Hou, Z.
Osaka J. Math.
37 (2000), 651–666

CONVERGENCE OF LOGISTIC PARAMETERS IN BAYESIAN APPROACH

ZHEN-LIN HOU

(Received August 26, 1998)

1. Introduction

In this paper, we study the statistical model of the independent, 0-1-valued observations with the following distributions:

$$P(Y_i = 1) = \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}\ , \quad P(Y_i = 0) = \frac{1}{1+e^{\alpha+\beta x_i}}$$
$$(i = 1, 2, \ldots, n),$$

where x_i's are known real numbers called the *observation points*. It is sometimes more natural to consider the parameters α and β in the above logistic way than something like $e^{\alpha}/(1 + e^{\alpha})$'s. For example, let us cosider a random variable Y on $\{0, 1\}$ with small $P[Y = 1]$, where the value 1 stands for a serious accident which we must avoid definitely. Since we are sensitive on the value $P[Y = 1]$, we take the measurement $\log P[Y = 1]$ instead of the value itself. In this case, the logistic parametrization is suitable. In the same reason, it is natural to assume that the prior distribution α and β is uniform, that is, the joint prior density for (α, β) is given by $p(\alpha, \beta) \equiv 1$ on $\mathbf{R}^2$. Then we discuss the posterior distribution on (α, β) under a set of observations $Y_i = y_i$ $(i = 1, \ldots, n)$.

By the Bayes formula, the posterior probability density, is given by

$$p(\alpha, \beta | y_1, \ldots, y_n) = c^{-1} \prod_{i=1}^{n} \left(\frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}\right)^{y_i} \left(\frac{1}{1+e^{\alpha+\beta x_i}}\right)^{1-y_i}$$

if and only if the normalizing constant exists, that is

$$c := \iint \prod_{i=1}^{n} \left(\frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}\right)^{y_i} \left(\frac{1}{1+e^{\alpha+\beta x_i}}\right)^{1-y_i} d\alpha d\beta < \infty.$$

We obtain in Theorem 1 a necessary and sufficient condition for the existence of the posterior probability distribution, or equivalently, for $c < \infty$.

Theorem 1. *A necessary and sufficient condition for $c < \infty$ is that $1 \leq m \leq n-1$ and*

$$\min\left\{\sum_{i\in S} x_i : \sharp S = m\right\} < \sum_{i=1}^{n} x_i y_i < \max\left\{\sum_{i\in S} x_i : \sharp S = m\right\},$$

where we put $m = \sum_{i=1}^{n} y_i$ and $S \subset \{1, 2, \ldots, n\}$.

Under this condition, we consider α and β to be random variables and use the notations A and B for α and β in this sense to avoid a confusion with their sample values α and β.

We are interested in the convergence of the random variables A, B under the observations $y_1, y_2, \ldots, y_{kn}$ satisfying that k number of the observation points are fixed where the same number n of observations are allocated and the ratio of 1 among them converges as $n \to \infty$ to a value in (0, 1). That is, we assume the following set of observation points:

$$x_{i,j} \quad (i = 1, \ldots, k\ ;\ j = 1, \ldots, n)$$

with

$$x_{i,1} = \cdots = x_{i,n} := x_i \quad (i = 1, \ldots, k)$$

and $x_i < x_{i+1}$ for $i = 1 \cdots k-1$ with a fixed integer k not less than 2. Let $y_{i,j}$ be the set of corresponding observations, for which we assume that

$$p_i := \lim_{n\to\infty} \frac{t_i}{n} \quad (i = 1, \ldots, k)$$

exist for

$$t_i := \sum_{j=1}^{n} y_{i,j}$$

and it holds that $0 < p_i < 1 \quad (i = 1, \ldots, k)$.

Then, the posterior density $p(\alpha, \beta | t_1, \ldots, t_k)$ for (A, B) under these observations satisfies that

$$\begin{aligned} & p(\alpha, \beta | t_1, \ldots, t_k) \\ &= {c_n}^{-1} \prod_{i=1}^{k} \prod_{j=1}^{n} \left(\frac{e^{\alpha+\beta x_{i,j}}}{1 + e^{\alpha+\beta x_{i,j}}}\right)^{y_{i,j}} \left(\frac{1}{1 + e^{\alpha+\beta x_{i,j}}}\right)^{1-y_{i,j}} \\ &= {c_n}^{-1} \frac{\exp\{\sum_{i=1}^{k} t_i(\alpha + \beta x_i)\}}{\prod_{i=1}^{k} \{1 + \exp(\alpha + \beta x_i)\}^n} \end{aligned} \tag{1}$$

$$= c_n{}^{-1} \exp\left[n \sum_{i=1}^{k} \left\{ \frac{t_i}{n}(\alpha + \beta x_i) - \log(1 + \exp(\alpha + \beta x_i)) \right\} \right]$$

where c_n is the normalizing constant. We put

$$f(\alpha, \beta) := \sum_{i=1}^{k} \left[p_i(\alpha + \beta x_i) - \log\{1 + \exp(\alpha + \beta x_i)\} \right] \tag{2}$$

$$G_n(\alpha, \beta) := \sum_{i=1}^{k} \left[\frac{t_i}{n}(\alpha + \beta x_i) - \log\{1 + \exp(\alpha + \beta x_i)\} \right].$$

The *maximal likelihood estimator* $(\hat{\alpha_n}, \hat{\beta_n})$ is, by definition, a point (α, β) which maximize $G_n(\alpha, \beta)$. Similary, $(\hat{\alpha}, \hat{\beta})$ is defined to be (α, β) which maximize $f(\alpha, \beta)$.

Theorem 2. *The maximal likelihood estimator $(\hat{\alpha}_n, \hat{\beta}_n)$ exists uniquely.*

Theorem 3. *It holds that $(\hat{\alpha}, \hat{\beta})$ exists uniquely, $(\hat{\alpha}_n, \hat{\beta}_n)$ converges to $(\hat{\alpha}, \hat{\beta})$ as $n \to \infty$.*

Theorem 4. *The random variable (A, B) converges to $(\hat{\alpha}, \hat{\beta})$ in law.*

Corollary 1 (Lehmann [5], A. Ibragimov and R.Z. Khas'Minskii [10]). *Assume that $t_i/n = p_i + o(n^{-1})$ $(i = 1, \ldots, k)$ as $n \to \infty$. Then the distribution of the random variable $((A - \hat{\alpha})/\sqrt{n}, (B - \hat{\beta})/\sqrt{n})$ converges to the 2-dimensional centered normal distribution with the covariance matrix M^{-1}, where*

$$M = \begin{pmatrix} u & v \\ v & w \end{pmatrix}$$

with

$$u = \sum_{i=1}^{k} \frac{\exp(\hat{\alpha} + \hat{\beta} x_i)}{(1 + \exp(\hat{\alpha} + \hat{\beta} x_i))^2}$$

$$v = \sum_{i=1}^{k} \frac{x_i \exp(\hat{\alpha} + \hat{\beta} x_i)}{(1 + \exp(\hat{\alpha} + \hat{\beta} x_i))^2}$$

$$w = \sum_{i=1}^{k} \frac{x_i^2 \exp(\hat{\alpha} + \hat{\beta} x_i)}{(1 + \exp(\hat{\alpha} + \hat{\beta} x_i))^2}.$$

The aim of this paper is to justify the Bayesian approach for the logistic parameters by proving the consistency in Theorem 4 and the approximate normality in Corollary 1. The consistency for the natural parameters $e^{\alpha+\beta x_i}/(1 + e^{\alpha+\beta x_i})$ $(i = 1, \ldots, k)$

with the uniform distribution on $[0, 1]^k$ as their joint prior distribution is just the law of large number. One of the difficulties in our case is that the prior distribution is not a finite measure, so that we have to start with a condition for the posterior distribution to be a probability measure. As we already remarked, the logistic parameters are sometimes more natural than the natural parameters. This fact is also discussed in [1]. We refer to [2], [3], [4] for the meanings of Bayesian approach. Heberman [6] discussed the logit model with continuum observations, but did not discuss the binary data case which we discuss in this paper. Johan W. Pratt [7] discussed log likelihood for his model, but the model did not contain our case. Cox [9] gave a way to get maximum likelihood estimates, but he did not discuss the existence and uniqueness. Our results contains some of V.T. Farewell [11].

2. Proof of Theorem 1

For a given set of observation points x_i $(i = 1, \ldots, n)$ and a set of corresponding observations $y_i \in \{0, 1\}$ with $m := \sum_{i=1}^n y_i$ and $M := \sum_{i=1}^n x_i y_i$, we define a subset Ω of $\mathbf{R}^2$ as the closed convex set generated by the set

$$\left\{ \left(\sharp S,\ \sum_{i \in S} x_i \right) ;\ S \subset \{1, \ldots, n\} \right\}.$$

Let $\partial\Omega$ be the boundary of Ω. Then, the claimed condition in Theorem 1 is equivalent to $P := (m, M) \in \Omega \setminus \partial\Omega$, so that it is sufficient to prove that $c < \infty$ if and only if $P \in \Omega \setminus \partial\Omega$.

We put

$$Q_j = (\alpha_j,\ \beta_j) := \left(j,\ \min\left\{ \sum_{i \in S} x_i\ ;\ \sharp S = j \right\} \right)$$

for $j = 0, 1, \ldots, n$, and

$$Q_j = (\alpha_j,\ \beta_j) := \left(2n - j,\ \max\left\{ \sum_{i \in S} x_i\ ;\ \sharp S = 2n - j \right\} \right)$$

for $j = n, n+1, \ldots, 2n$. Then, it is easy to see that $\partial\Omega$ is the polygon $Q_0 Q_1 \cdots Q_{2n-1} Q_{2n}$ with $Q_{2n} = Q_0$. Let

$$\overrightarrow{PQ_j} = (r_j \cos\theta_j,\ r_j \sin\theta_j) \quad (j = 0, 1, \ldots, 2n - 1)$$

with $r_i \geq 0$ and $\theta_0 < \theta_1 < \cdots < \theta_{2n-1} < \theta_0 + 2\pi =: \theta_{2n}$.

Now we prove the "if" part. Assume that $P \in \Omega \setminus \partial\Omega$. Since P is in the interior of the convex set Ω, we have

$$\tau := \max_{0\leq j\leq 2n-1} \frac{\theta_{j+1}-\theta_j}{2} < \frac{\pi}{2}$$
$$c_0 := \min_{0\leq j\leq 2n-1} r_j > 0.$$

Define

$$\Omega_j := \left\{(r\cos\phi,\ r\sin\phi);\ \frac{\theta_{j-1}+\theta_j}{2} \leq \phi < \frac{\theta_j+\theta_{j+1}}{2}\ ,\ r>0\right\}$$

for $j = 0, 1, \ldots, 2n-1$, where $\theta_{-1} := \theta_{2k-1} - 2\pi$. Then it holds that

$$\bigcup_{j=0}^{2n-1} \Omega_j = \mathbf{R}^2 \setminus \{(0,0)\}$$

and that

$$\begin{aligned}
c &= \iint \frac{\exp(\alpha m+\beta M)}{\prod_{i=1}^n (1+\exp(\alpha+\beta x_i))} d\alpha d\beta \\
&= \iint \frac{\exp(\alpha m+\beta M)}{\sum_S \exp\left(\alpha \sharp S + \beta\sum_{i\in S} x_i\right)} d\alpha d\beta \\
&= \sum_{j=0}^{2n-1} \iint_{\Omega_j} \frac{1}{\sum_S \exp\left((\sharp S - m)\alpha + \left(\sum_{i\in S} x_i - M\right)\beta\right)} d\alpha d\beta \\
&\leq \sum_{j=0}^{2n-1} \iint_{\Omega_j} \exp((m-\alpha_j)\alpha + (M-\beta_j)\beta) d\alpha d\beta.
\end{aligned}$$

Since $|\theta_j - \phi| \leq \tau < \pi/2$ for any $(r\cos\phi,\ r\sin\phi) \in \Omega_j$, we have

$$(\alpha_j - m)\alpha + (\beta_j - M)\beta \geq c_0 r\cos\tau$$

for any $(\alpha, \beta) = (r\cos\phi,\ r\sin\phi) \in \Omega_j$. Thus,

$$c \leq \sum_{j=0}^{2k-1} \iint_{\Omega_j} \exp(-c_0 r\cos\tau) r dr d\phi < \infty.$$

Now we prove the "only if" part. Assume that $P \in \partial\Omega$. That is, P is one of the vertices of the polygon $Q_0 Q_1 \cdots Q_{2n-1} Q_0$. Let $P = Q_j$ and γ be the angle $Q_{j-1} P Q_{j+1}$ in the region of Ω. Then $\gamma \leq \pi$. Therefore, it is possible to take a half line $l = \{(m + r\cos\theta, M + r\sin\theta); r \geq 0\}$ satisfying that $\angle Q_{j-1} P l \leq \pi/2$ and

$\angle Q_{j+1}Pl \leq \pi/2$. This implies that $\angle Q(S)Pl \leq \pi/2$ for any $S \subset \{1, \dots, n\}$ with $Q(S) := (\sharp S, \ \sum_{i \in S} x_i) \neq P$.

Let

$$\Gamma := \{(\alpha, \beta) \in \mathbf{R}^2; |\alpha \sin\theta - \beta\cos\theta| \leq 1, \ \alpha\cos\theta + \beta\sin\theta < 0\}.$$

Then, for any $(\alpha, \beta) \in \Gamma$ and $S \subset \{1, \dots, n\}$, it hollds that

$$\alpha(u - m) + \beta(v - M) \leq \rho,$$

where $(u, v) := Q(S)$ and ρ is the diameter of Ω. Thus, we have

$$\begin{aligned} c &= \iint \frac{\exp(\alpha m + \beta M)}{\prod_{i=1}^n (1 + \exp(\alpha + \beta x_i))} d\alpha d\beta \\ &= \iint \frac{\exp(\alpha m + \beta M)}{\sum_S \exp\left(\alpha\sharp S + \beta\sum_{i\in S} x_i\right)} d\alpha d\beta \\ &= \iint \frac{1}{\sum_S \exp(\alpha(u - m) - \beta(v - M))} d\alpha d\beta \\ &\geq \iint_\Gamma \frac{1}{\sum_S \exp(\alpha(u - m) - \beta(v - M))} d\alpha d\beta \\ &\geq \iint_\Gamma \frac{1}{2^n e^\rho} d\alpha d\beta \\ &= 2^{-n} e^{-\rho} \iint_\Gamma d\alpha d\beta = \infty \ . \end{aligned}$$

EXAMPLE 1. We consider the case where $x_1 = x_2 = \cdots = x_{n_1} = u \neq v = x_{n_1+1} = x_{n_1+2} = \cdots = x_{n_1+n_2}$ and

$$\sum_{i=1}^{n_1} y_i = m_1 \quad , \quad \sum_{i=n_1+1}^{n_1+n_2} y_i = m_2$$

with $0 < m_1 < n_1$ and $0 < m_2 < n_2$. Then we have

$$\begin{aligned} c &= \iint \frac{\exp(m_1(\alpha + u\beta))}{(1 + \exp(\alpha + u\beta))^{n_1}} \frac{\exp(m_2(\alpha + v\beta))}{(1 + \exp(\alpha + v\beta))^{n_2}} d\alpha d\beta \\ &= \frac{1}{|u - v|} B(n_1 - m_1, m_1) B(n_2 - m_2, m_2). \end{aligned}$$

3. Proof of Theorem 2

Note that

$$\frac{\partial G_n}{\partial \alpha} = \sum_{i=1}^{k} \left(\frac{t_i}{n} - 1 + \frac{1}{1+\exp(\alpha+\beta x_i)} \right) =: g_1(\alpha, \beta)$$

$$\frac{\partial G_n}{\partial \beta} = \sum_{i=1}^{k} \left(x_i \left(\frac{t_i}{n} - 1 \right) + \frac{x_i}{1+\exp(\alpha+\beta x_i)} \right) =: g_2(\alpha, \beta).$$

Since

$$\frac{\partial g_1(\alpha,\ \beta)}{\partial \alpha} = -\sum_{i=1}^{k} \frac{\exp(\alpha+\beta x_i)}{(1+\exp(\alpha+\beta x_i))^2} < 0$$

$$g_1(-\infty,\ \beta) = \sum_{i=1}^{k} \frac{t_i}{n_i} > 0$$

$$g_1(\infty,\ \beta) = \sum_{i=1}^{k} \left(\frac{t_i}{n_i} - 1 \right) < 0$$

for any α, β, there exists a unique $\overline{\alpha} = \overline{\alpha}(\beta)$ for any β such that $g_1(\overline{\alpha}, \beta) \equiv 0$.

Then since

$$\begin{aligned}\frac{d\overline{\alpha}}{d\beta} &= -\frac{\partial g_1/\partial \beta}{\partial g_1/\partial \alpha} \\ &= -\frac{\sum_{i=1}^{k} \left\{ x_i \exp(\overline{\alpha}+\beta x_i)/[1+\exp(\overline{\alpha}+\beta x_i)]^2 \right\}}{\sum_{i=1}^{k} \left\{ \exp(\overline{\alpha}+\beta x_i)/[1+\exp(\overline{\alpha}+\beta x_i)]^2 \right\}},\end{aligned}$$

we have

$$\begin{aligned}\frac{dg_2(\overline{\alpha}, \beta)}{d\beta} &= \frac{\partial g_2}{\partial \alpha}\frac{d\overline{\alpha}}{d\beta} + \frac{\partial g_2}{\partial \beta} \\ &= \left(\sum_{i=1}^{k} \frac{\exp(\overline{\alpha}+\beta x_i)}{[1+\exp(\overline{\alpha}+\beta x_i)]^2} \right)^{-2} \\ &\quad \times \left\{ \left(\sum_{i=1}^{k} \frac{\exp(\overline{\alpha}+\beta x_i)}{[1+\exp(\overline{\alpha}+\beta x_i)]^2} \right) \left(\sum_{i=1}^{k} \frac{{x_i}^2 \exp(\overline{\alpha}+\beta x_i)}{[1+\exp(\overline{\alpha}+\beta x_i)]^2} \right) \right. \\ &\quad \left. - \left(\sum_{i=1}^{k} \frac{x_i \exp(\overline{\alpha}+\beta x_i)}{[1+\exp(\overline{\alpha}+\beta x_i)]^2} \right)^2 \right\} \\ &< 0\end{aligned}$$

by the Cauchy-Schwarz inequality.

We consider $\overline{\alpha}/\beta$ as $\beta \to \infty$. Let $p \in [-\infty, \ +\infty]$ be any one of limit points of $\overline{\alpha}/\beta$ as $\beta \to \infty$. We denote by $\lim_{\beta * \to \infty}$ the limit as $\beta \to \infty$ along a subset such that $\overline{\alpha}/\beta \to p$.

Case 1: If $-p < x_1$, then

$$\begin{aligned} 0 &= \lim_{\beta * \to \infty} g_1(\overline{\alpha}, \ \beta) \\ &= \sum_{i=1}^{k} \left(\frac{t_i}{n} - 1 \right) < 0, \end{aligned}$$

which is absurd.

Case 2: If $-p > x_k$, then

$$\begin{aligned} 0 &= \lim_{\beta * \to \infty} g_1(\overline{\alpha}, \ \beta) \\ &= \sum_{i=1}^{k} \frac{t_i}{n} > 0, \end{aligned}$$

which is absurd.

Case 3: If there exists x_{i_0} such that $x_{i_0} < -p < x_{i_0+1}$, then we have

$$\begin{aligned} 0 &= \lim_{\beta * \to \infty} g_1(\overline{\alpha}, \ \beta) \\ &= \sum_{i=i_0+1}^{k} \left(\frac{t_i}{n} - 1 \right) + \sum_{i=1}^{i_0} \frac{t_i}{n} \ . \end{aligned}$$

Hence,

$$\begin{aligned} \lim_{\beta * \to \infty} g_2(\overline{\alpha}, \ \beta) &= \sum_{i=i_0+1}^{k} x_i \left(\frac{t_i}{n} - 1 \right) + \sum_{i=1}^{i_0} x_i \frac{t_i}{n} \\ &< x_{i_0} \left[\sum_{i=i_0+1}^{k} \left(\frac{t_i}{n} - 1 \right) + \sum_{i=1}^{i_0} \frac{t_i}{n} \right] = 0. \end{aligned}$$

Case 4: If $p = x_{i_0}$ for some $i_0 = 1, 2, \ldots, k$, then

$$\begin{aligned} 0 &= \lim_{\beta * \to \infty} g_1(\overline{\alpha}, \ \beta) \\ &= \sum_{i=i_0+1}^{k} \left(\frac{t_i}{n} - 1 \right) + \sum_{i=1}^{i_0-1} \frac{t_i}{n} + \lim_{\beta * \to \infty} \frac{1}{1 + \exp(\overline{\alpha} + p\beta)} \ . \end{aligned}$$

Hence,

$$\lim_{\beta*\to\infty} g_2(\overline{\alpha},\ \beta) =$$

$$\sum_{i=i_0+1}^{k} x_i\left(\frac{t_i}{n}-1\right)+\sum_{i=1}^{i_0-1} x_i\frac{t_i}{n}+x_{i_0}\lim_{\beta*\to\infty}\frac{1}{1+\exp(\overline{\alpha}+p\beta)}$$

$$< x_{i_0}\left[\sum_{i=i_{i_0}+1}^{k}\left(\frac{t_i}{n}-1\right)+\sum_{i=1}^{i_0-1}\frac{t_i}{n}+\lim_{\beta*\to\infty}\frac{1}{1+\exp(\overline{\alpha}+\lambda\beta)}\right]=0.$$

Thus, $\lim_{\beta*\to\infty} g_2(\overline{\alpha},\ \beta) < 0$.

In the same way, we can prove that $\lim_{\beta**\to-\infty} g_2(\overline{\alpha},\ \beta) > 0$. Therefore, there exists a unique $\hat{\beta}_n$ such that $g_2(\overline{\alpha},\ \hat{\beta}_n) = 0$. Putting $\hat{\alpha}_n = \overline{\alpha}(\hat{\beta}_n)$, we have proved that $(\hat{\alpha}_n, \hat{\beta}_n)$ is the unique point which maximizes the function $G_n(\alpha, \beta)$.

4. Proof of Theorem 3

The unique existence of $(\hat{\alpha}, \hat{\beta})$ can be proved exactly in the same way as for that of $(\hat{\alpha}_n, \hat{\beta}_n)$.

Let us take $\delta > 0$ and n_0 such that for any $n \geq n_0$,

$$\delta \leq \frac{t_i}{n} \leq 1-\delta \quad (i = 1, \dots, k).$$

Lemma 1. *Let*

$$\varphi(x, p) := px - \log(1+e^x)$$

be a function on $x \in \mathbf{R}$ and $p \in \mathbf{R}$ with $0 < \delta \leq p \leq 1-\delta < 1$ for some $\delta > 0$. Then, we have

(i)
$$\begin{aligned}\max_{x\in\mathbf{R}}\varphi(x,p) &= p\log p+(1-p)\log(1-p)\\ &\leq \delta\log\delta+(1-\delta)\log(1-\delta)<0,\end{aligned}$$

(ii)
$$\max_{\delta\leq p\leq 1-\delta}\varphi(x,p)\leq -\delta|x|$$

and

(iii)
$$\left|\frac{\varphi(x,p')}{\varphi(x,p)}-1\right| \leq C|p'-p| \textit{ for some constant } C>0.$$

Proof. (i) Since

$$\frac{\partial\varphi}{\partial x} = p-1+\frac{1}{1+e^x}$$

is a monotone decreasing function in x and takes value 0 at $x = \log p - \log(1-p)$, we have

$$\begin{aligned}\max_{x\in\mathbf{R}} \varphi(x, p) &= \varphi(\log p - \log(1-p), p)\\ &= p\log p + (1-p)\log(1-p)\\ &\le \delta\log\delta + (1-\delta)\log(1-\delta) < 0.\end{aligned}$$

(ii) For any $x \ge 0$, we have

$$\varphi(x, p) \le px - \log e^x \le -\delta x.$$

On the other hand, for any $x < 0$, we have

$$\varphi(x, p) \le px \le \delta x.$$

Thus we have (ii).

(iii) Since

$$\left|\frac{\partial \log\varphi}{\partial p}\right| = \left|\frac{x}{\varphi}\right| \le \frac{1}{\delta}$$

by (ii), we have

$$|\log\varphi(x, p') - \log\varphi(x, p)| \le \frac{1}{\delta}|p' - p|,$$

which implies (iii). □

Lemma 2. *For any $x_i \ne x_j$, there exists a constant $C > 0$ such that*

$$(\alpha + \beta x_i)^2 + (\alpha + \beta x_j)^2 \ge C(\alpha^2 + \beta^2)$$

holds for any α and β.

Proof. We have

$$\begin{aligned}&(\alpha + \beta x_i)^2 + (\alpha + \beta x_j)^2\\ &\quad = 2\left(\alpha + \beta\frac{x_i + x_j}{2}\right)^2 + 2\left(\beta\frac{x_i - x_j}{2}\right)^2\\ &\quad \ge C_1\beta^2\end{aligned}$$

and

$$(\alpha + \beta x_i)^2 + (\alpha + \beta x_j)^2$$

$$
\begin{aligned}
&\geq \frac{x_j^2}{x_i^2+x_j^2}(\alpha+\beta x_i)^2 + \frac{x_i^2}{x_i^2+x_j^2}(\alpha+\beta x_j)^2 \\
&= \frac{(\alpha x_j+\beta x_i x_j)^2+(\alpha x_i+\beta x_i x_j)^2}{x_i^2+x_j^2} \\
&= \frac{2\{\alpha(x_i-x_j)/2\}^2+2\{\alpha(x_i+x_j)/2+\beta x_i x_j\}^2}{x_i^2+x_j^2} \\
&\geq C_2\alpha^2
\end{aligned}
$$

with some positive constants C_1 and C_2. Thus we have

$$(\alpha+\beta x_i)^2+(\alpha+\beta x_j)^2 \geq C(\alpha^2+\beta^2)$$

with $C := (1/2)\min\{C_1, C_2\} > 0$. □

Lemma 3. *There exists a constant $D > 0$ such that*

$$G_n(\alpha,\beta) \leq -D(\alpha^2+\beta^2)^{1/2}$$

for any $n \geq n_0$ and $(\alpha,\beta) \in \mathbf{R}^2$.

Proof. Since

$$G_n(\alpha,\beta) = \sum_{i=1}^{k} \varphi\left(\alpha+\beta x_i, \frac{t_i}{n}\right),$$

where φ is defined in Lemma 1, we have

$$
\begin{aligned}
G_n(\alpha,\beta) &\leq -\delta(|\alpha+\beta x_i|+|\alpha+\beta x_j|) \\
&\leq -\delta\{(\alpha+\beta x_i)^2+(\alpha+\beta x_j)^2\}^{1/2} \\
&\leq -\delta C(\alpha^2+\beta^2)^{1/2} \\
&= -D(\alpha^2+\beta^2)^{1/2}
\end{aligned}
$$

with $D = \delta C$ by Lemmas 1 and 2. □

Now we shall complete the proof of Theorem 3, since

$$G_n(0,0) = -k\log 2$$

and by Lemma 3, for any (α,β) with $\alpha^2+\beta^2 > (k\log 2/C)^2$

$$G_n(\alpha,\beta) < -k\log 2,$$

it holds that

$$\hat{\alpha}_n^2 + \hat{\beta}_n^2 \le \left(\frac{k \log 2}{C}\right)^2.$$

Since G_n converges to f uniformly in any bounded region as $n \to \infty$, for any subsequence $\{n'\}$ of $\{n\}$ such that

$$\alpha^* := \lim_{n' \to \infty} \hat{\alpha}_{n'}\ , \quad \beta^* := \lim_{n' \to \infty} \hat{\beta}_{n'}$$

exist, it holds that

$$\begin{aligned} \lim_{n' \to \infty} G_n(\hat{\alpha}_{n'}, \hat{\beta}_{n'}) &= \lim_{n' \to \infty} f(\hat{\alpha}_{n'}, \hat{\beta}_{n'}) \\ &= f(\alpha^*, \beta^*) \ \le \ f(\hat{\alpha}, \hat{\beta}). \end{aligned}$$

On the other hand, since

$$\begin{aligned} & |f(\hat{\alpha}, \hat{\beta}) - G_n(\hat{\alpha}_n, \hat{\beta}_n)| \\ = & \left| \max_{\alpha^2+\beta^2 \le (k \log 2/D)^2} f(\alpha, \beta) - \max_{\alpha^2+\beta^2 \le (k \log 2/D)^2} G_n(\alpha, \beta) \right| \\ \le & \sup_{\alpha^2+\beta^2 \le (k \log 2/D)^2} |f(\alpha, \beta) - G_n(\alpha, \beta)| \to 0 \end{aligned}$$

as $n \to \infty$, $f(\alpha^*, \beta^*) = f(\hat{\alpha}, \hat{\beta})$. The uniqueness of the (α, β) which maximizes $f(\alpha, \beta)$ implies that $(\alpha^*, \beta^*) = (\hat{\alpha}, \hat{\beta})$. This also implies that $\hat{\alpha}_n \to \hat{\alpha}$ and $\hat{\beta}_n \to \hat{\beta}$ as $n \to \infty$, which completes the proof.

EXAMPLE 2. For Example 1, we have

$$\begin{aligned} \hat{\alpha}_n &= \frac{v}{v-u} \log \frac{m_1}{n_1 - m_1} + \frac{u}{u-v} \log \frac{m_2}{n_2 - m_2} \\ \hat{\beta}_n &= \frac{1}{u-v} \log \frac{m_1(n_2 - m_2)}{m_2(n_1 - m_1)} \ . \end{aligned}$$

5. Proof of Theorem 4

Lemma 4. *It holds that*

$$\sum_{i=1}^{k} \varphi\left(\alpha + \beta x_i, \frac{t_i}{n}\right) = f(\alpha, \beta)(1 + O(\delta_n))$$

where $\delta_n := \max_i |(t_i/n) - p_i|$ *and* $O(\delta_n)$ *is uniform in* α *and* β *as* $n \to \infty$.

Proof. Take $\delta > 0$ such that $2\delta < \min_i p_i$ and $\max_i p_i + 2\delta < 1$. Then by (1), there exists n_0 such that for any $n \geq n_0$, it holds that

$$\left|\frac{t_i}{n} - p_i\right| < \delta \quad (i = 1, \dots, k).$$

Then by (iii) of Lemma 1, there exists a constant C such that

$$\varphi\left(\alpha + \beta x_i, \frac{t_i}{n}\right) = \varphi(\alpha + \beta x_i\ , p_i)(1 + \xi_{i,n})$$

with $|\xi_{i,n}| \leq C|(t_i/n) - p_i|$ for any $i = 1, \dots, k$. Therefore, we have

$$\sum_{i=1}^{k} \varphi\left(\alpha + \beta x_i, \frac{t_i}{n}\right) = f(\alpha, \beta)(1 + \xi_n)$$

with

$$|\xi_n| \leq C \max_i \left|\frac{t_i}{n} - p_i\right| = O(\delta_n). \qquad \square$$

To prove Theorem 4, it is sufficient to prove that for any given $\varepsilon > 0$,

$$\lim_{n\to\infty} \iint_{(\hat\alpha-\varepsilon,\hat\alpha+\varepsilon)\times(\hat\beta-\varepsilon,\hat\beta+\varepsilon)} p(\alpha, \beta | t_1, \dots, t_k) d\alpha d\beta = 1.$$

Note that

$$p(\alpha, \beta | t_1, \dots, t_k) = c_n{}^{-1} \exp\left[n \sum_{i=1}^{k} \varphi\left(\alpha + \beta x_i, \frac{t_i}{n}\right)\right]$$

with

$$c_n := \iint \exp\left[n \sum_{i=1}^{k} \varphi\left(\alpha + \beta x_i, \frac{t_i}{n}\right)\right] d\alpha d\beta.$$

By Theorem 3, Lemmas 1 and 3,

$$\max_{(\alpha,\beta)\in\mathbf{R}^2} f(\alpha, \beta) = f(\hat\alpha, \hat\beta) < 0$$

$$\lim_{\alpha^2+\beta^2\to\infty} f(\alpha, \beta) = -\infty. \tag{1}$$

For any $\Delta > 0$, let

$$\Omega(\Delta) := \{(\alpha, \beta) \in \mathbf{R}^2; f(\alpha, \beta) > \Lambda - \Delta\},$$

where we put $\Lambda := f(\hat{\alpha}, \hat{\beta})$. Since by Theorem 3, $(\hat{\alpha}, \hat{\beta})$ is the unique point which maximizes f together with (3) and the fact that f is continuous, we can take Δ such that

$$\Omega(5\Delta) \subset (\hat{\alpha} - \varepsilon, \hat{\alpha} + \varepsilon) \times (\hat{\beta} - \varepsilon, \hat{\beta} + \varepsilon). \tag{2}$$

Since $\Omega(\Delta)$ is a nonempty bounded open set, it has a positive area, say $S > 0$. Moreover, by (1) and Lemma 4, there exists n_1 such that for any $n \geq n_1$ and $(\alpha, \beta) \in \Omega(\Delta)$,

$$\sum_{i=1}^{k} \varphi\left(\alpha + \beta x_i, \frac{t_i}{n}\right) > \Lambda - 2\Delta.$$

Hence for any $n \geq n_1$, we have

$$\iint_{\Omega(\Delta)} \exp\left[n \sum_{i=1}^{k} \varphi\left(\alpha + \beta x_i, \frac{t_i}{n}\right)\right] d\alpha d\beta \geq e^{(\Lambda - 2\Delta)n} S. \tag{3}$$

On the other hand, by (1), (2), (3) and Lemma 1, there exists n_2 such that for any $n \geq n_2$ and $(\alpha, \beta) \notin \Omega(5\Delta)$,

$$\sum_{i=1}^{k} \varphi\left(\alpha + \beta x_i, \frac{t_i}{n}\right) < \Lambda - 4\Delta.$$

Also by (1), Lemmas 3 and 4, there exists n_3 such that for any $n \geq n_3$ and $(\alpha, \beta) \in \mathbf{R}^2$,

$$\sum_{i=1}^{k} \varphi\left(\alpha + \beta x_i, \frac{t_i}{n}\right) < \frac{1}{2} f(\alpha, \beta) \leq -\frac{1}{2} C(\alpha^2 + \beta^2)^{1/2}.$$

Hence, for any η with $0 < \eta < 1$, $(\alpha, \beta) \notin \Omega(5\Delta)$, and $n \geq n_4 := n_2 \vee n_3$ we have

$$\sum_{i=1}^{k} \varphi\left(\alpha + \beta x_i, \frac{t_i}{n}\right) \leq -\frac{1}{2} C\eta(\alpha^2 + \beta^2)^{1/2} + (1 - \eta)(\Lambda - 4\Delta).$$

Therefore, taking a small $\eta > 0$ such that

$$(1 - \eta)(\Lambda - 4\Delta) < \Lambda - 3\Delta,$$

we have

$$\sum_{i=1}^{k} \varphi\left(\alpha + \beta x_i, \frac{t_i}{n}\right) \leq -C'(\alpha^2 + \beta^2)^{1/2} + \Lambda - 3\Delta$$

for any $(\alpha, \beta) \notin \Omega(5\Delta)$ and $n \geq n_4$ with some constant $C' > 0$. Hence, we have

$$\begin{aligned}
&\iint\limits_{\mathbf{R}^2 \setminus \Omega(5\Delta)} \exp\left[n \sum_{i=1}^{k} \varphi\left(\alpha + \beta x_i, \frac{t_i}{n}\right)\right] d\alpha d\beta \\
&\qquad \leq \iint \exp\left[-C'n(\alpha^2 + \beta^2)^{1/2} + (\Lambda - 3\Delta)n\right] d\alpha d\beta \\
&\qquad \leq e^{(\Lambda - 3\Delta)n} \iint \exp\left[-C'(\alpha^2 + \beta^2)^{1/2}\right] d\alpha d\beta \\
&\qquad \leq C'' e^{(\Lambda - 3\Delta)n} \qquad (4)
\end{aligned}$$

for any $n \geq n_4$ with some constant $C'' > 0$.

Let

$$I_n := \iint\limits_{(\hat{\alpha}-\varepsilon, \hat{\alpha}+\varepsilon) \times (\hat{\beta}-\varepsilon, \hat{\beta}+\varepsilon)} p(\alpha, \beta | t_1, \ldots, t_k) d\alpha d\beta.$$

Then by (4), we have

$$\begin{aligned}
I_n &\geq \iint\limits_{\Omega(5\Delta)} p(\alpha, \beta | t_1, \ldots, t_k) d\alpha d\beta \\
&= c_n{}^{-1} \iint\limits_{\Omega(5\Delta)} \exp\left[n \sum_{i=1}^{k} \varphi\left(\alpha + \beta x_i, \frac{t_i}{n}\right)\right] d\alpha d\beta.
\end{aligned}$$

Putting

$$J(j) := \iint\limits_{\Omega(j\Delta)} \exp\left[n \sum_{i=1}^{k} \varphi\left(\alpha + \beta x_i, \frac{t_i}{n}\right)\right] d\alpha d\beta \qquad (5)$$

and

$$L(j) := \iint\limits_{\mathbf{R}^2 \setminus \Omega(j\Delta)} \exp\left[n \sum_{i=1}^{k} \varphi\left(\alpha + \beta x_i, \frac{t_i}{n}\right)\right] d\alpha d\beta, \qquad (6)$$

we have

$$\begin{aligned}
I_n &\geq c_n{}^{-1} J(5) = \frac{J(5)}{J(5) + L(5)} \\
&\geq \frac{J(1)}{J(1) + L(5)} = \frac{1}{1 + \{L(5)/J(1)\}}.
\end{aligned}$$

Let $n_0 := n_1 \vee n_4$. Then, for any $n \geq n_0$, we have by (5) and (6) that

$$J(1) \geq e^{(\Lambda-2\Delta)n} S \text{ and } L(5) \leq C'' e^{(\Lambda-3\Delta)n}.$$

Thus,

$$I_n \geq \frac{1}{1 + C'' S^{-1} e^{-\Delta n}}$$

from which $\lim_{n\to\infty} I_n = 1$ follows. □

Lehmann gave conditions B(1)–B(4) for the asymptotic normality in [5]. The condition B(1) follows Theorem 4, the other conditions B(2)–B(4) are verified easily. Thus we have Corolloary 1.

The author wishes to thank his supervisor Professor Teturo Kamae (Osaka City University) for his useful suggestions and discussions with him. The author also thanks the referee for useful advices.

References

[1] D.R. Cox and E.J. Snell: Analysis of Binary Data, CHAPMAN & HALL, 1994.
[2] D.R. Cox and D.V. Hinkley: Theoretical Statistics, CHAPMAN & HALL, 1973.
[3] M.A. Tanner: Tools for Statistical Inference, Lecture Notes in Stastistics **67**, Springer-Verlag, 1991.
[4] P. McCullagh and J.A. Nelder: Generalized Linear Models, CHAPMAN & HALL, 1996.
[5] E.L. Lehmann: Theory of Point, JOHN WILEY, 1983.
[6] P. McCullagh: *Regression Models for ordinal* (S.J. Haberman, Discussion of Dr. McCullagh's Paper), J. R Statist SocB, **42(2)** (1980), 109–142.
[7] J.W. Pratt: *Concavity of the log likelihood*, J. Amer. Stat. Association, **76** (1981), 103–106.
[8] H. Strasser: *Consistency of maximum likelihood and bayes estimates*, The Annals of Statistics, **9**-5 (1981), 1107–1113.
[9] D.R. Cox: *Regression models and life-tables* (*with discussion*), J. Roy. Soc. Ser. B, **34** (1972), 187–220.
[10] A. Ibragimov and R.Z. Khas'Minskii: *Asymptotic behavior of some statsctical estimators II. Limit theorems for the a posteriori density and bayes'estimators*, Theory of probability and its applications, **18** (1973), 76–91.
[11] V.T. Farewell: *Some results on the estimation of logistic model based on retrospective data*, Biometrika, **66(1)** (1979), 27–32.

Department of Mathematics
Osaka City University
Sugimoto 3-3-138, Sumiyoshi-ku
OSAKA 558-8585, Japan