87-F-10

On Selection of Statistical Models

by

Yukio Suzuki
The University of Tokyo

October 1987

PREPARED UNDER
THE PROJECT ON STATISTICS
RESEARCH INSTITUTE FOR THE JAPANESE ECONOMY

ON SELECTION OF STATISTICAL MODELS

YUKIO SUZUKI

THE UNIVERSITY OF TOKYO

JAPAN

1. PRINCIPLE OF SELECTION OF STATISTICAL MODELS

The m kinds of models $M_1, M_2, \ldots, M_m$ are considered as the possible models concerning the distribution of the observed random variable $\tilde{x}$. Each model M_i is supposed to be specified by the following assumptions.

(1) M_i has the unknown parameter θ_i with the parameter space Θ_i.

(2) For given M_i and θ_i, $\tilde{x}$ has the known p. d. f. $p(x|M_i, \theta_i)$ with respect to some measure μ on the measureable space $(\mathfrak{X}, B(\mathfrak{X}))$, where $\mathfrak{X}$ and $B(\mathfrak{X})$ are the sample space and a σ-field of $\mathfrak{X}$, respectively.

(3) The prior p. d. f. $p(\theta_i|M_i)$ (w. r. t. ν_i) of θ_i, given M_i, is known, where ν_i is a measure on $(\Theta_i, B(\Theta_i))$, $B(\Theta_i)$ being a σ-field of Θ_i.

(4) The prior probability of M_i is available and denoted by $p(M_i)$ ($i=1, \ldots, m$). Of course, $\sum_{i=1}^{m} p(M_i)=1$.

From the above assumptions we can obtain the posterior probability $p(M_i|x)$ of the model M_i for the observed value $\tilde{x}=x$ by

$$p(M_i|x) = \frac{p(M_i)p(x|M_i)}{\sum_{j=i}^{m} p(M_j)p(x|M_j)} \qquad (i=1,\ldots, m) \tag{1.1}$$

where

$$p(x|M_i)=\int_{\Theta_i} p(\theta_i|M_i)p(x|M_i, \theta_i)\nu_i(d\theta_i) \qquad (i=1,\ldots, m) \tag{1.2}$$

From the Bayesian viewpoint, our selection should be made solely on the basis of the posterior probabilities $p(M_i|x)$ ($i=1,\ldots, m$). Of course, we may decide to take further observations if any of $p(M_1|x), \ldots, p(M_m|x)$ is not sufficiently near to 1. Thus, the principle of selection is stated as follows:

(i) If $\max_i p(M_i|x)$ is sufficiently near to one, it is reasonable for us to choose the model M_{i*} such that $\max_i p(M_i|x)=p(M_{i*}|x)$.

(ii) On the contrary, if none of $p(M_i|x)$ $(i=1, \ldots, m)$ are sufficiently near to one, it is difficult for us to choose a model with confidence. In order to overcome this indecisive situation, we should gather more sample information. Thus we are naturally led to sequential (or multi-stage) selection procedures (Suzuki [2]).

As is seen from the above principle, the posterior probabilities of models $M_i (i=1, \ldots, m)$ are fundamentally important. Hence, throughout this paper, our concerns are concentrated on the derivation of posterior probabilities of models.

In the following sections, we will treat several examples of selection of models, on the basis of the above principle. The selection of shape parameters of gamma distributions and Weibull distributions are treated in sections 2 and 3, respectively. In section 4, problems of selection concerning normal distributions are discussed. In section 5, the selection of regressors in a linear regression model is considered. Further, in section 6, we treat the selection of regressors in a linear regression model in case of vague prior information, where we use the concept of intermediate prior distributions and the corresponding concept of intermediate posterior distributions which were first introduced by the author in [1]. In section 7, the selection of orders of polynomial regression models is treated as a special case of section 6.

2. SELECTION OF SHAPE PARAMETER OF GAMMA DISTRIBUTIONS

Let $\tilde{x}=(\tilde{x}_1, \ldots, \tilde{x}_n)$ be the observed random variable and let $\mathcal{X}=R^n_+=\{(x_1, \ldots, x_n)|x_i>0 \ (i=1, \ldots, n)\}$ be the sample space. The m kinds of models

$M_1, \ldots, M_m$ are the possible candidates which are specified as follows.

(1) M_i has the unknown parameter θ_i with the parameter space $\Theta_i = \mathbb{R}_+$ and has the known shape parameter $\alpha(i)$.

(2) For given M_i and θ_i, $\tilde{x}_1, \ldots, \tilde{x}_n$ are independent and $\tilde{x}_j \sim \mathrm{Gamma}(\alpha(i), \theta_i^{-1})$, $(j=1, \ldots, n)$, where $\alpha(i)$ is known, that is,

$$(2.1) \qquad p(\boldsymbol{x}|M_i, \theta_i) = \prod_{j=1}^{n} \left\{\frac{1}{\Gamma(\alpha(i))}\theta_i^{\alpha(i)} x_j^{\alpha(i)-1} \exp(-\theta_i x_j)\right\}$$

$$= \left(\frac{1}{\Gamma(\alpha(i))}\right)^n \theta_i^{n\alpha(i)} \left(\prod_{j=1}^{n} x_j\right)^{\alpha(i)-1} \exp\left(-\theta_i \sum_{j=1}^{n} x_j\right)$$

Without loss of generality, we can assume that $0<\alpha(1)<\alpha(2)<\ldots<\alpha(m)$.

(3) The prior distribution of $\tilde{\theta}_i$, given M_i, is $\mathrm{Gamma}(\alpha_0(i), \beta_0(i))$, that is,

$$(2.2) \qquad p(\theta_i|M_i) = \frac{1}{\Gamma(\alpha_0(i))}(\beta_0(i))^{-\alpha_0(i)} \theta_i^{\alpha_0(i)-1} \exp\left(-\frac{\theta_i}{\beta_0(i)}\right) \qquad (\theta_i \in R_+)$$

Clearly, this is a conjugate prior distribution which is easily seen from (2.1).

(4) $p(M_1), p(M_2), \ldots, p(M_m)$ are prior probabilities of models $M_1, \ldots, M_m$.

From the above assumptions, we obtain by (1.2)

$$(2.3) \quad p(\boldsymbol{x}|M_i)$$

$$= \frac{(\beta_0(i))^{-\alpha_0(i)}}{\Gamma(\alpha_0(i))(\Gamma(\alpha(i)))^n} \left(\prod_{j=1}^{n} x_j\right)^{\alpha(i)-1} \int_0^\infty \theta_i^{\alpha_0(i)+n\alpha(i)-1} \exp\left(-\left(\frac{1}{\beta_0(i)} + \sum_{j=1}^{n} x_j\right)\theta_i\right) d\theta_i$$

$$= \frac{\Gamma(\alpha_0(i)+n\alpha(i))}{\Gamma(\alpha_0(i))(\Gamma(\alpha(i)))^n} (\beta_0(i))^{-\alpha_0(i)} \left(\prod_{j=1}^{n} x_j\right)^{\alpha(i)-1} \left(\frac{1}{\beta_0(i)} + \sum_{j=1}^{n} x_j\right)^{-(\alpha_0(i)+n\alpha(i))}$$

$$= \frac{\Gamma(\alpha_0(i)+n\alpha(i))}{\Gamma(\alpha_0(i))(\Gamma(\alpha(i)))^n} (\beta_0(i))^{n\alpha(i)} \left(\prod_{j=1}^{n} x_j\right)^{\alpha(i)-1} \left(1+\beta_0(i) \sum_{j=1}^{n} x_j\right)^{-(\alpha_0(i)+n\alpha(i))}$$

where $\boldsymbol{x}=(x_1, \ldots, x_n)$. Thus, for the observed value $\tilde{\boldsymbol{x}}=\boldsymbol{x}$, we have $p(M_i|\boldsymbol{x})$ $(i=1, \ldots, m)$ easily by the formula (1.1).

REMARK 2.1. From (2.3) it is easily seen that the statistic $\left(\prod_{j=1}^{n} \tilde{x}_j, \sum_{j=1}^{n} \tilde{x}_j\right)$

is sufficient for our selection problem.

REMARK 2.2. The statistic $\sum_{j=1}^{n} \tilde{x}_j$ is sufficient with respect to the parameter θ_i in the model M_i, but is not sufficient for our selection problem.

3. SELECTION OF SHAPE PARAMETERS OF WEIBULL DISTRIBUTIONS

Let $\tilde{\boldsymbol{x}}=(\tilde{x}_1,\ldots,\tilde{x}_n)$ be the observed random variable and let $\mathcal{X}=\mathbf{R}_+^n$ be the sample space. The m kinds of models $M_1,\ldots, M_m$ are the possible candidates which are specified as follows.

(1) M_i has the unknown parameter θ_i with $\Theta_i=R_+$ and has the known shape parameter $\alpha(i)$.

(2) For given M_i and θ_i, $\tilde{x}_1,\ldots, \tilde{x}_n$ are independently and identically distributed with $\tilde{x}_j \sim \text{Weibull}(\alpha(i), \theta_i^{-\frac{1}{\alpha(i)}})$, $(j=1,\ldots, n)$, that is,

$$(3.1)\qquad p(x|M_i, \theta_i)=\prod_{j=1}^{n} \{\alpha(i)\theta_i x_j^{\alpha(i)-1}\exp(-\theta_i x_j^{\alpha(i)})\}$$

$$=(\alpha(i)\theta_i)^n\left(\prod_{j=1}^{n} x_j\right)^{\alpha(i)-1}\exp\left(-\theta_i\sum_{j=1}^{n} x_j^{\alpha(i)}\right)\qquad (\boldsymbol{x}\in R_+^n)$$

Also we assume that $0<\alpha(1)<\alpha(2)<\ldots<\alpha(m)$.

(3) The prior distribution of θ_i, given M_i, is $\text{Gamma}(\alpha_0(i), \beta_0(i))$. This is a conjugate prior distribution as is easily seen from (3.1). Thus,

$$(3.2)\qquad p(\theta_i|M_i)=\frac{1}{\Gamma(\alpha_0(i))}(\beta_0(i))^{-\alpha_0(i)}\theta_i^{\alpha_0(i)-1}\exp\left(-\frac{\theta_i}{\beta_0(i)}\right)\qquad (\theta_i\in R_+)$$

(4) Prior probabilities $p(M_1),\ldots, p(M_m)$ are available.

From the above assumptions, we obtain by (1.2)

$$(3.3)\qquad p(\boldsymbol{x}|M_i)$$

$$=\frac{(\alpha(i))^n}{\Gamma(\alpha_0(i))(\beta_0(i))^{\alpha_0(i)}}\left(\prod_{j=1}^{n} x_j\right)^{\alpha(i)-1}\int_0^\infty \theta_i^{\alpha_0(i)+n-1}\exp\left\{-\left(\frac{1}{\beta_0(i)}+\sum_{j=1}^{n} x_j^{\alpha(i)}\right)\theta_i\right\}d\theta_i$$

$$=\frac{(\alpha(i))^n\,\Gamma(\alpha_0(i)+n)}{\Gamma(\alpha_0(i))(\beta_0(i))^{\alpha_0(i)}}\left(\prod_{j=1}^{n} x_j\right)^{\alpha(i)-1}\left(\frac{1}{\beta_0(i)}+\sum_{j=1}^{n} x_j^{\alpha(i)}\right)^{-(\alpha_0(i)+n)}$$

$$=\frac{\Gamma(\alpha_0(i)+n)}{\Gamma(\alpha_0(i))}(\beta_0(i)\alpha(i))^n(\prod_{j=1}^{n} x_j)^{\alpha(i)-1}(1+\beta_0(i)\sum_{j=1}^{n} x_j^{\alpha(i)})^{-(\alpha_0(i)+n)}$$

and the posterior probabilities $p(M_i|x)$ from (1.1) and (3.3).

REMARK 3.1. From (3.3) we can see that the statistic $(\prod_{j=1}^{n}\tilde{x}_j, \sum_{j=1}^{n}\tilde{x}_j^{\alpha(1)}, \ldots, \sum_{j=1}^{n}\tilde{x}^{\alpha(m)})$ is sufficient for our selection problem.

REMARK 3.2. It is easily seen that the statistic $\sum_{j=1}^{n}\tilde{x}_j^{\alpha(i)}$ is sufficient w. r. t. θ_i in the model M_i $(i=1,\ldots, m)$, but is not sufficient for our selection problem.

REMARK 3.3. The statistic $(\prod_{j=1}^{n}\tilde{x}_j, \prod_{j=1}^{n}\tilde{x}_j^{\alpha(i)})$ is also sufficient w. r. t. θ_i in the medel M_i $(i=1,\ldots, m)$, but is not sufficient for our selection problem.

4. SELECTION OF MODELS CONCERNING NORMAL DISTRIBUTIONS

4.1. CASE I

let $\tilde{X}=(\tilde{\boldsymbol{x}}_1,\ldots, \tilde{\boldsymbol{x}}_n)$ be the observed random variable and let $\mathcal{X}=\mathbf{R}^{np}$ be the sample space. The m kinds of models $M_1,\ldots, M_m$ are the possible models which are specified as follows.

(1) M_i has the unknown parameter θ_i with $\Theta_i=\mathbf{R}^p$ and the known variance matrix $\Sigma(i)$.

(2) For given M_i and $\boldsymbol{\theta}_i$, $\tilde{\boldsymbol{x}}_1,\ldots, \tilde{\boldsymbol{x}}_n$ are independent and $\tilde{\boldsymbol{x}}_j \sim N(\boldsymbol{\theta}_i, \Sigma(i))$, $(j=1,\ldots, n)$. Thus,

$$p(X|M_i, \boldsymbol{\theta}_i)=\prod_{j=1}^{n} n(\boldsymbol{x}_j|\boldsymbol{\theta}_i, \Sigma(i))$$

and it is easily shown that

$$(4.1)\qquad p(X|M_i,\boldsymbol{\theta}_i)=(2\pi)^{-\frac{(n-1)p}{2}} n^{-\frac{p}{2}}|\Sigma(i)|^{-\frac{n-1}{2}}\exp\{-\tfrac{1}{2}\mathrm{tr}(S\Sigma^{-1}(i))\}n(\bar{\boldsymbol{x}}|\boldsymbol{\theta}_i, \tfrac{1}{n}\Sigma(i))$$

where $\bar{\boldsymbol{x}}=\frac{1}{n}\sum_{j=1}^{n}\boldsymbol{x}_j$, $S=\sum_{j=1}^{n}(\boldsymbol{x}_j-\bar{\boldsymbol{x}})'(\boldsymbol{x}_j-\bar{\boldsymbol{x}})$ and $n(\cdot|\boldsymbol{\theta}, \Sigma)$ indicates the p. d. f. of $N(\theta, \Sigma)$.

(3) The prior distribution of $\tilde{\boldsymbol{\theta}}_i$, given M_i, is $N(\boldsymbol{\mu}_0(i)), \Sigma_0(i))$,

($(\mu_0(i), \Sigma_0(i))$: known), that is,

(4.2) $$p(\theta_i | M_i) = n(\theta_i | \mu_0(i), \Sigma_0(i))$$

Obviously, this is a conjugate prior p. d. f..

(4) The prior probabilities $p(M_1), \ldots, p(M_m)$ are available.

From these assumptions, we have, by (1.2), (4.1) and (4.2),

(4.3) $$p(X|M_i) = (2\pi)^{-\frac{(n-1)p}{2}} n^{-\frac{p}{2}} |\Sigma(i)|^{-\frac{n-1}{2}} \exp\{-\frac{1}{2}\mathrm{tr}(S\Sigma^{-1}(i))\}$$

$$\cdot \int_{R^p} n(\theta_i | \mu_0(i), \Sigma_0(i)) n(\bar{x} | \theta_i, \frac{1}{n}\Sigma(i)) d\theta_i$$

$$= c' |\Sigma(i)|^{-\frac{n-1}{2}} \exp\{-\frac{1}{2}\mathrm{tr}(S\Sigma^{-1}(i))\} n(\bar{x} | \mu_0(i), \Sigma_0(i) + \frac{1}{n}\Sigma(i))$$

$$= c |\Sigma(i)|^{-\frac{n-1}{2}} |\Sigma_1(i)|^{-\frac{1}{2}} \exp\{-\frac{1}{2}[\mathrm{tr}(\Sigma^{-1}(i)S) + (\bar{x} - \mu_0(i))\Sigma_1^{-1}(i)$$

$$\cdot (\bar{x} - \mu_0(i))']\}$$

where $c' = (2\pi)^{-\frac{(n-1)p}{2}} n^{-\frac{p}{2}}$, $c = (2\pi)^{-\frac{np}{2}} n^{-\frac{p}{2}}$

(4.4) $$\Sigma_1(i) = \Sigma_0(i) + \frac{1}{n}\Sigma(i)$$

Therefore, we have

(4.5) $$p(M_i | X) = \frac{p(M_i) p(X|M_i)}{\sum_{j=1}^{m} p(M_j) p(X|M_j)}$$

$$\propto p(M_i) |\Sigma(i)|^{-\frac{n-1}{2}} |\Sigma_1(i)|^{-\frac{1}{2}} \exp\{-\frac{1}{2}[\mathrm{tr}(\Sigma^{-1}(i)S) + (\bar{x} - \mu_0(i))$$

$$\cdot \Sigma_1^{-1}(i)(\bar{x} - \mu_0(i))']\} \qquad (i = 1, \ldots, m)$$

4.2. CASE II

Let $\tilde{X} = (\tilde{x}_1, \ldots, \tilde{x}_n)$ be the observed random variable and let $\mathfrak{X} = R^{np}$ be the sample space. The possible models $M_1, \ldots, M_m$ are specified as follows.

(1) M_i has the unknown parameter $(\mu_i, \tau_i) \in \Theta_i = R^p \times R_+$ and has the known parameter $\Sigma(i) \in S_p^+$, where S_p^+ is the set of all positive definite matrices of order p.

(2) For given (M_i, μ_i, τ_i), $\tilde{x}_1, \ldots, \tilde{x}_n$ are independent and $\tilde{x}_j \sim N(\mu_i, \tau_i^{-1}\Sigma(i))$,

$(j=1,\ldots, n)$. Thus, we have

(4.6) $$p(X|M_i, \mu_i, \tau_i)=\prod_{j=1}^{n} n(\boldsymbol{x}_j|\boldsymbol{\mu}_i, \tau_i^{-1}\Sigma(i))$$
$$=(2\pi)^{-\frac{np}{2}}|\Sigma(i)|^{-\frac{n}{2}}\tau_i^{\frac{np}{2}}\exp\{-\frac{\tau_i}{2}\sum_{j=1}^{n}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)\Sigma^{-1}(i)(\boldsymbol{x}_j-\boldsymbol{\mu}_i)'\}.$$

Since it holds that

$$\sum_{j=1}^{n}(\boldsymbol{x}_j-\boldsymbol{\mu}_i)\Sigma^{-1}(i)(\boldsymbol{x}_j-\boldsymbol{\mu}_i)'=tr(\Sigma^{-1}(i)S)+n(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_i)\Sigma^{-1}(i)(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_i)',$$

(4.6) is rewritten as

(4.6)' $$p(X|M_i, \boldsymbol{\mu}_i, \tau_i)=(2\pi)^{-\frac{(n-1)p}{2}}n^{-\frac{p}{2}}\tau_i^{\frac{(n-1)p}{2}}|\Sigma(i)|^{-\frac{n-1}{2}}\exp\{-\frac{\tau_i}{2}tr(\Sigma^{-1}(i)S)\}$$
$$\cdot(2\pi)^{-\frac{p}{2}}|\frac{1}{n\tau_i}\Sigma(i)|^{-\frac{1}{2}}\exp\{-\frac{n\tau_i}{2}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_i)\Sigma^{-1}(i)(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_i)'\}$$

where $\bar{\boldsymbol{x}}=\frac{1}{n}\sum_{j=1}^{n}\boldsymbol{x}_j$ and $S=\sum_{j=1}^{n}(\boldsymbol{x}_j-\bar{\boldsymbol{x}})'(\boldsymbol{x}_j-\bar{\boldsymbol{x}})$.

(3) The prior distribution of $(\tilde{\mu}_i, \tilde{\tau}_i)$, given M_i, is N-Gamma$(\boldsymbol{\mu}_0(i), \Sigma_0(i), \alpha_0(i), \beta_0(i))$, that is,

(4.7) $$\tilde{\boldsymbol{\mu}}_i|(\tilde{\tau}_i=\tau_i)\sim N(\mu_0(i), \tau_i^{-1}\Sigma_0(i))$$

(4.8) $$\tilde{\tau}_i\sim \mathrm{Gamma}(\alpha_0(i), \beta_0(i))$$

where $\boldsymbol{\mu}_0(i)$, $\Sigma_0(i)$, $\alpha_0(i)$ and $\beta_0(i)$ are known. This is a conjugate prior distribution as is easily recognized from (4.6)'.

(4) Prior probabilities $p(M_1),\ldots, p(M_m)$ are available.

From these assumptions we obtain

(4.9) $$P(X|M_i)$$
$$=\int_{R_+}\int_{R^p}p(\boldsymbol{\mu}_i, \tau_i)p(X|M_i, \boldsymbol{\mu}_i, \tau_i)d\boldsymbol{\mu}_i d\tau_i$$
$$=\int_{R_+}\int_{R^p}n(\boldsymbol{\mu}_i|\boldsymbol{\mu}_0(i), \tau_i^{-1}\Sigma_0(i))\frac{1}{\Gamma(\alpha_0(i))}(\beta_0(i))^{-\alpha_0(i)}\tau_i^{\alpha_0(i)-1}\exp(-\frac{\tau_i}{\beta_0(i)})$$
$$\cdot c'|\Sigma(i)|^{-\frac{n-1}{2}}\tau_i^{\frac{(n-1)p}{2}}\exp\{-\frac{\tau_i}{2}tr(\Sigma^{-1}(i)S)\}n(\bar{\boldsymbol{x}}|\boldsymbol{\mu}_i, \frac{1}{n\tau_i}\Sigma(i))d\boldsymbol{\mu}_i d\tau_i$$
$$=c'\frac{(\beta_0(i))^{-\alpha_0(i)}}{\Gamma(\alpha_0(i))}|\Sigma(i)|^{-\frac{n-1}{2}}\int_0^\infty\{\int_{R^p}n(\mu_i|\mu_0(i),\frac{1}{\tau_i}\Sigma_0(i))n(\bar{\boldsymbol{x}}|\boldsymbol{\mu}_i, \frac{1}{n\tau_i}\Sigma(i))d\boldsymbol{\mu}_i$$

$$\cdot \tau_i^{\alpha_0(i)+\frac{(n-1)p}{2}-1} \exp\{-\left[\frac{1}{\beta_0(i)}+\frac{1}{2}tr(\Sigma^{-1}(i)S)\right]\tau_i\} d\tau_i$$

$$= c' \frac{(\beta_0(i))^{-\alpha_0(i)}}{\Gamma(\alpha_0(i))} |\Sigma(i)|^{-\frac{n-1}{2}} \int_0^\infty n(\bar{\boldsymbol{x}}|\boldsymbol{\mu}_0(i), \frac{1}{\tau_i}\Sigma_1(i)) \tau_i^{\alpha_0(i)+\frac{(n-1)p}{2}-1}$$

$$\exp\{-\left[\frac{1}{\beta_0(i)}+\frac{1}{2}tr(\Sigma^{-1}(i)S)\right]\tau_i\} d\tau_i$$

$$= c \frac{(\beta_0(i))^{-\alpha_0(i)}}{\Gamma(\alpha_0(i))} |\Sigma(i)|^{-\frac{n-1}{2}} |\Sigma_1(i)|^{-\frac{1}{2}} \int_0^\infty \tau_i^{\alpha_0(i)+\frac{np}{2}-1} \exp\{-\left[\frac{1}{\beta_0(i)}\right.$$

$$\left.+\frac{1}{2}tr(\Sigma^{-1}(i)S)+\frac{1}{2}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0(i))\Sigma_1^{-1}(i)(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0(i))'\right]\tau_i\} d\tau_i$$

$$= c \frac{\Gamma(\alpha_0(i)+\frac{np}{2})}{\Gamma(\alpha_0(i))} \cdot \frac{(\beta_1(i))^{\alpha_0(i)+\frac{np}{2}}}{(\beta_0(i))^{\alpha_0(i)}} |\Sigma(i)|^{-\frac{n-1}{2}} |\Sigma_1(i)|^{-\frac{1}{2}}$$

where $c'=(2\pi)^{-\frac{(n-1)p}{2}} n^{-\frac{p}{2}}$, $c=(2\pi)^{-\frac{np}{2}} n^{-\frac{p}{2}}$,

$$(4.10) \qquad (\beta_1(i))^{-1}=(\beta_0(i))^{-1}+\frac{1}{2}tr(\Sigma^{-1}(i)S)+\frac{1}{2}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0(i))\Sigma_1^{-1}(i)(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0(i))'$$

$$(4.11) \qquad \Sigma_1(i)=\Sigma_0(i)+\frac{1}{n}\Sigma(i)$$

and, thus,

$$(4.12) \qquad p(M_i|X) \propto p(M_i) \frac{\Gamma(\alpha_0(i)+\frac{np}{2})(\beta_1(i))^{\alpha_0(i)+\frac{np}{2}}}{\Gamma(\alpha_0(i))(\beta_0(i))^{\alpha_0(i)}} |\Sigma(i)|^{-\frac{n-1}{2}} |\Sigma_1(i)|^{-\frac{1}{2}}$$

$$(i=1,\ldots, m)$$

4.3. CASE III

Let $\tilde{X}=(\tilde{\boldsymbol{x}}_1,\ldots, \tilde{\boldsymbol{x}}_n)$ be observed random variable and let $\mathcal{X}$ be R^{np}. The possible models $M_1,\ldots, M_m$ are specified in the following.

(1) M_i has the unknown parameter $(\boldsymbol{\mu}_i, \Sigma(i)) \in R^p \times S_p^+$.

(2) For given $(M_i, \boldsymbol{\mu}_i, \Sigma(i))$, $\tilde{\boldsymbol{x}}_1,\ldots, \tilde{\boldsymbol{x}}_n$ are independent and $\tilde{\boldsymbol{x}}_j \sim N(\boldsymbol{\mu}_i, \Sigma(i))$, $(j=1,\ldots, m)$. Thus, like (4.1) or (4.6)',

$$(4.13) \qquad p(X|M_i, \boldsymbol{\mu}_i, \Sigma(i)) = \prod_{j=1}^{n} n(\boldsymbol{x}_j|\boldsymbol{\mu}_i, \Sigma(i))$$

$$= c'|\Sigma(i)|^{-\frac{n-1}{2}} \exp\{-\frac{1}{2}tr(S\,\Sigma^{-1}(i))\}$$

$$\cdot n(\bar{\mathbf{x}}|\mu_i, \frac{1}{n}\Sigma(i))$$

where $c'=(2\pi)^{-\frac{(n-1)p}{2}} n^{-\frac{p}{2}}$, $\bar{\mathbf{x}}=\frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_j$, $S=\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})'(\mathbf{x}_j-\bar{\mathbf{x}})$.

(3) The prior distribution of $(\tilde{\mu}_i, \tilde{\Sigma}^{-1}(i))$, given M_i, is the normal-Wishart distribution, N-Wishart $(\mu_0(i), \tau_0(i), \Sigma_0^{-1}(i), p, \nu_0(i))$, with $\nu_0(i)>p-1$, that is,

(4.14) $$\tilde{\mu}_i|(\tilde{\Sigma}(i)=\Sigma(i)) \sim N(\mu_0(i), \tau_0^{-1}(i)\Sigma(i))$$

$$\Sigma^{-1}(i) \sim \text{Wishart}(\Sigma_0^{-1}(i), p, \nu_0(i))$$

(4) The prior probabilities $p(M_1),\ldots, p(M_m)$ are available.

From these assumptions we have

(4.15) $$p(X|M_i)=\int_{S_p^+}\int_{R^p} p(\mu_i, \Sigma^{-1}(i))p(X|M_i, \mu_i, \Sigma^{-1}(i))d\mu_i d\Sigma^{-1}(i)$$

$$=c'\int_{S_p^+} w(\Sigma^{-1}(i)|\Sigma_0^{-1}(i), p, \nu_0(i))|\Sigma(i)|^{-\frac{n-1}{2}}\exp\{-\tfrac{1}{2}\mathrm{tr}(S\Sigma^{-1}(i))\}$$

$$\cdot\int_{R^p} n(\mu_i|\mu_0(i), \tau_0^{-1}(i)\Sigma(i))n(\bar{x}|\mu_i, \tfrac{1}{n}\Sigma(i))d\mu_i d\Sigma^{-1}(i)$$

$$=c'\int_{S_p^+} w(\Sigma^{-1}(i)|\Sigma_0^{-1}(i), p, \nu_0(i))|\Sigma(i)|^{-\frac{n-1}{2}}\exp\{-\tfrac{1}{2}\mathrm{tr}(S\Sigma^{-1}(i))\}$$

$$\cdot n(\bar{x}|\mu_0(i), (\frac{1}{\tau_0(i)}+\frac{1}{n})\Sigma(i))d\Sigma^{-1}(i)$$

$$=cC(p, \nu_0(i))(\frac{1}{\tau_0(i)}+\frac{1}{n})^{-\frac{p}{2}}|\Sigma_0(i)|^{\frac{\nu_0(i)}{2}}$$

$$\cdot\int_{S_p^+}|\Sigma^{-1}(i)|^{\frac{\nu_1(i)-p-1}{2}}\exp\{-\tfrac{1}{2}\mathrm{tr}(\Sigma_1(i)\Sigma^{-1}(i))\}$$

$$=c\frac{C(p, \nu_0(i))}{C(p, \nu_1(i))}\left(\frac{n\tau_0(i)}{\tau_0(i)+n}\right)^{\frac{p}{2}}|\Sigma_0(i)|^{\frac{\nu_0(i)}{2}}|\Sigma_1(i)|^{-\frac{\nu_1(i)}{2}}$$

where $w(\cdot|\Sigma, p, \nu)$ is the p. d. f. of the Whishart distribution with the parameter (Σ, p, ν) and

(4.16) $$c'=(2\pi)^{-\frac{(n-1)p}{2}} n^{-\frac{p}{2}}, \qquad c=(2\pi)^{-\frac{np}{2}} n^{-\frac{p}{2}},$$

(4.17) $$C(p,\ \nu)=2^{\frac{p\nu}{2}} \pi^{\frac{p(p-1)}{4}} \prod_{j=1}^{p} \left(\frac{\nu+1-j}{2}\right)$$

(4.18) $$\nu_1(i)=\nu_0(i)+n$$

(4.19) $$\Sigma_1(i)=\Sigma_0(i)+S+\frac{n\tau_0(i)}{\tau_0(i)+n}(\bar{\mathbf{x}}-\mu_0(i))'(\bar{\mathbf{x}}-\mu_0(i))$$

Therefore, for the observation $\tilde{X}=X$, we have

(4.20) $$p(M_i|X)\propto p(M_i)\frac{C(p,\ \nu_0(i))}{C(p,\ \nu_1(i))}\left(\frac{\tau_0(i)}{\tau_0(i)+n}\right)^{\frac{p}{2}}|\Sigma_0(i)|^{\frac{\nu_0(i)}{2}}\,|\Sigma_1(i)|^{-\frac{\nu_1(i)}{2}}$$

(i=1,..., m)

Since

(4.21) $$\frac{C(p,\ \nu_0)}{C(p,\ \nu_1)}=2^{\frac{p(\nu_1-\nu_0)}{2}} \prod_{j=1}^{p} \frac{\Gamma((\nu_1+1-j)/2)}{\Gamma((\nu_0+1-j)/2)}$$

and $\nu_1(i)-\nu_0(i)=n$, we have

(4.22) $$\frac{C(p,\ \nu_0(i))}{C(p,\ \nu_1(i))}=2^{\frac{np}{2}} \prod_{j=1}^{p} \frac{\Gamma((\nu_1(i)+1-j)/2)}{\Gamma((\nu_0(i)+1-j)/2)}$$ (i=1,..., m)

Hence we obtain

(4.23) $$p(M_i|X)\propto p(M_i)C(i)\left(\frac{\tau_0(i)}{\tau_0(i)+n}\right)^{\frac{p}{2}}|\Sigma_0(i)|^{\frac{\nu_0(i)}{2}}\,|\Sigma_1(i)|^{-\frac{\nu_1(i)}{2}}$$

(i=1,..., m)

where

(4.24) $$C(i)=\prod_{j=1}^{p} \frac{\Gamma((\nu_1(i)+1-j)/2)}{\Gamma((\nu_0(i)+1-j)/2)}$$ (i=1,..., m)

REMARK 4.1. In the above specification, the difference among models $M_1,\ldots,$ M_m is solely due to the assumption (3): the specification of the prior distribution of $(\tilde{\mu}_i,\ \tilde{\Sigma}^{-1}(i))$, (i=1,..., m). Therefore, our selection problem is nothing but the selection among prior distributions on the basis of the observation $\tilde{X}=X$.

5. SELECTION OF REGRESSORS IN A LINEAR REGRESSION MODEL IN CASE OF INFORMATIVE PRIOR INFORMATION

Assume that $x_1,\ldots, x_m$ is the set of possible regressors for a regression model. That is, the maximal model is

(5.1) $$\tilde{y}=\beta_0+\sum_{i=1}^{m}\beta_i x_i+\tilde{\varepsilon}$$

When the observations on $\tilde{y}$ are considered for the explanatory variables $(x_{1j},\ldots, x_{mj})$ $(j=1,\ldots, n)$, we have, from (5.1),

(5.2) $$\tilde{\boldsymbol{y}}=\boldsymbol{\beta}X+\tilde{\boldsymbol{\varepsilon}}$$

where

$$\boldsymbol{y}=(y_1,\ldots, y_n),\ \boldsymbol{\beta}=(\beta_0,\ \beta_1,\ldots,\ \beta_m),\ \boldsymbol{\varepsilon}=(\varepsilon_1,\ldots,\ \varepsilon_n)$$

(5.3) $$X=\begin{pmatrix}\mathbf{1}\\ \boldsymbol{x}_1\\ \vdots\\ \boldsymbol{x}_m\end{pmatrix},\qquad \mathbf{1}=(1,\ldots,\ 1)\in\mathbb{R}^n,\qquad \boldsymbol{x}_i=(x_{i1},\ldots,\ x_{in})\qquad (i=1,\ldots,\ m)$$

Now, let us introduce the concept of selection matrix which seems to be useful to describe our problem.

DEFINITION For any integer k $(0\leq k\leq m)$ and any set of k integers $\{i_1,\ i_2,\ldots, i_k\}$ $(1\leq i_1<i_2\ldots<i_k\leq m)$, we define the $(k+1,\ m+1)$ matrix $D(i_1,\ldots,\ i_k)$ by

(5.4) $$D(i_1,\ldots,\ i_k)=\begin{pmatrix}\boldsymbol{e}_1\\ \boldsymbol{e}_{i_1+1}\\ \vdots\\ \boldsymbol{e}_{i_k+1}\end{pmatrix};\qquad D(\phi)=\boldsymbol{e}_1\ \text{ for } k=0$$

and call it a selection matrix, where $\boldsymbol{e}_i$ is the unit vector of $\mathbb{R}^{m+1}$ of which the i-th component is $1(i=1,\ldots,\ m+1)$. We denote

$$\mathbb{D}=\{D(i_1,\ldots,\ i_k)\,|\,1\leq i_1<\ldots<i_k\leq m;\ k=1,\ldots,\ m\}\cup\{D(\phi)\}$$

REMARK 5.1. Clearly, $D(1,\ldots,\ m)=I_{m+1}$(the unit matrix of order $m+1$).

The regression model corresponding to the set of k explanatory variables $x_{i_1},\ldots, x_{i_k}$ is written as

$$(5.5) \qquad \tilde{\mathbf{y}}=(\beta_0, \beta_{i_1},\ldots, \beta_{i_k})\begin{pmatrix}1\\ x_{i_1}\\ \vdots\\ x_{i_k}\end{pmatrix}+\tilde{\boldsymbol{\varepsilon}} \;;\quad \tilde{\mathbf{y}}=\beta_0 1+\tilde{\boldsymbol{\varepsilon}} \qquad \text{for } k=0$$

This is rewritten as

$$(5.6) \qquad \tilde{\mathbf{y}}=\beta D' DX+\tilde{\varepsilon} \qquad (D=D(i_1,\ldots, i_k) \text{ or } D(\phi))$$

Moreover, if we denote $\boldsymbol{\beta} D'=\boldsymbol{\beta}_D$, using the notation D as a subscript,

$$(5.7) \qquad \tilde{\mathbf{y}}=\tilde{\boldsymbol{\beta}}_D DX+\tilde{\boldsymbol{\varepsilon}}_D \qquad (D\in\mathbf{D})$$

Since the correspondence between the selection $\{x_{i},\ldots, x_{i_k}\}$and the selection matrix $D(i_1,\ldots, i_k)$ is one to one, the regression model (5.7) will be called model D simply and conveniently.

In the following our attention is confined to the case where

$$(5.8) \qquad \text{rank } X=m+1$$

and, given $\tau_D\in R_+$,

$$(5.9) \qquad \tilde{\boldsymbol{\varepsilon}}_D \sim N(\mathbf{0}, \tau_D^{-1} I_n) \qquad (D\in\mathbf{D})$$

Now, let us specify each model D ($\in\mathbf{D}$) as follows.

(1) Model D has the unknown parameter $(\boldsymbol{\beta}_D, \tau_D)\in R^{r(D)}\times R_+$, where r(D) is the rank of matrix D.

(2) For given D, X and $(\boldsymbol{\beta}_D, \tau_D)$, we have, by (5.7) and (5.9),

$$(5.10) \qquad \tilde{y} \sim N(\boldsymbol{\beta}_D DX, \tau_D^{-1} I_n)$$

that is,

$$(5.10)' \qquad p(\mathbf{y}|D, X, \boldsymbol{\beta}_D, \tau_D)=n(\mathbf{y}|\boldsymbol{\beta}_D DX, \tau_D^{-1} I_n)$$

(3) The prior distribution of $(\tilde{\boldsymbol{\beta}}_D, \tilde{\tau}_D)$ in the model D is the normal-gamma distribution with (known) parameters $(\mu_0(D), \Sigma_0(D), \alpha_0(D), \beta_0(D))$, that is,

$$(5.11) \qquad (\tilde{\boldsymbol{\beta}}_D, \tilde{\tau}_D) \sim \text{N-Gamma}(\mu_0(D), \Sigma_0(D), \alpha_0(D), \beta_0(D))$$

or

$$(5.12) \qquad \tilde{\boldsymbol{\beta}}_D | (\tilde{\tau}_D=\tilde{\tau}_D) \sim N(\mu_0(D), \tau_D^{-1}\Sigma_0(D))$$

(5.13) $$\tilde{\tau}_D \sim \text{Gamma}(\alpha_0(D),\ \beta_0(D))$$

(4) The prior probability of model D is denoted by p(D): $\sum_{D\in\mathbf{D}} p(D)=1$. If $\mathbf{D}_0=\{D|p(D)>0,\ D\in\mathbf{D}\}\subsetneqq\mathbf{D}$, we can confine our attention on $\mathbf{D}_0$.

From the above specifications, we easily obtain

(5.14) $$p(\mathbf{y}|D,\ X,\ \tau_D)=\int_{R^{r(D)}} n(\boldsymbol{\beta}_D|\boldsymbol{\mu}_0(D),\ \tau_D^{-1}\ \Sigma_0(D))n(\mathbf{y}|\boldsymbol{\beta}_D DX,\ \tau_D^{-1}I_n)d\beta_D$$
$$=n(\mathbf{y}|\boldsymbol{\mu}_0(D)DX,\ \tau_D^{-1}S^{-1}(D))$$

where

(5.15) $$S(D)=I_n-X'D'H^{-1}(D)DX,\ H(D)=\Sigma_0^{-1}(D)+DXX'D'$$

Further, from (5.13) and (5.14), we can easily obtain

(5.16) $$p(\mathbf{y}|D,\ X)=\int_0^\infty p(\tau_D)p(\mathbf{y}|D,\ X,\ \tau_D)d\tau_D$$
$$=\int_0^\infty \text{gamma}(\tau_D|\alpha_0(D),\ \beta_0(D))n(\mathbf{y}|\boldsymbol{\mu}_0(D)DX,\ \tau_D^{-1}S^{-1}(D))d\tau_D$$
$$=\pi^{-\frac{n}{2}}(2\alpha_0(D))^{-\frac{n}{2}}\frac{\Gamma(\alpha_0(D)+\frac{n}{2})}{\Gamma(\alpha_0(D))}|\alpha_0(D)\beta_0(D)S(D)|^{\frac{1}{2}}$$
$$\cdot(1+\frac{1}{2\alpha_0(D)}(\mathbf{y}-\boldsymbol{\mu}_0(D)DX)(\alpha_0(D)\beta_0(D)S(D))(\mathbf{y}-\boldsymbol{\mu}_0(D)DX)')^{-\frac{2\alpha_0(D)+n}{2}}$$

That is, for given (D, X), $\tilde{\mathbf{y}}$ follows a n-dimensional t-distribution with D. F. $2\alpha_0(D)$, i. e.

(5.17) $$\tilde{\mathbf{y}}|(D,\ X)\sim t(\boldsymbol{\mu}_0(D)DX,\ (\alpha_0(D)\beta_0(D)S(D))^{-1},\ n,\ 2\alpha_0(D))$$

Thus, we have the posterior probability of model D for the observed data (X, **y**) from (5.16) and

(5.18) $$p(D|X,\ \mathbf{y})=\frac{p(D)p(\mathbf{y}|D,\ X)}{\sum_{C\in\mathbf{D}} p(C)p(\mathbf{y}|C,\ X)}$$
$$\propto p(D)(\alpha_0(D))^{-\frac{n}{2}}\frac{\Gamma(\alpha_0(D)+\frac{n}{2})}{\Gamma(\alpha_0(D))}(\alpha_0(D)\beta_0(D))^{\frac{n}{2}}|S(D)|^{\frac{1}{2}}$$
$$\cdot(1+\frac{\beta_0(D)}{2}(\mathbf{y}-\boldsymbol{\mu}_0(D)DX)S(D)(\mathbf{y}-\boldsymbol{\mu}_0(D)DX)')^{-\frac{2\alpha_0(D)-n}{2}}$$

$$\propto p(D)\frac{\Gamma(\alpha_0(D)+\frac{n}{2})}{\Gamma(\alpha_0(D))}(\beta_0(D))^{\frac{n}{2}}|S(D)|^{\frac{1}{2}}$$

$$\cdot(1+\frac{\beta_0(D)}{2}\|y-\mu_0(D)DX\|^2_{S(D)})^{-\frac{2\alpha_0(D)+n}{2}}$$

REMARK 5.2. The above formulation of selection of regressors has some difficulties. Among others, it must be very laborious to specify the prior distribution of $(\tilde{\beta}_0, \tilde{\tau}_D)$ for each $D\in\mathbf{D}$, unless $\mathbf{D}_0=\{D|p(D)>0, D\in\mathbf{D}\}$ is reasonably small.

6. SELECTION OF REGRESSORS IN A LINEAR REGRESSION MODEL IN CASE OF VAGUE PRIOR INFORMATION

In this section we will treat the same problem as in the preceding section in the case where we are almost ignorant of $\boldsymbol{\beta}_D$ and τ_D for each model $D\in\mathbf{D}$.

To express our state of ignorance, we tacitly introduce an intermediate (proper) prior distribution and then consider its limit to obtain the posterior probabilities $p(D|X, \mathbf{y})$ for vague prior information.

DEFINITION An intermediate prior distribution of $(\tilde{\boldsymbol{\beta}}_D, \tilde{\tau}_D)$ is defined for parameters a and b $(a>0, b>0)$ as follows.

(1) $\tilde{\boldsymbol{\beta}}_D$ and $\tilde{\tau}_D$ are independent $(D\in\mathbf{D})$.

(2) $p(\boldsymbol{\beta}_D|D, a)=\frac{1}{2a}I(\boldsymbol{\beta}_D|A_{r(D)}(a_{r(D)}))$

where

$$A_{r(D)}(c)=(-c, c)\times\dots\times(-c, c)\subset\mathbf{R}^{r(D)} \qquad (c>0)$$

that is, it is a $r(D)$-dimensional product set of interval $(-c, c)$, and $I(\cdot|B)$ is the indicater function of set B. In addition, $a_{r(D)}$ is given by

$$(2a_{r(D)})^{r(D)}=2a \text{ or } a_{r(D)}=\frac{1}{2}(2a)^{\frac{1}{r(D)}}$$

The prior distribution of $\boldsymbol{\beta}_D$, given D and a, is the uniform distribution on the r(D)-dimensional interval $A_{r(D)}(a_{r(D)})$.

(3) $$p(\tau_D|D,\ b)=\frac{1}{2b\tau_D}I(\tau_D|(e^{-b},\ e^{b}))$$

that is,

$$\log\tilde{\tau}_D \sim U(-b,\ b) \qquad \text{(uniform distribution on } (-b,\ b))$$

Thus, the intermediate prior p. d. f. of $(\hat{\beta}_D,\ \hat{\tau}_D)$ for a and b, is

(6.1) $$p(\beta_D,\ \tau_D|D,\ a,\ b)=p(\beta_D|D,\ a)p(\tau_D|D,\ b)$$

$$=\frac{1}{4ab\tau_D}I(\beta_D|A_{r(D)}(a_{r(D)}))I(\tau_D|(e^{-b},\ e^{b}))$$

For this intermediate prior p. d. f. (6.1), we obtain the p. d. f. of $\tilde{\mathbf{y}}$, given D, X, a and b, and then the posterior probabilities of model D ($D \in \mathbf{D}$), given X, $\mathbf{y}$, a and b, as follows.

(6.2) $$p(\mathbf{y}|D,\ X,\ a,\ b)=\int_0^\infty\int_{R^{r(D)}}p(\beta_D|D,\ a)p(\tau_D|D,\ b)p(\mathbf{y}|D,\ X,\ \beta_D,\ \tau_D)d\beta_D d\tau_D$$

$$=\frac{1}{4ab}q(\mathbf{y}|D,\ X,\ a,\ b)$$

where

(6.3) $$q(\mathbf{y}|D,\ X,\ a,\ b)=\int_{e^{-b}}^{e^{b}}\int_{A_{r(D)}(a_{r(D)})}\frac{1}{\tau_D}n(\mathbf{y}|\beta_D DX,\ \tau_D^{-1}I_n)d\beta_D d\tau_D$$

(6.4) $$p(D|X,\ \mathbf{y},\ a,\ b)=\frac{p(D)p(\mathbf{y}|D,\ X,\ a,\ b)}{\sum_{C\in\mathbf{D}} p(C)p(\mathbf{y}|C,\ X,\ a,\ b)}$$

$$=\frac{p(D)(\mathbf{y}|D,\ X,\ a,\ b)}{\sum_{C\in\mathbf{D}} p(C)q(\mathbf{y}|C,\ X,\ a,\ b)}$$

Now, in order to obtain the posterior probability of model D for vague prior information, we consider the limits of $q(\mathbf{y}|D,\ X,\ a,\ b)$ and $p(D|X,\ \mathbf{y},\ a,\ b)$ when $a\to\infty$ and $b\to\infty$. Thus,

(6.5) $$q(\mathbf{y}|D,\ X)=\lim_{a,b\to\infty} q(\mathbf{y}|D,\ X,\ a,\ b)$$

$$=\int_0^\infty\int_{R^{r(D)}}\frac{1}{\tau_D}n(\mathbf{y}|\beta_D DX,\ \tau_D^{-1}I_n)d\beta_D d\tau_D$$

$$=(2\pi)^{-\frac{n}{2}}\int_0^\infty \tau_D^{\frac{n}{2}-1}\int_{\mathbb{R}^{r(D)}}\exp\{-\frac{\tau_D}{2}(\mathbf{y}-\beta_D DX)(\mathbf{y}-\beta_D DX)'\}d\beta_D d\tau_D$$

It is easily shown that

(6.6) $$(\mathbf{y}-\beta_D DX)(\mathbf{y}-\beta_D DX)'=(\beta_D-\hat{\beta}_D)(DXX'D')(\beta_D-\hat{\beta}_D)'+\|y-\hat{\beta}_D DX\|^2$$

where

(6.7) $$\hat{\beta}_D=\mathbf{y}X'D'(DXX'D')^{-1}$$

Hence, we have

(6.8) $$q(y\ D,\ X)=(2\pi)^{-\frac{n}{2}}\int_0^\infty \tau_D^{\frac{n}{2}-1}\exp(-\frac{\tau_D}{2}\|\mathbf{y}-\hat{\beta}_D DX\|^2)\int_{\mathbb{R}^{r(D)}}\exp\{-\frac{\tau_D}{2}(\beta_D-\hat{\beta}_D)$$

$$(DXX'D')(\beta_D-\hat{\beta}_D)'\}d\beta_D d\tau_0$$

$$=(2\pi)^{-\frac{n-r(D)}{2}}\int_0^\infty \tau_D^{\frac{n}{2}-1}|\tau_D(DXX'D')|^{-\frac{1}{2}}\exp(-\frac{\tau_D}{2}\|\mathbf{y}-\hat{\beta}_D DX\|^2)d\tau_D$$

$$=(2\pi)^{-\frac{n-r(D)}{2}}|DXX'X'|^{-\frac{1}{2}}\int_0^\infty \tau_D^{\frac{n-r(D)}{2}-1}\exp(-\frac{\tau_D}{2}\|\mathbf{y}-\hat{\beta}_D DX\|^2)d\tau_D$$

$$=\pi^{-\frac{n-r(D)}{2}}\Gamma(\frac{n-r(D)}{2})|DXX'D'|^{-\frac{1}{2}}\|y-\beta_D DX\|^{-(n-r(D))}$$

and thus

(6.9) $$p(D|X,\ \mathbf{y})=\lim p(D|X,\ \mathbf{y},\ a,\ b)$$

$$=\frac{p(D)q(y|D,\ X)}{\sum_{C\ D} p(C)q(y|C,\ X)}$$

$$=\frac{p(D)\pi^{\frac{r(D)}{2}}\Gamma(\frac{n-r(D)}{2})|DXX'D'|^{-\frac{1}{2}}\|y-\hat{\beta}_D DX\|^{-(n-r(D))}}{\sum_{C\in D} p(C)\pi^{\frac{r(C)}{2}}\Gamma(\frac{n-r(C)}{2})|CXX'C'|^{-\frac{1}{2}}\|y-\hat{\beta}_C CX\|^{-(n-r(C))}}$$

or

(6.10) $$p(D|X,\ y)\propto p(D)\pi^{\frac{r(D)}{2}}\Gamma(\frac{n-r(D)}{2})|DXX'D'|^{-\frac{1}{2}}\|y-\hat{\beta}_D DX\|^{-(n-r(D))}$$

$(D\in D)$

REMARK 6.1. The dimension $\dim(\Theta_D)$ of the parameter space Θ_D of model D depends on model D ($\in\mathbb{D}$), since

$$\dim(\Theta_D)=\dim(\mathbb{R}^{r(D)}\times R_+)=r(D)+1$$

Our intermediate prior p. d. f. (6.1) (in particular, (2) of DEFINITION) is devised to make the degree of uncertainty in each model D the same in spite of the difference of $\dim(\Theta_D)$. For instance, if we express the degree of uncertainty by the concept of entropy,

(6.11) the entropy of $p(\boldsymbol{\beta}_D, \tau_D|D, a, b)$

$$=-\int_{\Theta_D}(\log p(\boldsymbol{\beta}_D, \tau_D|D, a, b))p(\boldsymbol{\beta}_D, \tau_D|D, a, b)d\boldsymbol{\beta}_D d\tau_D$$

$$=-\int_{\mathbb{R}^{r(D)}}(\log p(\boldsymbol{\beta}_D|D, a))p(\boldsymbol{\beta}_D|D, a)d\boldsymbol{\beta}_D$$

$$-\int_0^\infty(\log p(\tau_D|D, b))p(\tau_D|D, b)d\tau_D$$

$$=\log(2a)+\log(2b) \qquad (D\in\mathbb{D})$$

which is independent of $D\in\mathbb{D}$.

7. SELECTION OF THE ORDER OF POLYNOMIAL REGRESSIONS IN CASE OF VAGUE PRIOR INFORMATION

Let the following polynomial regression models $M_0, M_1, \ldots, M_m$ be the candidates of our selection problem:

(7.1) $$M_k: \quad \tilde{y}_i=\sum_{j=0}^{k}\beta_j x_i^j+\tilde{\varepsilon}_i \qquad (i=1,\ldots, n)$$

or

(7.2) $$M_k: \quad \tilde{\mathbf{y}}=\boldsymbol{\beta}_k X_k+\tilde{\boldsymbol{\varepsilon}} \qquad (k=0,\ldots, m)$$

where

$$\mathbf{y}=(y_1,\ldots, y_n),\ \boldsymbol{\beta}_k=(\beta_0, \beta_1,\ldots, \beta_k),\ \boldsymbol{\varepsilon}=(\varepsilon_1,\ldots, \varepsilon_n)$$

and

(7.3) $$X_k=\begin{pmatrix}\mathbf{1}\\ \mathbf{x}_1\\ \vdots\\ \mathbf{x}_k\end{pmatrix} \qquad \mathbf{x}_j=(x_1^j,\ldots, x_n^j) \quad (j=1,\ldots, m) \qquad ; X_0=\mathbf{1}$$

As in the preceding section, model M_k is written with selection matrix D=D(1, ..., k) for k=1,..., m and D=D(ϕ) for k=0 as follows:

(7.4) $$\tilde{y}=\beta D'DX+\tilde{\varepsilon}_D$$

or

(7.5) $$\tilde{y}=\beta_D DX+\tilde{\varepsilon}_D \qquad (\beta_D=\beta D')$$

where $\beta\equiv\beta_m$, $X\equiv X_m$. Hence, the model M_k can be regarded as the model D(1,..., k) and thus the selection among M_0, M_1,..., M_m can be regarded as the selection from $\mathbb{D}_0$:

(7.6) $$\mathbb{D}_0=\{D(\phi),\ D(1),\ D(1,\ 2),\ldots,\ D(1,\ 2,\ldots,\ m)\}$$

Therefore, our selection problem in case of vague prior in each model is nothing but a special case of the selection problem which was treated in the preceding section.

REFERENCES:

[1] Suzauki, Y. (1983). On Bayesian approach to model selection, ISI Contributed Papers, Madrid 1983, Vol. 1, 288-291.

[2] Suzuki, Y. (1987). Theory of Statistics, Asakura Publishing Company. (in Japanese).