. (1982). Use of the Box-Cox transformation with binary response models, *Biometrika*, **69**, 309-314.
[9] Haberman, S. J. (1977). Product models for frequency tables involving indirect observation, *Ann. Statist.*, **5**, 1124-1147.
[10] Mahalanobis, P. C. (1936). On the generalized distance in statistics, *Proc. Nat. Inst. Sci. Ind.*, **2**, 49-55.
[11] Rao, C. R. (1972). *Linear Statistical Inference and its Applications*, Wiley, New York.
[12] Takeuchi, K., Yanai, H. and Mukherjee, B. N. (1982). *The Foundations of Multivariate Analysis*, Wiley, Tokyo.
[13] Yasuda, N. (1968). Estimation of inbreeding coefficient from phenotype frequencies by a method of maximum likelihood scoring, *Biometrics*, **24**, 915-934.