Ann. Inst. Statist. Math.
37 (1985), Part B, 369-387

BAYESIAN BINARY REGRESSION INVOLVING TWO EXPLANATORY VARIABLES

YOSIYUKI SAKAMOTO AND MAKIO ISHIGURO

(Received Apr. 9, 1984; revised May 31, 1985)

Summary

The purpose of the present paper is to propose a practical Bayesian procedure for the estimation of binary response probability where the explanatory variable is bivariate. The procedure is an extension of the procedure for univariate case which was proposed by the present authors [2] and is based on a model which approximates the logistic transformation of response probability by a quadratic orthogonal spline function on the two-dimensional space of explanatory variable. The flexibility of the model is guaranteed by assuming a spline function on sufficiently fine mesh. To obtain stable estimates we introduce a prior distribution of the parameters of the model. The prior distribution has several parameters (hyper-parameters) which are chosen to minimize an Bayesian information criterion ABIC. The procedure is applicable to cases where each explanatory variable takes continuous values provided that the probability of the occurrence changes smoothly. The practical utility of the procedure is demonstrated by examples of applications to five sets of data.

1. Introduction

The most important problem of the analysis of binary data is to study how the probability of the occurrence of a certain phenomenon depends on explanatory variables. The present authors recently proposed in [2] a Bayesian method to estimate the probability which depends on the value of a single explanatory variable. The method has a wide range of applications and the typical examples are the dose response curve estimation, the estimation of the intensity function in the point process analysis, and the analysis of public opinion poll data. It also suggests a new approach to the discriminant analysis.

The purpose of the present paper is to generalize the Bayesian method to the case where the explanatory variable is bivariate. We assume that the probability of the occurrence is a smooth function of the explanatory variable. For the convenience of readers, we will briefly review the procedure proposed in [2] in Section 2. Our Bayesian model is proposed in Section 3 and a numerical procedure is described in Section 4. The practical utility of the present procedure is demonstrated in Section 5 by examples of applications to five sets of data.

2. Review of the Bayesian binary regression model in univariate case

To estimate the conditional probability $p^*(x)$ of the occurrence of a certain phenomenon given a value of an explanatory variable x, we assumed in [2] that $p^*(x)$ can be approximated by a piece-wise constant function, or the 0-th order spline function, $p(x)$ defined by

$$(1)\qquad p(x)=\exp q_j/(1+\exp q_j) \qquad \text{if } a_{j-1}<x\leqq a_j,\ j=1,\cdots,c\,,$$

where $a_0<a_1<\cdots<a_c$ are closely set segment points. Given a set of data (h_i, x_i), $i=1,\cdots,n$, where h_i is 1 or 0 in accordance with the occurrence or non-occurrence of the phenomenon and x_i is a value of the explanatory variable, the likelihood of the parameter vector $\boldsymbol{q}=(q_1,\cdots, q_c)^t$ of the model (1) is given by

$$(2)\qquad L(\boldsymbol{q})=\prod_{j=1}^{c}\left(\frac{\exp q_j}{1+\exp q_j}\right)^{n(1,j)}\left(\frac{1}{1+\exp q_j}\right)^{n(0,j)},$$

where $n(k, j)$ denotes the number of data of satisfying both $h_i=k$ ($k=0,1$) and $a_{j-1}<x_i\leqq a_j$. This likelihood and a prior distribution defined by

$$(3)\qquad \pi(\boldsymbol{q}|v^2,\boldsymbol{q}_0)=\frac{1}{\sqrt{2\pi}\,v}\exp\left\{-\frac{1}{2v^2}|D(\boldsymbol{q}-\boldsymbol{q}_0)|^2\right\}$$

of the parameter vector $\boldsymbol{q}$ constitute our Bayesian model. This prior distribution is introduced to guarantee the smoothness of the estimates of $\{q_j\}$ with respect to j. Here we used the notation

$$(4)\qquad D=\begin{pmatrix} 1 & & & & & & 0\\ -2 & 1 & & & & & \\ 1 & -2 & 1 & & & & \\ & \ddots & \ddots & \ddots & & & \\ & & \ddots & \ddots & \ddots & & \\ 0 & & & \ddots & \ddots & \ddots & \\ & & & & 1 & -2 & 1 \end{pmatrix}$$

and

$$(5)\qquad D\boldsymbol{q}_0=(-q_{-1}+2q_0,\ -q_0,\ 0,\cdots,0)^t\,,$$

where v^2 and $\boldsymbol{q}_0$ are adjustable parameters (hereafter referred as hyper-parameters) and are chosen so that they minimize the statistic ABIC defined by

$$\text{ABIC} = -2\log \int L(\boldsymbol{q})\pi(\boldsymbol{q}\,|\,v^2, \boldsymbol{q}_0)d\boldsymbol{q}\,, \tag{6}$$

which was introduced by Akaike [1]. Once the values of v^2 and $\boldsymbol{q}_0$ are specified, the estimate of $\boldsymbol{q}$ is given by the value of $\boldsymbol{q}$ which maximizes $L(\boldsymbol{q})\pi(\boldsymbol{q}\,|\,v^2, \boldsymbol{q}_0)$, that is, by the mode of the posterior distribution of $\boldsymbol{q}$.

3. A Bayesian binary regression model in bivariate case

A direct generalization of the procedure reviewed in the preceding section to bivariate case is difficult because of the following reason. In the univariate case, we have to set the value of c greater than 50 to obtain a practical approximation of $p(x)$ to $p^*(x)$ and have to carry out the maximization of $L(\boldsymbol{q})\pi(\boldsymbol{q}\,|\,v^2, \boldsymbol{q}_0)$ in c-dimensional space [2]. It follows that if we adopt the piece-wise constant structure of response probability on two-dimensional, say x-y, plane, then we will have to carry out a numerical maximization in 50^2 dimensional space or more. It is difficult and impractical approach.

We propose, instead, the use of two-dimensional orthogonal spline functions for the expression of the logistic transformation of the probability function, or the following expression:

$$p(x, y\,|\,\tilde{\boldsymbol{q}}) = \frac{\exp\left\{\sum_{j=1}^{J}\sum_{k=1}^{K} q_{jk}s_j(x)s_k(y)\right\}}{1+\exp\left\{\sum_{j=1}^{J}\sum_{k=1}^{K} q_{jk}s_j(x)s_k(y)\right\}} \tag{7}$$

for the probability function itself. Here $\tilde{\boldsymbol{q}} = (q_{11}, q_{12}, \cdots, q_{1K}, q_{21}, \cdots, q_{JK})^t$ and $s_j(x)$ and $s_k(y)$ are assumed to be the basis of the quadratic spline functions with suitably fixed segment points. And q_{jk} is the coefficient of the term $s_j(x)s_k(y)$.

When data (h_i, x_i, y_i), $h_i = 0, 1$, $i = 1, 2, \cdots, n$ are given, the likelihood of the model (7) is given by

$$L(\tilde{\boldsymbol{q}}) = \prod_{i=1}^{n} p(x_i, y_i\,|\,\tilde{\boldsymbol{q}})^{h_i}\{1-p(x_i, y_i\,|\,\tilde{\boldsymbol{q}})\}^{1-h_i}\,. \tag{8}$$

Note that J and K should be set large enough to guarantee the flexibility of the model (7). The smoothness of the estimate of $\tilde{\boldsymbol{q}}$ is guaranteed by assuming the prior distribution

$$\pi(\boldsymbol{q}\,|\,q_{11}, q_{12}, q_{21}, w^2) = \left(\frac{1}{\sqrt{2\pi}}\right)^{JK-3}(\det D^tD/w^2)^{1/2}\exp\left\{-\frac{1}{2w^2}|D\boldsymbol{q}+\boldsymbol{r}|^2\right\} \tag{9}$$

for the main part

(10) $$\boldsymbol{q}=(q_{13}, q_{14}, \cdots, q_{1K}, q_{22}, q_{23}, \cdots, q_{2K}, q_{31}, q_{32}, \cdots, q_{JK})^t$$

of the parameter vector $\tilde{\boldsymbol{q}}$. Here $|\ \ |$ denotes the Euclidean norm, and the matrix D and the vector $\boldsymbol{r}$ are defined by

(11) $$D\boldsymbol{q}=\begin{pmatrix} \boldsymbol{q}_x \\ \boldsymbol{q}_y \\ \boldsymbol{q}_{xy} \end{pmatrix} \quad \text{and} \quad \boldsymbol{r}=\begin{pmatrix} \boldsymbol{r}_x \\ \boldsymbol{r}_y \\ \boldsymbol{r}_{xy} \end{pmatrix},$$

where $\boldsymbol{q}_x$ and $\boldsymbol{r}_x$ are $(J-2)$-vectors defined by

(12) $$\boldsymbol{q}_x=\begin{pmatrix} q_{31} \\ q_{41}-2q_{31} \\ q_{51}-2q_{41}+q_{31} \\ \vdots \\ q_{j1}-2q_{j-1\,1}+q_{j-2\,1} \\ \vdots \\ q_{J1}-2q_{J-1\,1}+q_{J-2\,1} \end{pmatrix} \quad \text{and} \quad \boldsymbol{r}_x=\begin{pmatrix} -2q_{21}+q_{11} \\ q_{21} \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

respectively; $\boldsymbol{q}_y$ and $\boldsymbol{r}_y$ are $(K-2)$-vectors defined by

(13) $$\boldsymbol{q}_y=\begin{pmatrix} q_{13} \\ q_{14}-2q_{13} \\ q_{15}-2q_{14}+q_{13} \\ \vdots \\ q_{1\,k}-2q_{1\,k-1}+q_{1\,k-2} \\ \vdots \\ q_{1\,K}-2q_{1\,K-1}+q_{1\,K-2} \end{pmatrix} \quad \text{and} \quad \boldsymbol{r}_y=\begin{pmatrix} -2q_{12}+q_{11} \\ q_{12} \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Also, $\boldsymbol{q}_{xy}$ and $\boldsymbol{r}_{xy}$ are

(14) $$\boldsymbol{q}_{xy}=\begin{pmatrix} q_{22} \\ q_{23}-q_{22}-q_{13} \\ q_{24}-q_{23}-q_{14}+q_{13} \\ \vdots \\ q_{2\,K}-q_{2\,K-1}-q_{1\,K}+q_{1\,K-1} \\ q_{32}-q_{31}-q_{22} \\ q_{33}-q_{32}-q_{23}+q_{22} \\ \vdots \\ q_{J\,K}-q_{J\,K-1}-q_{J-1\,K}+q_{J-1\,K-1} \end{pmatrix} \quad \text{and} \quad \boldsymbol{r}_{xy}=\begin{pmatrix} -q_{21}-q_{12}+q_{11} \\ q_{12} \\ 0 \\ \vdots \\ 0 \\ q_{21} \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

respectively. Then we have

$$(15)\qquad |D\boldsymbol{q}+\boldsymbol{r}|^2=\sum_{k=3}^{K}(q_{1\,k}-2q_{1\,k-1}+q_{1\,k-2})^2+\sum_{j=3}^{J}(q_{j\,1}-2q_{j-1\,1}+q_{j-2\,1})^2$$
$$+\sum_{j=2}^{J}\sum_{k=2}^{K}(q_{j\,k}-q_{j\,k-1}-q_{j-1\,k}+q_{j-1\,k-1})^2\,.$$

Note that thus defined D has the unit determinant. The vector $\boldsymbol{r}$ in (9) is a function of q_{11}, q_{12} and q_{21}, which are the parameters of the prior distribution, or hyper-parameters. Also w^2 is another hyper-parameter.

The prior distribution defined by (9) and (15) is derived from the consideration that the logistic transformation $q(x, y)$ of $p(x, y|\tilde{\boldsymbol{q}})$ should be approximated, at least locally, by a plane. Note that (15) is minimized when q_{jk} is a linear function of j and k.

Since $\tilde{\boldsymbol{q}}$ is composed of q_{11}, q_{12}, q_{21} and the elements of $\boldsymbol{q}$, we will adopt the notation $L(\boldsymbol{q}|q_{11}, q_{12}, q_{21})$ for $L(\tilde{\boldsymbol{q}})$. When values of hyper-parameters are suitably chosen, our estimate of $\boldsymbol{q}$ is defined by the value that maximizes posterior distribution of $\boldsymbol{q}$ which is proportional to $L(\boldsymbol{q}|q_{11}, q_{12}, q_{21})\pi(\boldsymbol{q}|q_{11}, q_{12}, q_{21}, w^2)$.

For the choice of the hyper-parameters q_{11}, q_{12}, q_{21} and w^2 we adopt the statistic ABIC defined by

$$(16)\qquad \text{ABIC}=-2\log\int L(\boldsymbol{q}|q_{11}, q_{12}, q_{21})\pi(\boldsymbol{q}|q_{11}, q_{12}, q_{21}, w^2)d\boldsymbol{q}\,.$$

The values of hyper-parameters are chosen so that ABIC is minimized.

4. Numerical consideration

Since the exact integration in (16) is difficult, we first find an approximation of the integrand

$$(17)\qquad L(\boldsymbol{q}|q_{11}, q_{12}, q_{21})\pi(\boldsymbol{q}|q_{11}, q_{12}, q_{21}, w^2)\,.$$

Taking logarithm of $L(\boldsymbol{q}|q_{11}, q_{12}, q_{21})$, we have

$$(18)\qquad l(\boldsymbol{q}|q_{11}, q_{12}, q_{21})=\log L(\boldsymbol{q}|q_{11}, q_{12}, q_{21})$$
$$=\sum_{i=1}^{n}[h_i\log p(x_i, y_i|\tilde{\boldsymbol{q}})+(1-h_i)\log\{1-p(x_i, y_i|\tilde{\boldsymbol{q}})\}]$$
$$=\sum_{i=1}^{n}h_i\sum_{j=1}^{J}\sum_{k=1}^{K}q_{jk}s_j(x_i)s_k(y_i)$$
$$-\sum_{i=1}^{n}\log\left[1+\exp\left\{\sum_{j=1}^{J}\sum_{k=1}^{K}q_{jk}s_j(x_i)s_k(y_i)\right\}\right].$$

The first and second partial derivatives of (18) are given by

$$\frac{\partial l}{\partial q_{rt}}=\sum_{i=1}^{n}\{h_i-p(x_i, y_i|\tilde{q})\}s_r(x_i)s_t(y_i)$$

and

$$(19)\qquad \frac{\partial^2 l}{\partial q_{rt}\partial q_{uv}}=-\sum_{i=1}^{n}s_r(x_i)s_t(y_i)s_u(x_i)s_v(y_i)p(x_i, y_i|\tilde{q})\{1-p(x_i, y_i|\tilde{q})\}\,,$$

respectively. Here we used the relation

$$(20)\qquad \frac{\partial p(x_i, y_i|\tilde{q})}{\partial q_{uv}}=p(x_i, y_i|\tilde{q})s_u(x_i)s_v(y_i)-p(x_i, y_i|\tilde{q})^2 s_u(x_i)s_v(y_i)\,.$$

From (19) and (20) it is easily shown that the third order derivatives are uniformly bounded for the entire space of the parameter $\boldsymbol{q}$. This observation suggests that $\log L(\boldsymbol{q}|q_{11}, q_{12}, q_{21})$ for fixed q_{11}, q_{12} and q_{21} is closely approximated, at least locally, by the Taylor expansion up to the second order defined by

$$\begin{aligned}T(\boldsymbol{q}|q_{11}, q_{12}, q_{21}, \boldsymbol{q}^0)=&\log L(\boldsymbol{q}^0|q_{11}, q_{12}, q_{21})\\&+\sum_{i=1}^{n}\sum_{r,t}\{h_i-p(x_i, y_i|\tilde{\boldsymbol{q}}^0)\}s_r(x_i)s_t(y_i)(q_{rt}-q_{rt}^0)\\&-\frac{1}{2}\sum_{i=1}^{n}\sum_{r,t}\sum_{u,v}p(x_i, y_i|\tilde{\boldsymbol{q}}^0)\{1-p(x_i, y_i|\tilde{\boldsymbol{q}}^0)\}\\&\times s_r(x_i)s_t(y_i)s_u(x_i)s_v(y_i)(q_{rt}-q_{rt}^0)(q_{uv}-q_{uv}^0)\,,\end{aligned}$$

where $\boldsymbol{q}^0$ is a properly chosen initial point and q_{rt}^0's are its components. $\tilde{\boldsymbol{q}}^0$ is the vector which is composed of q_{11}, q_{12}, q_{21} and the elements of $\boldsymbol{q}^0$. The summation $\sum_{r,t}$ is to be carried out over all elements of $\boldsymbol{q}$. Given a value of w^2, the maximization of (17) is realized by the iterative maximization of

$$T(\boldsymbol{q}|q_{11}, q_{12}, q_{21}, \boldsymbol{q}^0)+\log \pi(\boldsymbol{q}|q_{11}, q_{12}, q_{21}, w^2)$$

starting with properly chosen $\boldsymbol{q}^0$, then renewing $\boldsymbol{q}^0$ by the maximizing value.

Assume that the maximizing value $\boldsymbol{q}^*$ is found, then we approximate (17) by

$$(21)\qquad \exp\{T(\boldsymbol{q}|q_{11}, q_{12}, q_{21}, \boldsymbol{q}^*)\}\pi(\boldsymbol{q}|q_{11}, q_{12}, q_{21}, w^2)\,.$$

Though the integration of (17) is difficult, the integration of (21) is analytically found and we obtain an approximate value to ABIC. Let us denote by $-(\boldsymbol{q}-\boldsymbol{q}^*)^t H^*(\boldsymbol{q}-\boldsymbol{q}^*)/2$ the second order term to $T(\boldsymbol{q}|q_{11}, q_{12}, q_{21}, \boldsymbol{q}^*)$, where H^* is the minus Hessian of $\log L(\boldsymbol{q}|q_{11}, q_{12}, q_{21})$ evaluated at $\boldsymbol{q}=\boldsymbol{q}^*$. Then (21) has the expression:

$$\left(\frac{1}{\sqrt{2\pi}}\right)^{JK-3}(\det D^tD/w^2)^{1/2}L(\boldsymbol{q}^*|q_{11}, q_{12}, q_{21})\pi(\boldsymbol{q}^*|q_{11}, q_{12}, q_{21}, w^2)$$
$$\times\exp\left\{-\frac{1}{2}(\boldsymbol{q}-\boldsymbol{q}^*)^t(H^*+D^tD/w^2)(\boldsymbol{q}-\boldsymbol{q}^*)\right\}$$

and an approximate value to ABIC is obtained as

$$\begin{aligned}(22)\qquad \mathrm{ABIC}&=(JK-3)\log 2\pi-\log\det D^tD/w^2\\&\quad-2\log L(\boldsymbol{q}^*|q_{11}, q_{12}, q_{21})\pi(\boldsymbol{q}^*|q_{11}, q_{12}, q_{21}, w^2)\\&\quad-2\log\int\exp\left\{-\frac{1}{2}(\boldsymbol{q}-\boldsymbol{q}^*)^t(H^*+D^tD/w^2)(\boldsymbol{q}-\boldsymbol{q}^*)\right\}d\boldsymbol{q}\\&=-2\log L(\boldsymbol{q}^*|q_{11}, q_{12}, q_{21})\pi(\boldsymbol{q}^*|q_{11}, q_{12}, q_{21}, w^2)\\&\quad+\log\det(w^2H^*+D^tD)\,.\end{aligned}$$

Note that H^* depends on the choice of q_{11}, q_{12} and q_{21}. However, the dependence of the approximate ABIC on q_{11}, q_{12} and q_{21} through the second term of the most right hand side of (22) is relatively minor compared to that through the first term. This means that when w^2 is fixed, the choice of q_{11}, q_{12} and q_{21} is practically realized by the simultaneous maximization by Newton method of $L(\boldsymbol{q}|q_{11}, q_{12}, q_{21})\pi(\boldsymbol{q}|q_{11}, q_{12}, q_{21}, w^2)$ with respect to q_{11}, q_{12}, q_{21} and $\boldsymbol{q}$. The value of w^2 is selected by the grid search procedure.

The amount of the computation is proportional to the sample size n as is clear from the form of the equation (18). This burden will be significantly decreased if (x_i, y_i)'s take their values only on a few fixed points. Some kind of data, such as data of the first to third examples in the next section, have such a property.

If not so, it will be practical to adopt preliminary discretization of data (x_i, y_i)'s. If this discretization is fine enough compared to the segmentation of the spline function this trick is practically harmless. The results of the simulations in the next section are obtained by such manipulation.

5. Numerical examples

The first example is the data shown in Table 1 which were obtained in a public opinion poll conducted by the Institute of Statistical Mathematics in Tokyo in December 1981. They asked whether a respondent prospects that the existing one-party cabinet will endure or not. Table 1 is the table of responses of 568 respondents to this question by age and 'durables' which means here the number of durable consumer goods in the respondent's possession and takes an integer value between 0 and 17. Table 1 takes an unusual form but contains all the

Table 1. The number of the respondents having conservative prospect out of each entry in parenthesis according to age and the number of durables.

Age \ Durables	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
20	0(1)								0(1)			0(2)	2(2)		1(1)			
21			0(1)	1(1)	0(1)			1(1)		1(1)			1(2)	1(1)		0(2)		
22	1(1)				0(1)		2(2)	1(2)			1(1)	3(4)		2(2)				
23			0(1)					2(3)		1(2)		0(2)	2(2)					0(1)
24			1(1)	0(1)		1(2)	0(1)		1(1)	0(1)		1(1)	2(2)					
25			0(1)	1(1)	1(2)	1(1)		2(2)			1(1)							
26						0(1)	1(2)	0(1)	1(3)		1(2)	1(2)		1(1)				
27	1(1)						0(1)	2(2)	2(2)	2(2)	1(2)	1(1)		1(2)				
28				0(1)	0(1)	1(2)	0(1)	1(1)	3(3)	1(1)	1(2)	2(2)						
29						1(2)	0(1)	1(3)	1(1)	1(1)		1(2)	0(1)		1(1)			
30							1(2)	2(3)	2(3)	2(2)	0(1)	2(2)	2(2)		1(1)			
31						1(1)	2(3)	4(4)	0(1)			2(2)		1(1)				
32						0(1)	1(2)	2(3)	1(2)	3(3)	1(2)	1(1)			1(1)	1(1)		0(1)
33					1(2)	1(2)	2(2)	2(2)	2(2)	1(3)	1(1)	2(3)	1(1)	1(1)				
34		1(1)		1(1)			3(4)	1(2)	1(3)	0(2)	1(2)	1(1)	1(1)	1(1)	1(1)			
35							2(3)	2(2)	1(2)			0(1)	0(1)		1(1)			
36			1(1)	0(1)				2(2)	1(1)		1(1)	1(1)						
37						1(1)	0(1)		0(1)	3(4)		1(1)		1(1)	1(2)	0(1)		
38					2(3)		1(1)	2(3)	1(1)	4(5)		4(4)						
39					1(1)	1(1)	2(3)		3(4)	1(1)		1(1)	2(2)	2(3)				
40						1(1)		3(3)	1(1)	0(2)		0(1)	0(1)		1(1)	1(1)		
41		2(2)			1(1)		0(2)		0(1)			1(1)		1(1)	1(1)			
42						1(2)		1(1)		1(1)	1(1)		0(1)					1(1)
43							1(1)	2(2)		1(1)	1(2)	0(1)	2(3)		1(1)	1(1)		
44	1(1)						1(1)	3(3)	1(1)	1(1)	2(2)	1(1)	0(2)				0(1)	
45					0(1)		0(1)		1(2)	2(2)	1(1)	0(1)	2(3)	1(1)				
46					1(1)		1(1)	2(2)	1(1)	2(2)		1(1)		2(2)	2(2)	1(1)	0(1)	
47					1(1)	0(1)	2(2)	2(3)	1(1)		1(1)	3(4)	1(1)					
48					1(1)		1(1)	1(1)		1(1)	1(1)	1(1)			2(2)			
49							1(1)		1(1)	0(1)	2(2)	1(1)				2(2)		
50						1(1)			0(1)	1(3)	0(1)		1(2)	1(1)				
51	0(1)							1(2)	2(2)	1(1)	3(4)	2(2)			2(2)		1(1)	
52					0(1)	1(1)		1(1)	1(3)	0(1)	1(1)	3(4)	2(4)					
53	0(1)				1(1)	0(1)		2(2)	1(1)	0(1)			0(1)	2(2)				
54	1(1)							1(3)	1(1)	2(4)	0(1)		1(1)				1(1)	
55								0(1)		1(1)	1(1)	1(2)	1(1)	2(2)				
56						1(2)		1(1)	0(1)		1(1)	0(2)		2(3)				
57					1(1)	1(2)			1(3)	1(1)	0(1)	1(1)						
58	1(1)		1(1)			1(1)	0(1)		0(1)	2(3)		2(2)			2(2)			
59						0(1)	1(1)		1(1)		1(2)		1(2)				1(1)	
60	1(1)				0(1)			1(1)		1(2)							1(1)	
61	1(1)					1(1)	0(1)	1(1)	1(2)	1(2)		2(2)						
62							1(1)	1(1)	1(3)	2(4)		0(1)						
63						1(1)	1(1)			1(1)			1(2)		1(1)			
64					1(1)	2(2)	1(1)			2(3)						1(1)	0(1)	
65						1(1)	0(1)							0(1)				
66						0(1)		1(2)	0(1)									
67					1(1)		0(1)	1(1)							1(1)			
68						0(1)	0(1)				0(1)							
69							0(1)	0(1)		0(1)				0(1)		1(1)		
70																0(1)		
71			0(1)		1(1)	0(1)			1(1)	0(1)	1(1)	0(1)	0(2)					
72						0(1)	1(1)	1(2)					0(1)					
73	0(1)					1(1)		1(1)		1(1)								
74																		
75							1(1)		0(2)									
76						1(1)												
77						0(1)										1(1)		
78						0(1)	1(1)	1(2)										
79																		
80	1(1)	0(1)		1(1)														
81																		
82					1(1)				0(1)									
83																		
84																		
85								0(1)										
86																		
87																		
88																		
89																		
90																		
91																		
92																		
93																		
94								0(1)										

information which will be furnished by a familiar $2\times75\times18$ three-way table. In Table 1, an entry in parenthesis in each cell represents the marginal frequency with respect to the two explanatory variables, age and 'durables', and a numerical value preceding the parenthesis represents the number of respondents belonging to the category 'will endure' out of the respondents in the parenthesis. Empty cells mean zero frequency.

Our purpose is to see the relation of the probability supporting the opinion 'will endure' to a pair of values of age and 'durables'. Figure 1.a is a digital representation of the final result for those data and Figure 1.b is a gray shading display of the result. We put $J=K=10$ for this and following all examples in this section. In these figures we plotted age in abscissa and the 'durables' in ordinate. Though our procedure gives the estimate at any point on the x-y plane, the magnitude of the probabilities only at 75×18 grids is represented in

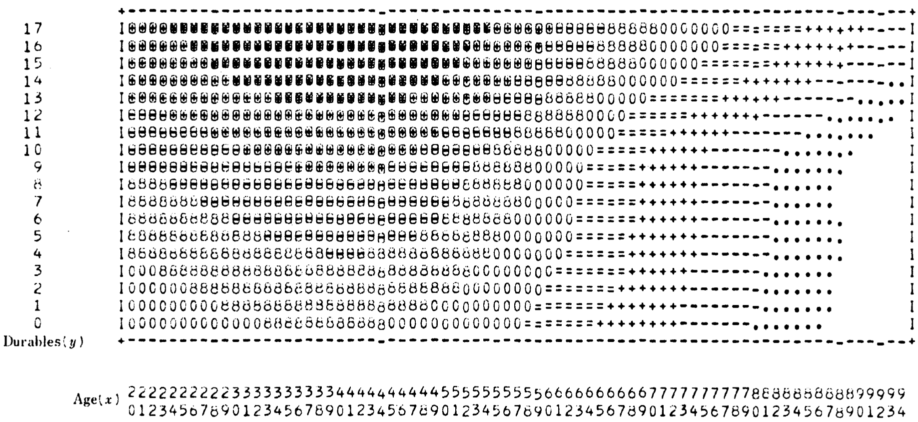
```
              +---------------------------------------------------------------------------+
17            IVVVVWWWXXXYYYYZZZZZZZZZZZZZZYYYXXWWVUUTSSRQPPONMLLKJJIHHGGFEEDDCCBAA9987766I
16            IUUUVVVWWWXXXYYYYYZZZZZZZZZYYYXXXWVVUUTSSRQQPONMLLKJJIHHGGFEEDDCBBAA98877665I
15            ITUUUUVVVWWWXXXXYYYYYYYYYYYYYXXWWVVUUTSSRQQPONMLLKJIIHHGFFEEDCCBBA9988766554I
14            ISTTTUUUVVVWWWWXXXXXXXXXXXXXXWWVVUUTSSRRQPONNMLKJJIHGGFEEDDCBBAA998776654433I
13            ISSSSTTTUUUVVVVWWWWWWWWWWWWWVVUUTTSSRQQPONNMLKJIIHGFFEDDCCBAA99877665443321 1I
12            IRRRRSSSSTTTUUUUVVVVVVVVVVVUUUTTSRRQQPONNMLKKJIHGFFEDCCBBA99887665543322100+I
11            IQQQQRRRRSSSTTTTUUUUUUUUUUUTTTSSRRGQPOONMLLKJIHGGFEDCCBAA998776654433211 00++-I
10            IPPPPPQQQRRRSSSSSTTTTTTTTSSSSRRQQPOONMMLKJJIHGFEEDCBBA99877665443322100++-..I
 9            IOOOOPPPPQQQRRRRRSSSSSSSSSRRRQQPPONNMLLKJJIHGFEEDCBBA998776654433211 00++-.. I
 8            INNNNOOOPPPPQQQQRRRRRRRRRRRQQPPOONNMLLKJJIHHGFEDDCBAA98877655443222110++--.. I
 7            IMMMNNNNOOOPPPPQQQQQQQQQQQQPPOONNMMLKKJIIHGFFEDCCBAA9887765544322110++--.. I
 6            ILLMMMMNNNNOOOPPPPPPPQPPPPPPOOONNMMLLKJJIHHGFFEDCCBAA98877655443321100+--.. I
 5            IKLLLLLMMMNNNNOOOOOOPPPPOOOOONNMMLLKKJJIIHGGFEEDCBBAA98877655443321100++-.. I
 4            IKKKKKLLLLMMMMNNNNNNNOOOONNNNNMMMLLKKJJIIHGGFFEDDCBBA998877655443321100++-.. I
 3            IJJJKKKKKLLLLMMMMMMNNNNNNMMMMLLLKKJJJIIHGGFFEDDCCBAA99877665543322110++--.. I
 2            IIJJJJJKKKKKLLLLLMMMMMMMMLLLLLKKKJJIIHHGGFFEEDCCBBA998877655443322100++--.. I
 1            IIIIIJJJJJKKKKKKLLLLLLLLLLKKKJJJIIIHHGGFFEEDDCBBAA9987766554432 21100++-..  I
 0            IHIIIIIIJJJJJJKKKKKKKKKKKJJJJIIIIHHGGGFFEEDDCBBAA998877655443322100++--..  I
Durables(y)   +---------------------------------------------------------------------------+

       Age(x) 222222222233333333334444444444555555555566666666667777777777888888888899999
              012345678901234567890123456789012345678901234567890123456789012345678901234
```

Note:

```
0.3896D+00 ~ 0.3990D+00  :
0.3990D+00 ~ 0.4083D+00  : ..........
0.4083D+00 ~ 0.4177D+00  : ----------
0.4177D+00 ~ 0.4270D+00  : ++++++++++
0.4270D+00 ~ 0.4364D+00  : 0000000000
0.4364D+00 ~ 0.4457D+00  : 1111111111
0.4457D+00 ~ 0.4551D+00  : 2222222222
0.4551D+00 ~ 0.4644D+00  : 3333333333
0.4644D+00 ~ 0.4738D+00  : 4444444444
0.4738D+00 ~ 0.4831D+00  : 5555555555
0.4831D+00 ~ 0.4925D+00  : 6666666666
0.4925D+00 ~ 0.5018D+00  : 7777777777
0.5018D+00 ~ 0.5112D+00  : 8888888888
0.5112D+00 ~ 0.5205D+00  : 9999999999
0.5205D+00 ~ 0.5299D+00  : AAAAAAAAAA
0.5299D+00 ~ 0.5392D+00  : BBBBBBBBBB
0.5392D+00 ~ 0.5486D+00  : CCCCCCCCCC
0.5486D+00 ~ 0.5580D+00  : DDDDDDDDDD
0.5580D+00 ~ 0.5673D+00  : EEEEEEEEEE
0.5673D+00 ~ 0.5767D+00  : FFFFFFFFFF
0.5767D+00 ~ 0.5860D+00  : GGGGGGGGGG
0.5860D+00 ~ 0.5954D+00  : HHHHHHHHHH
0.5954D+00 ~ 0.6047D+00  : IIIIIIIIII
0.6047D+00 ~ 0.6141D+00  : JJJJJJJJJJ
0.6141D+00 ~ 0.6234D+00  : KKKKKKKKKK
0.6234D+00 ~ 0.6328D+00  : LLLLLLLLLL
0.6328D+00 ~ 0.6421D+00  : MMMMMMMMMM
0.6421D+00 ~ 0.6515D+00  : NNNNNNNNNN
0.6515D+00 ~ 0.6608D+00  : OOOOOOOOOO
0.6608D+00 ~ 0.6702D+00  : PPPPPPPPPP
0.6702D+00 ~ 0.6795D+00  : QQQQQQQQQQ
0.6795D+00 ~ 0.6889D+00  : RRRRRRRRRR
0.6889D+00 ~ 0.6982D+00  : SSSSSSSSSS
0.6982D+00 ~ 0.7076D+00  : TTTTTTTTTT
0.7076D+00 ~ 0.7169D+00  : UUUUUUUUUU
0.7169D+00 ~ 0.7263D+00  : VVVVVVVVVV
0.7263D+00 ~ 0.7356D+00  : WWWWWWWWWW
0.7356D+00 ~ 0.7450D+00  : XXXXXXXXXX
0.7450D+00 ~ 0.7543D+00  : YYYYYYYYYY
0.7543D+00 ~ 0.7637D+00  : ZZZZZZZZZZ
```

Figure 1.a. Digital representation of the estimates corresponding to the minimum of ABIC attained at $w=1/16$.



Note:

```
0.3896D+00 ~ 0.4270D+00   :
0.4270D+00 ~ 0.4644D+00   : ..........
0.4644D+00 ~ 0.5018D+00   : ----------
0.5018D+00 ~ 0.5392D+00   : ++++++++++
0.5392D+00 ~ 0.5767D+00   : ==========
0.5767D+00 ~ 0.6141D+00   : 0000000000
0.6141D+00 ~ 0.6515D+00   : 8888888888
0.6515D+00 ~ 0.6889D+00   : 8888888888
0.6889D+00 ~ 0.7263D+00   : 8888888888
0.7263D+00 ~ 0.7637D+00   : 8888888888
```

Figure 1.b. Gray shading display of the estimates in Figure 1.a.

40 different stages in Figure 1.a and in different 10 grades in Figure 1.b. In Figure 1.b the darkness of the symbol is proportional to the magnitude of the corresponding probability.

The two-way table data between the responses to the same question and age were analyzed in [2] and it was found that the curve of the probability supporting the opinion 'will endure' is unimodal with respect to age and the probability takes the maximum value 0.7143 at the age of 42. Figure 1.a clarifies further detailed aspects that the conservative view increases with the 'durables' and that the effect of the 'durables' on the probability is not uniform over all age.

The second example is the data shown in Table 2 which were obtained from the survey of the Japanese national character carried out by the Institute of Statistical Mathematics. The survey was originated with the first nation-wide survey in 1953. Since then, a similar survey has been conducted every five years. Each of the seven surveys carried out thus far consists of 2000 to 3000 respondents, aged 20 and over. Table 2 is a table with the same manner of representation as Table 1 and shows the number of people whose principle in their daily life was "don't think about money or fame; just live a life that suits your own tastes". If we regard the survey period and age in Table 2 as the explanatory variables to apply the procedure to those data, we get the final result shown in Figure 2. This figure shows that the

Table 2. The number of the respondents having the life style "Don't think about money or fame; just live a life that suits your own tastes" out of each entry in parenthesis according to age and survey period.

	Survey period						
Age	1953	1958	1963	1968	1973	1978	1983
20	49(98)	14(28)	29(56)	34(67)	37(73)	12(36)	23(44)
21	27(94)	9(34)	40(69)	36(55)	28(53)	23(39)	16(38)
22	32(69)	8(25)	30(60)	29(68)	39(79)	28(41)	11(31)
23	16(70)	9(28)	19(52)	36(68)	53(101)	9(25)	19(25)
24	20(72)	14(29)	28(72)	42(67)	43(81)	18(32)	22(43)
25	17(71)	9(27)	32(79)	38(91)	36(88)	19(44)	11(27)
26	17(76)	11(32)	25(64)	29(69)	41(72)	20(37)	15(30)
27	18(72)	9(23)	23(77)	39(92)	33(60)	30(54)	15(35)
28	16(65)	7(21)	30(72)	26(72)	34(76)	36(70)	24(48)
29	14(61)	7(24)	27(69)	33(74)	29(63)	31(47)	17(43)
30	20(60)	11(26)	29(92)	25(79)	36(80)	32(60)	35(62)
31	16(63)	9(31)	25(64)	31(77)	37(77)	16(43)	24(43)
32	12(40)	11(30)	23(66)	28(79)	34(86)	16(39)	22(48)
33	11(42)	5(19)	24(63)	31(83)	23(67)	18(48)	19(50)
34	15(52)	9(22)	19(80)	36(81)	35(66)	21(48)	33(62)
35	16(56)	5(21)	22(77)	31(91)	27(74)	20(58)	27(66)
36	11(54)	3(17)	17(69)	19(77)	38(64)	22(42)	27(59)
37	12(41)	5(23)	33(70)	19(76)	28(73)	21(58)	23(45)
38	4(48)	5(16)	14(50)	25(77)	29(66)	24(44)	13(33)
39	8(49)	1(13)	15(62)	27(69)	23(74)	21(55)	17(51)
40	4(44)	5(15)	21(60)	25(73)	34(84)	17(38)	27(52)
41	6(45)	4(18)	25(65)	31(78)	31(66)	21(46)	23(58)
42	7(45)	4(18)	16(61)	33(65)	24(69)	29(58)	17(50)
43	8(44)	3(24)	13(45)	18(60)	27(69)	11(41)	22(53)
44	9(45)	6(21)	9(48)	19(68)	25(68)	13(48)	13(49)
45	1(33)	3(19)	11(44)	20(46)	23(74)	14(42)	14(38)
46	6(44)	7(21)	16(50)	15(50)	33(80)	22(54)	26(54)
47	7(41)	6(20)	8(43)	17(65)	20(57)	14(39)	25(62)
48	8(39)	2(9)	9(47)	5(36)	11(45)	10(37)	15(54)
49	4(30)	1(20)	12(55)	11(46)	20(54)	12(49)	13(47)
50	4(44)	0(16)	14(52)	11(41)	14(45)	14(43)	17(51)
51	9(36)	3(10)	12(59)	9(34)	16(51)	20(43)	12(54)
52	4(44)	2(13)	11(49)	11(44)	13(50)	14(39)	13(46)
53	6(36)	4(9)	10(46)	10(59)	9(38)	10(29)	6(43)
54	6(41)	3(11)	10(46)	9(45)	15(41)	12(26)	14(34)
55	8(39)	5(14)	10(39)	4(42)	19(47)	8(29)	20(51)
56	1(24)	1(14)	6(34)	10(52)	13(43)	11(19)	10(36)
57	2(29)	2(17)	7(21)	10(32)	10(37)	13(30)	9(37)
58	2(22)	0(6)	11(31)	9(41)	14(56)	8(26)	13(36)
59	3(21)	3(12)	4(27)	6(37)	16(40)	10(33)	13(32)
60	2(29)	3(12)	3(32)	10(46)	12(44)	7(25)	8(34)
61	5(21)	1(10)	7(25)	8(37)	7(26)	3(15)	11(38)
62	0(16)	2(8)	8(34)	3(30)	9(33)	7(22)	16(30)
63	3(23)	0(7)	5(40)	9(37)	7(29)	1(14)	12(30)
64	1(24)	3(6)	8(24)	7(42)	8(31)	8(22)	4(20)
65	2(13)	0(8)	5(25)	6(30)	7(23)	7(20)	5(13)
66	2(21)	0(9)	6(22)	4(20)	5(25)	3(17)	6(19)
67	0(7)	0(8)	1(26)	5(24)	6(26)	2(23)	8(20)
68	1(8)	1(8)	1(17)	3(26)	3(30)	3(16)	4(18)
69	2(10)	0(4)	6(23)	3(28)	0(17)	7(19)	6(31)
70	0(13)	1(5)	3(15)	1(24)	1(10)	3(19)	7(23)
71	0(6)	0(6)	0(16)	5(19)	2(15)	1(9)	5(27)
72	1(14)	0(7)	3(10)	2(13)	5(15)	2(19)	2(13)
73	1(5)	0(1)	4(25)	1(8)	3(18)	4(15)	6(21)
74	0(7)	1(4)	1(7)	2(19)	1(15)	0(11)	3(18)
75	1(5)	1(3)	1(7)	1(14)	0(9)	2(9)	3(11)
76	1(6)	0(2)	0(4)	1(10)	0(7)	2(8)	1(11)
77	0(2)		1(14)	0(3)	0(10)	3(12)	0(10)
78	0(3)	0(1)	1(4)	0(5)	0(9)	0(2)	0(9)
79		0(4)	0(5)	0(6)	2(4)	0(5)	1(10)
80		1(2)	0(2)	0(3)	0(8)	0(11)	0(5)
81			0(2)	0(7)	2(5)	1(4)	0(7)
82	0(1)		0(3)	0(2)	1(7)	0(2)	0(9)
83			0(1)	1(1)	1(6)	1(4)	1(1)
84		0(1)	0(2)		1(3)	1(2)	2(3)
85		0(1)			2(2)	0(1)	0(3)
86					0(2)	0(1)	
87	0(1)				0(1)		0(1)
88						0(1)	0(1)
89				0(1)	0(1)		
90					0(1)		
91					0(2)		
92							
93						0(1)	
94							
95				0(1)	0(1)		
96							
97							
98							
99					0(1)		

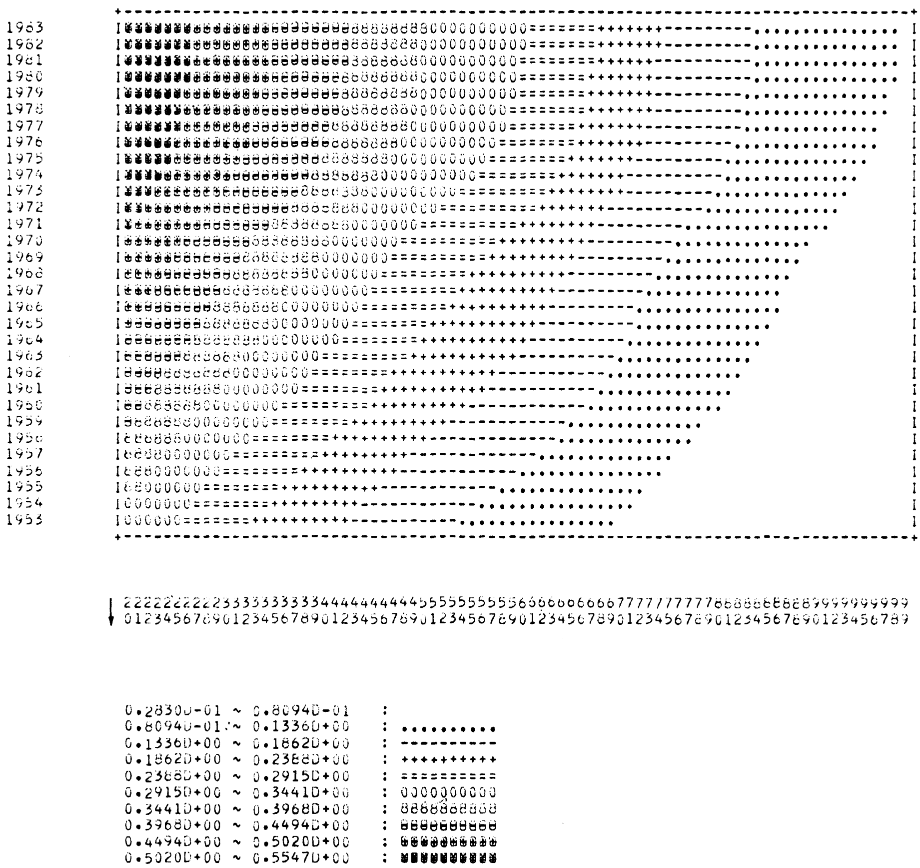


Figure 2. Estimates corresponding to the minimum of ABIC attained at $w=1/32$.

percentage of people choosing the principle has been increasing since the thickness of symbol at a given age group increases with the survey period. In an age-period-cohort table, such as Table 2, great interest is often focused on which of the three effects, age, period and cohort effects, is dominant on the growth of the principle. Figure 2 illustrates that the cohort effect is dominant as seen from the behavior of the contour in it. This finding coinsides with the result by Nakamura [4] obtained by the use of a Bayesian cohort model.

Next example is the analysis of a set of data supplied by the Meteorological Agency in Japan. The data are concerned with the number of days of snowfall observed at 46 spots (Figure 3) in Central Japan in January and February from 1978 through 1982. All the spots are situated within lat. 34°04′N to 37°55′N and long. 136°04′E to 140°54′E. Applying the present procedure to the data, we get Figure 4 as the final estimates of the probabilities of days of snowfall over the area. Figure 5 is a reproduction of the patterns in Figure 4 on

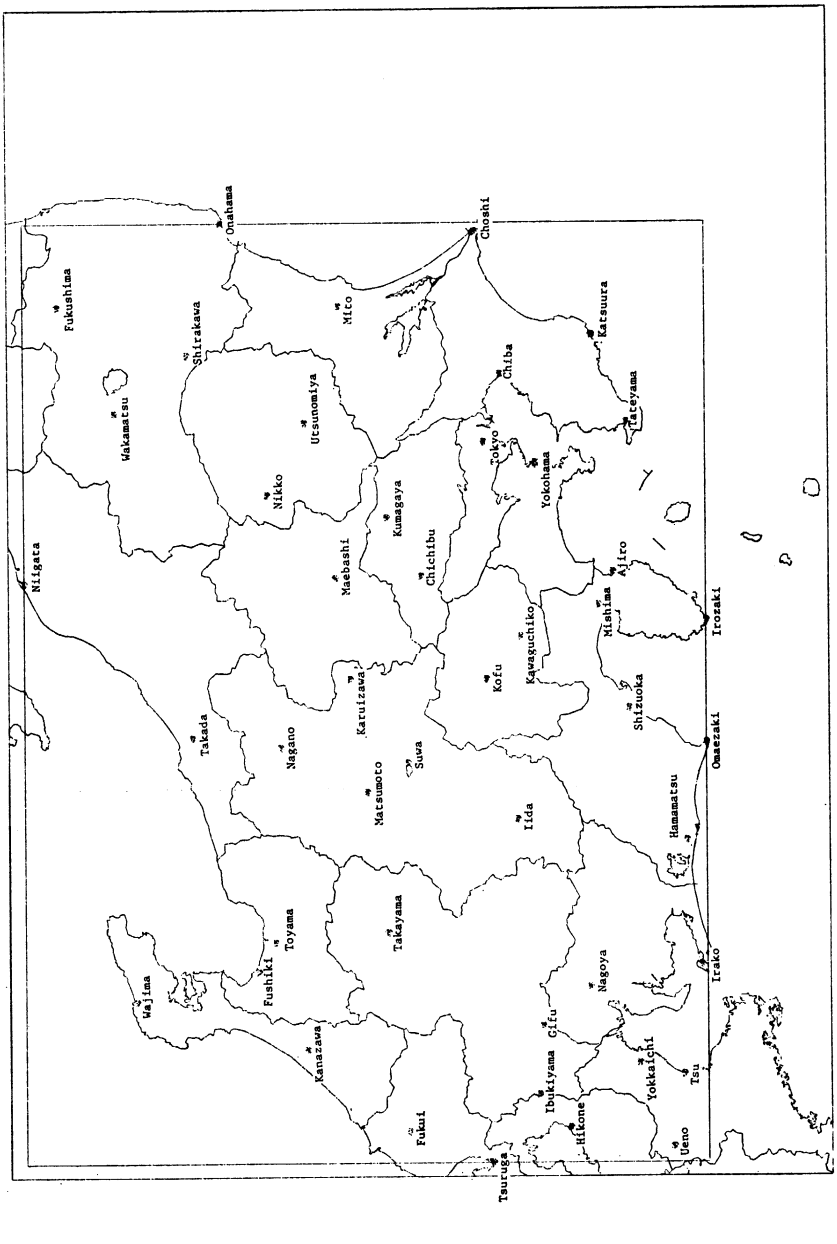


Figure 3. Locations of 46 meteorogical observatories.

Figure 4. Estimates of the probabilities of days of snowfall, which correspond to the minimum of ABIC attained at $w=1/2$.

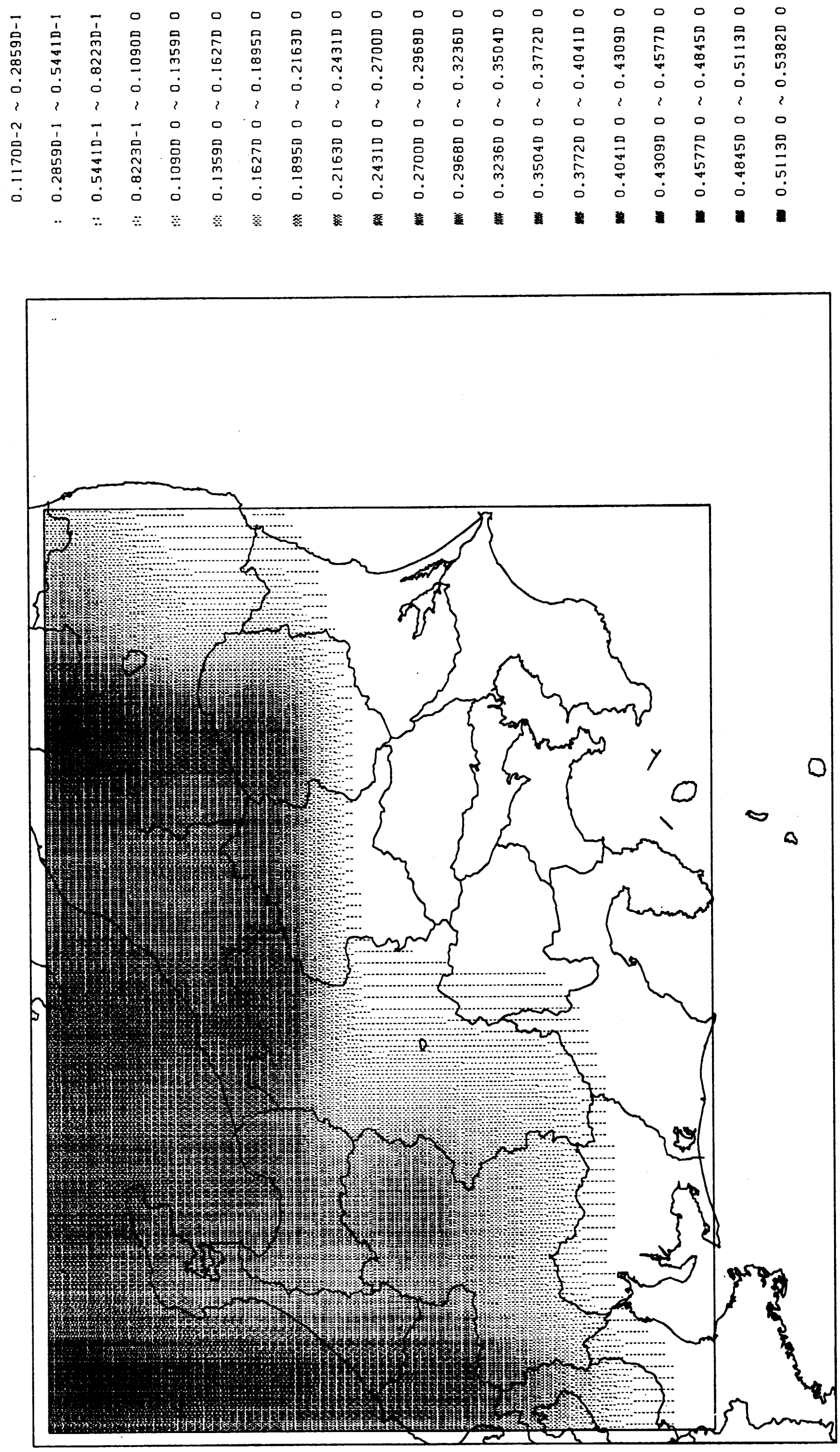


Figure 5. Reproduction of the final estimates in Figure 4.

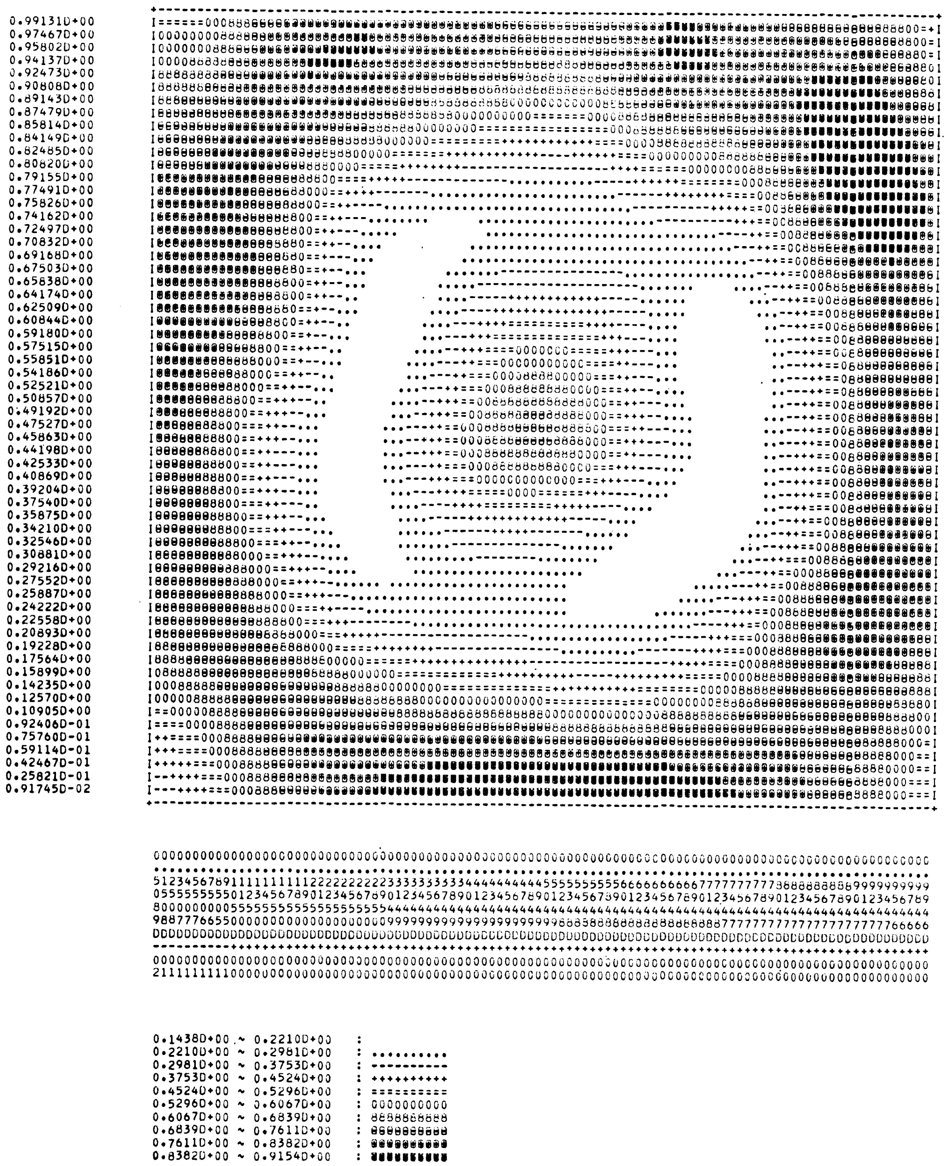


Figure 6. An example of the final estimates of $p_1^*(x, y)$ for $n=1000$.

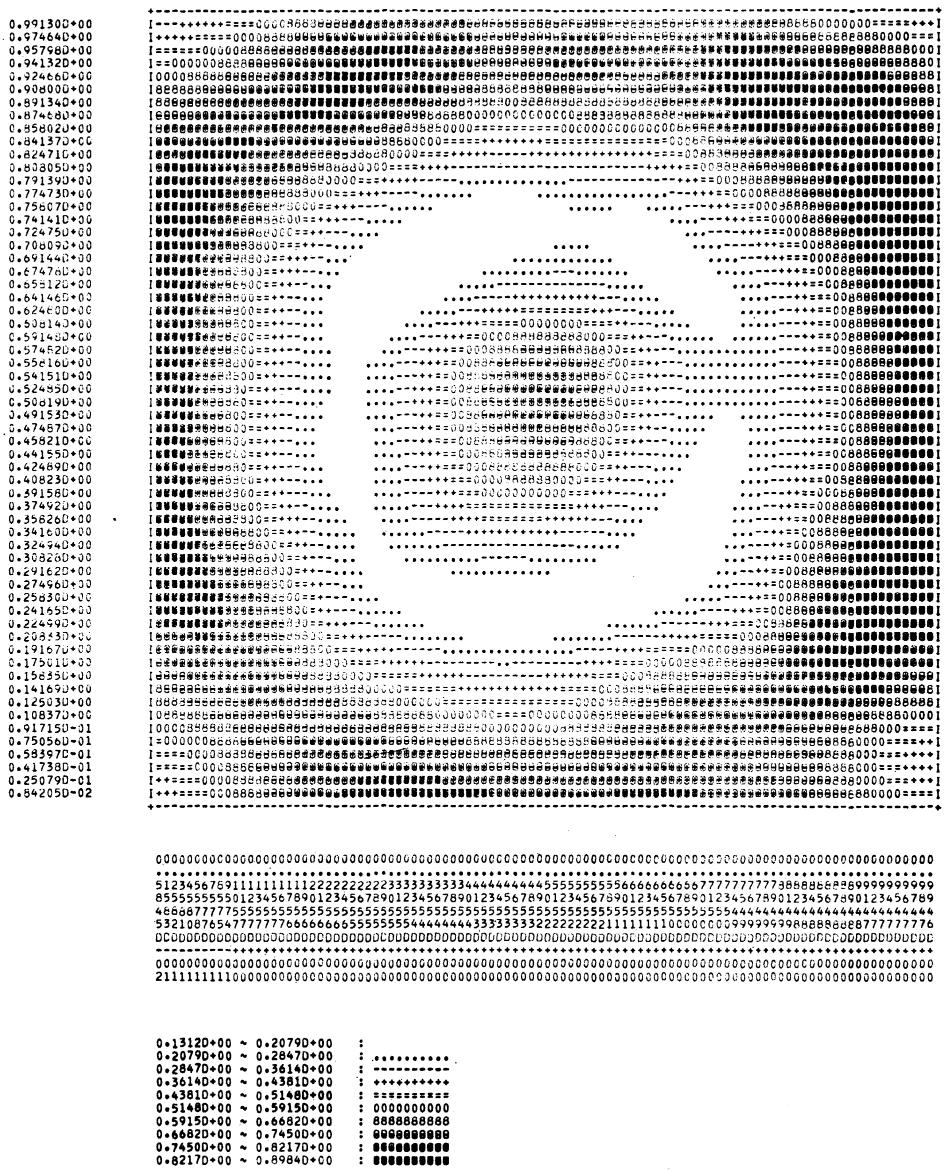


Figure 7. An example of the final estimates of $p_1^*(x, y)$ for $n=2000$.

the map. Apparently, the analysis of the data observed only at 46 spots has to result in a rough sketch of the real snowfall probability which may have local structure. Nevertheless, we can see that there is a noticeable difference in the probability of having snow between the Pacific side and the Japan Sea side. We remark that snowfall on the Japan Sea in Figure 5 may be neglected if it is not necessary. If we refine our procedure so that it gives confidence regions of the estimates, we shall see that the estimate on the Japan Sea is not reliable.

Finally we shall describe some results of simple experiments illustrating that the performance of the present procedure depends on the true structure and the sample size n. Figure 6 is an example of the final estimates from a random sample of size 1000 from a population whose true structure is given by

$$p_1^*(x, y)=\frac{2}{5}\sin\left[4\pi\left\{\sqrt{\left(x-\frac{1}{2}\right)^2+\left(y-\frac{1}{2}\right)^2}+\frac{1}{8}\right\}\right]+\frac{1}{2},$$

where both x and y are random numbers from the uniform distribution on the interval [0, 1]. When we put

$$r^2=\left(x-\frac{1}{2}\right)^2+\left(y-\frac{1}{2}\right)^2,$$

$p_1^*(x, y)$ has its maximum $\frac{9}{10}$ for $r=0$ and $r=\frac{1}{2}$ and has its minimum $\frac{1}{10}$ for $r=\frac{1}{4}$. But, Figure 6 does not reconstitute such features. The performance of the procedure improves with the sample size as is illustrated typically by a result shown in Figure 7 for $n=2000$. Of course, if the true structure is not so complicated, we can obtain fairly accurate estimates even from data of smaller sample size.

6. Conclusion

As mentioned in [2], there are two types of random variables, categorical and continuous, and then there are four types of relations between a response variable and its explanatory variables. Sakamoto [5], [6] proposed a practical procedure for the variable selection in the case where the response variable is categorical. To obtain good estimates of the conditional probabilities, the present authors proposed in [2] a Bayesian method for a binary response curve estimation in case of single explanatory variable. The present procedure is an extention of the method to the bivariate case. The demonstration in the preceding section suggests that the present procedure will efficiently work for any data set provided that the true function is properly smooth, and should have a wide range of applications. If we further refine the procedure so that it takes care of the multivariate cases, it will find a wider range of applications. This will be the subject for further study.

Acknowledgement

The authors are grateful to the referees for their valuable comments and to Mr. K. Katsura of the Institute of Statistical Mathematics for computing assistance.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Akaike, H. (1980). Likelihood and Bayes procedure, *Bayesian Statistics* (eds. J. M. Bernardo, M. H. De Groot, D. U. Lindley and A. F. M. Smith), University Press, Valencia, Spain.

[2] Ishiguro, M. and Sakamoto, Y. (1983). A Bayesian approach to binary response curve estimation, *Ann. Inst. Statist. Math.*, **35**, B, 115-137.

[3] Ishiguro, M. and Sakamoto, Y. (1984). A Bayesian approach to the probability density estimation, *Ann. Inst. Statist. Math.*, **36**, B, 523-538.

[4] Nakamura, T. (1982). A Bayesian cohort model for standard cohort table analysis, *Proc. Inst. Statist. Math.*, **29**, 77-97. (in Japanese)

[5] Sakamoto, Y. and Akaike, H. (1978). Analysis of cross classified data by AIC, *Ann. Inst. Statist. Math.*, **30**, B, 185-197.

[6] Sakamoto, Y. (1982). Efficient use of Akaike's information criterion for model selection in high dimensional contingency table analysis, *Metron*, **40**, 257-275.

Ann. Inst. Statist. Math.
39 (1987), Part B, 245

CORRECTIONS TO "BAYESIAN BINARY REGRESSION INVOLVING TWO EXPLANATORY VARIABLES"

YOSIYUKI SAKAMOTO AND MAKIO ISHIGURO

(This Annals Vol. 37, No. 2 (1985), pp. 369–387)

Page 370. At line 21, $\frac{1}{\sqrt{2\pi v}}$ should be $\left(\frac{1}{\sqrt{2\pi v}}\right)^c$.

Page 372. The line 2 should be
$\boldsymbol{q}=(q_{31},\cdots,q_{J1},q_{13},\cdots,q_{1K},q_{22},\cdots,q_{2K},q_{32},\cdots,q_{JK})^t$.

Page 375. At line 1, $\left(\frac{1}{\sqrt{2\pi}}\right)^{JK-3}(\det D^tD/w^2)^{1/2}$ should be deleted.

Page 375. At line 4, $(JK-3)\log 2\pi-\log\det D^tD/w^2$ should be deleted.

Page 375. At line 8, $\log\det(w^2H^*+D^tD)$ should be
$\log\det(H^*+D^tD/w^2)-(JK-3)\log 2\pi$.

THE INSTITUTE OF STATISTICAL MATHEMATICS