STATISTICAL RESEARCH REPORT
Institute of Mathematics
University of Oslo

No 5
July 1977

ON THE ESTIMATION OF PARAMETERS IN A BERNOULLI MODEL WITH DEPENDENCE

by

BO LINDQVIST

ABSTRACT

A model for Bernoulli trials with dependence is developed by Klotz (1972, 1973). The sequence $X_1, X_2, \ldots$ is considered as a stationary Markov chain on $\{0,1\}$, determined by the parameters $p = Pr(X_i = 1)$ and $c = \rho(X_{i-1}, X_i)$. The present paper shows that the maximum likelihood estimator $(\hat{p}, \hat{c})$ derived under the Markov chain assumption is strongly consistent for (p,c) and asymptotically normally distributed under quite weaker assumtions. The proof is based on a paper of Ranneby (1975). Finally, it is proved that $(\hat{p}, \hat{c})$ is asymptotically equivalent to $(\bar{X}, \tilde{c})$, where $\tilde{c}$ is the empirical correlation coefficient.

Key words: Bernoulli trials, dependence, maximum likelihood estimator, strong consistency, asymptotic distribution, asymptotic equivalence.

1. Introduction

A model for Bernoulli trials with Markov dependence is developed by Klotz (1972, 1973). The model involves the usual frequency parameter p and an additional dependence parameter λ. It is assumed that $X_1, X_2, \ldots, X_n$ is a Markov chain on $\{0,1\}$ such that $\Pr(X_1 = 1) = 1 - \Pr(X_1 = 0) = p$; with transition matrix

$$\Pi = \begin{pmatrix} (1-2p+\lambda p)/q & (1-\lambda)p/q \\ 1-\lambda & \lambda \end{pmatrix}$$

where $q = 1-p$. Since the stationary distribution of this chain is (q,p), we have $\Pr(X_i = 1) = 1-\Pr(X_i = 0) = p$ for $i = 1,2,\ldots,n$.

It is seen that $\lambda = p$ corresponds to independent trials, whereas $\lambda > p$ and $\lambda < p$ correspond to, respectively, clustering and lack of clustering among the ones and zeros.

Let c denote the correlation coefficient of X_1 and X_2. It is readily verified that

$$(1.1) \qquad c = (\lambda-p)/(1-p)$$

and that $\rho(X_i, X_j) = c^{|i-j|}$ for $1 \leq i,j \leq n$.

It follows from (1.1) that c may replace λ as the parameter measuring the clustering effect. Since c has a simple interpretation, we will throughout this paper use c instead of λ. The transition matrix Π may now be written on the more "symmetric" form

$$\Pi = \begin{pmatrix} q+cp & p(1-c) \\ q(1-c) & p+cq \end{pmatrix}$$

In order that all entries of Π are nonnegative we must require

(1.2) $\max(1-1/p,\ 1-1/q) \leq c \leq 1$

The natural parameter set of our model is thus the set of pairs (p,c) such that $0 < p < 1$ and such that (1.2) holds.

Klotz (1973) and Devore (1976) consider maximum likelihood estimation of (p,λ) using the method of Billingsley (1961). By the same method we shall derive the maximum likelihood estimator (M.L.E) of (p,c) .

Let N_{ij} be the number of indices k for which $X_{k-1} = i$ and $X_k = j$ $(k = 2,3,\ldots,n)$, so that $N_{00} + N_{10} + N_{01} + N_{11} = n-1$ Billingsley (1961) considers a modified likelihood function, neglecting the first term of the full likelihood. If $\Pi = (\pi_{ij})$ then it follows that the (modified) M.L.E. of π_{00} and π_{11} are

$$\pi^*_{00} = N_{00}/(N_{00} + N_{01}) \quad \text{and}$$
$$\pi^*_{11} = N_{11}/(N_{10} + N_{11})$$

Solving the system

$$\pi_{00} = q+cp$$
$$\pi_{11} = p+cq$$

we get

(1.3) $$p^* = (1-\pi^*_{00})/(2-\pi^*_{00} - \pi^*_{11})$$
$$c^* = \pi^*_{00} + \pi^*_{11} - 1$$

Let $N_1 = \sum_{i=1}^{n} X_i$. Then $N_1 - (N_{10} + N_{11}) = X_n$ and hence it is seen that π_{11}^* and $\hat{\pi}_{11} = N_{11}/N_1$ are asymptotically equivalent. If $N_0 = n - N_1$, then by symmetry also π_{00}^* and $\hat{\pi}_{00} = N_{00}/N_0$ are asymptotically equivalent. For matters of convenience we shall in the sequel consider the estimators $\hat{p}$ and $\hat{c}$ obtained by replacing $*$ by $\hat{}$ in (1.3).

The asymptotic distribution of $n^{\frac{1}{2}}(\hat{p} - p, \hat{c} - c)$ may be found directly from the theory of Billingsley (1961). However, we shall in section 2 derive the asymptotic distribution under a weaker assumption than Markov-dependence. We shall also show that $\hat{p}$ and $\hat{c}$ are strongly consistent estimators of p and c .

2. Non-Markovian dependence

As in section 1 we shall let the process $X_1, X_2, \ldots$ be stationary with

$$\Pr(X_1 = 1) = 1 - \Pr(X_1 = 0) = p$$

and

$$\rho(X_1, X_2) = c$$

We shall in the remainder of this paper be concerned with the following additional assumptions.

A1: $X_1, X_2, \ldots$ is a Markov chain.

A2: $X_1, X_2, \ldots$ is α-mixing with $\sum_{i=1}^{\infty} \alpha_X(i) < \infty$ (for definition, see e.g. Ranneby, 1975)

A3: $X_1, X_2, \ldots$ is an ergodic process.

Clearly A1 $\Rightarrow$ A2 $\Rightarrow$ A3.

<u>Theorem 2.1</u>

Let $\hat{p}$ and $\hat{c}$ be defined as in section 1. Then, under assumption A3,

$$\hat{p} \xrightarrow{a.s.} p \qquad \text{as } n \to \infty$$
$$\hat{c} \xrightarrow{a.s.} c \qquad \text{as } n \to \infty$$

i.e. $\hat{p}$ and $\hat{c}$ are strongly consistent estimators.

<u>Proof:</u> From Ranneby (1975) follows that

$$\hat{\pi}_{11} \xrightarrow{a.s.} Pr(X_2 = 1 | X_1 = 1)$$

But $Pr(X_2 = 1 | X_1 = 1) = Pr(X_1 = 1,\ X_2 = 1)/p$

$$= (c \cdot pq + p^2)/p = p + cq = \pi_{11}$$

By symmetry, $\hat{\pi}_{00} \xrightarrow{a.s.} \pi_{00}$. Now the theorem follows from (1.3). Q.E.D.

We now turn to the derivation of the asymptotic distribution of $n^{\frac{1}{2}}(\hat{p} - p,\ \hat{c} - c)$. If A2 holds, then by the result of Ranneby (1975),

$n^{\frac{1}{2}}(\hat{\pi}_{00} - \pi_{00}, \hat{\pi}_{11} - \pi_{11})$ is asymptotically normal with mean $(0,0)$ and convariance matrix $\Sigma = \Sigma^0 + \Sigma^1$, where Σ^0 and Σ^1 are given below.

$$\Sigma^0 = \begin{pmatrix} (1-c)(q+cp)p/q & 0 \\ 0 & (1-c)(p+cq)q/p \end{pmatrix}$$

and $\Sigma^1 = (\sigma_{ij}^{\ 1})$ where

$$\sigma_{ij}^{1} = p_i^{-1} p_j^{-1} \sum_{v=2}^{\infty} b_v(i,j);$$

$p_o = q,\ p_1 = p$ and

$$\begin{aligned} b_v(i,j) &= \Pr(X_1 = i,\ X_2 = i,\ X_v = j,\ X_{v+1} = j) \\ &+ \Pr(X_1 = j,\ X_2 = j,\ X_v = i,\ X_{v+1} = i) \\ &- \pi_{jj}[\Pr(X_1 = i,\ X_2 = i,\ X_v = j) + \Pr(X_1 = j,\ X_v = i,\ X_{v+1} = i)] \\ &- \pi_{ii}[\Pr(X_1 = j,\ X_2 = j,\ X_v = i) + \Pr(X_1 = i,\ X_v = j,\ X_{v+1} = j)] \\ &+ \pi_{ii}\,\pi_{jj}[\Pr(X_1 = i,\ X_v = j) + \Pr(X_1 = j, X_v = i)] \end{aligned}$$

From this we get:

<u>Theorem 2.2</u>

Under the assumption A2, $n^{\frac{1}{2}}(\hat{p} - p,\ \hat{c} - c)$ is asymptotically normal with mean (0,0) and covariance matrix $\Lambda = \Lambda^o + \Lambda^1$

Here

$$\Lambda^o = \begin{pmatrix} pq(1+c)/(1-c) & c(q-p) \\ c(q-p) & (1-c)(1+c(\frac{q^2}{p} + \frac{p^2}{q})) \end{pmatrix}$$

and

$\Lambda^1 = (\lambda_{ij}^{1})$ is given by

$$\lambda_{oo}^{1} = (q^2\sigma_{oo}^{1} + p^2\sigma_{11}^{1} - 2pq\,\sigma_{o1}^{1})/(1-c)^2$$

$$\lambda_{o1}^{1} = (p\sigma_{11}^{1} - q\sigma_{oo}^{1} + (p-q)\sigma_{o1}^{1})/(1-c)$$

$$\lambda_{11}^{1} = \sigma_{oo}^{1} + \sigma_{11}^{1} + 2\sigma_{o1}^{1}$$

<u>Proof</u>: Write

$$\hat{p} - p = \frac{(\hat{\pi}_{11}-\pi_{11})\pi_{01} - (\hat{\pi}_{00}-\pi_{00})\pi_{10}}{(2-\hat{\pi}_{00}-\hat{\pi}_{11})(2-\pi_{00}-\pi_{11})}$$

$$\hat{c} - c = (\hat{\pi}_{00}-\pi_{00}) + (\hat{\pi}_{11}-\pi_{11})$$

Λ is now easily derived from Σ, using wellknown results from asymptotic theory (see e.g. Billingsley, 1968).

Corollary 2.3.

Under assumption A1, the covariance matrix of $n^{\frac{1}{2}}(\hat{p}-p, \hat{c}-c)$ is given by Λ^0.

Proof: As is noted by Ranneby (1975), $\Sigma^1 = 0$ and hence $\Lambda^1 = 0$ if $X_1, X_2, \ldots$ is a Markov chain.

We close this section by giving an example involving a process $X_1, X_2, \ldots$ which is not a Markov chain.

Example

Let $Y_0, Y_1, \ldots$ be independent, identically distributed with

$$Pr(Y_1 = 1) = 1 - Pr(Y_1 = 0) = p .$$

Let further $X_1, X_2, \ldots$ be constructed as follows:

$$X_i = \begin{cases} Y_i & \text{with probability } 1-\delta \\ Y_{i-1} & \text{" } \quad \text{" } \quad \delta \end{cases}$$

$i = 1,2,\ldots \; ; \; 0 \leq \delta \leq \frac{1}{2}$

The process $X_1, X_2, \ldots$ is stationary and clearly satisfies assumption A2. (In fact, the process is 2-dependent). A computation shows that

$$c = \rho(X_1, X_2) = \delta(1-\delta)$$

The asymptotic distribution of $n^{\frac{1}{2}}(\hat{p}-p, \hat{c}-c)$ is given by theorem 2.2, and hence we must compute Σ^1. From the definition of $b_v(i,j)$ it is seen that $b_v(i,j)$ in our case is constant for $v = 4,5,\ldots$ and hence must be 0. The computation of b_2 and b_3 is straightforward, noting that the terms $\Pr(X_1 = i_1, \ldots, X_k = i_k)$ are found by conditioning on $(Y_0, \ldots, Y_k)$. We get

$$\sigma_{00}^1 = -2c^2p^2(1-cp)/q \; ; \; \sigma_{01}^1 = -2c^3pq \; ;$$

$$\sigma_{11}^1 = -2c^2q^2(1-cq)/p \; ; \; \lambda_{00}^1 = -2c^2/(1-c) \; ;$$

$$\lambda_{01}^1 = 2c^2(p-q) \; ; \; \lambda_{11}^1 = 2c^2(c(q-p)^2 - p^3 - q^3)/pq$$

and finally

$$\lambda_{00} = (1+2c)pq$$

$$\lambda_{01} = c(1-2c)(q-p)$$

$$\lambda_{11} = 1-3c^2 + c(1-c)(1-2c)\left(\frac{p^2}{q} + \frac{q^2}{p} - 1\right)$$

It is notable that the entries of Λ^1 are all $O(c^2)$ as $c \to 0$. Hence the distribution of $(\hat{p}, \hat{c})$ is rather robust against this kind of departure from Markov chain dependence.

3. Asymptotically equivalent estimators.

It would seem natural in our model to estimate p by the mean and c by the empirical correlation coefficient of X_{i-1} and X_i. Let therefore $\tilde{p}$ and $\tilde{c}$ be given by

$$\tilde{p} = N_1/n = \bar{X}$$

$$\tilde{c} = \frac{n^{-1}\sum_{i=2}^{n} X_{i-1}X_i - \bar{X}^2}{n^{-1}\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

We shall in this section prove that $(\hat{p},\hat{c})$ and $(\tilde{p},\tilde{c})$ are asymptotically equivalent under assumption A3.

Klotz (1973) proves the asymptotic equivalence of the M.L.E. of (p,λ) and the estimator $(\bar{X},\lambda^*(\bar{X}))$, where $\lambda^*(p)$ is the M.L.E. of λ when p is known. The proof is based on the fact that $\hat{p}$ and $\bar{X}$ have the same asymptotic variance and also utilizes the likelihood function derived under the Markov chain assumption. Hence Klotz' proof is not applicable under more general assumptions. We treat $\tilde{p}$ and $\tilde{c}$ separately. It is assumed throughout that assumption A3 is satisfied.

Theorem 3.1.

$\tilde{p}$ is strongly consistent and asymptotically equivalent to $\hat{p}$.

Proof: We have $\tilde{p} = N_1/n$ and hence strong consistency follows from Ranneby (1975). To prove asymptotical equivalence we must prove that $n^{\frac{1}{2}}(\hat{p} - \tilde{p}) \overset{P}{\to} 0$ as $n \to \infty$.

Now

$$\hat{p} = \frac{N_{01}/N_0}{(N_{01}/N_0) + (N_{10}/N_1)} = \frac{N_{01}N_1}{N_{01}N_1 + N_{10}N_0}$$

Clearly $N_{10} = N_{01} + T$, where $|T| \leq 1$. Hence we can write

$$\hat{p} = \frac{N_1}{n+T\,\frac{N_o}{N_{o1}}}$$

which leads us to the inequality

$$|\hat{p} - \tilde{p}| \leq \frac{N_o}{N_{o1}} \cdot \frac{N_1}{n} \cdot \frac{1}{n - \frac{N_o}{N_{o1}}}$$

Since $\frac{N_o}{N_{o1}} \overset{p}{\to} \frac{1}{\pi_{o1}}$, $\frac{N_1}{n} \overset{p}{\to} p$, (see Ranneby, 1975) it follows from Slutsky's theorem that $n^{\frac{1}{2}}(\hat{p} - \tilde{p}) \overset{p}{\to} 0$.

Theorem 3.2.

$\tilde{c}$ is strongly consistent and asymptotically equivalent to $\hat{c}$.

Proof: A simple computation shows that

$$\tilde{c} = \frac{nN_{11} - N_1^2}{N_o\,N_1} .$$

Hence strong consistency follows from Ranneby (1975).

Since $\hat{c} = \frac{N_1 N_{oo} + N_o N_{11} - N_o N_1}{N_o\,N_1}$ it follows that

$$(3.1) \quad \tilde{c} - \hat{c} = \frac{N_{o1}-N_{10}}{N_o} + \frac{N_o-N_{oo}-N_{o1}}{N_o} - \frac{N_1-N_{11}-N_{10}}{N_o}$$

From the definition of N_i and N_{ij} we can conclude that each term on the right hand side of (3.1) is $\leq 1/N_o$ in absolute value and hence

$$n^{\frac{1}{2}}|\tilde{c} - \hat{c}| \leq 3 \cdot N_o^{-\frac{1}{2}}(n/N_o)^{\frac{1}{2}}$$

But $N_0^{-\frac{1}{2}} \overset{p}{\to} 0$ and $n/N_0 \overset{p}{\to} 1/q$, so $n^{\frac{1}{2}}(\tilde{c} - \hat{c}) \overset{p}{\to} 0$ by Slutsky's theorem. Q.E.D.

Theorems 3.1 and 3.2 together state that the estimator $(\tilde{p},\tilde{c})$ is equivalent to $(\hat{p},\hat{c})$ and hence asymptotically efficient in the Markov chain case.

The asymptotic distribution of $(\tilde{p},\tilde{c})$ under assumption A2 is given by theorem 2.2. It is believed, However, that in some cases a direct derivation of the asymptotic distribution of $(\tilde{p},\tilde{c})$ may be easier than using theorem 2.2. In fact, in the example of section 2, the asymptotic variance of $\bar{X}$ may be found using theorem 18.5.4 of Ibragimov and Linnik (1971) to be

$$\mathrm{Var}X_1 + 2 \sum_{v=2}^{\infty} \mathrm{Cov}(X_1,X_v) = pq + 2\ \mathrm{Cov}(X_1,X_2) = (1+2c)pq$$

However, the derivation of the variance of $\tilde{c}$ and the covariance of $\bar{X}$ and $\tilde{c}$ do not seem to be simplified by a direct approach.

References.

Billingsley, P. (1961). Statistical inference for Markov processes. University of Chicago Press, Chicago.

Billingsley, P. (1968). Convergence of probability measures. Wiley, New York.

Devore, J.L. (1976). A note on the estimation of parameters in a Bernoulli model with dependence. Ann. Statist. 4, 990-992.

Ibragimov, I.A. and Linnik, Yu.V. (1971). Independent and stationary sequences of random variables. Wolters- Noordhoff, Groningen.

Klotz, J. (1972). Markov chain clustering of births by sex. Proc. Sixth Berkeley Symp. Math. Statist. Prob. 4, 173-185, University of California Press.

Klotz, J. (1973). Statistical inference in Bernoulli trials with dependence, Ann Statist., 1, 373-379.

Ranneby, B. (1975). Asymptotic properties of the empirical transition frequencies for a stationary chain with discrete time. Scand. J. Statist. 2, 103-105.