The Annals of Statistics
1976, Vol. 4, No. 5, 990–992

A NOTE ON THE ESTIMATION OF PARAMETERS IN A BERNOULLI MODEL WITH DEPENDENCE

BY JAY L. DEVORE

Oberlin College

A generalization of a Bernoulli process which incorporates a dependence structure was given by Klotz (1972, 1973), in which he considered $X_1, X_2, \cdots, X_n$ as a stationary two-state Markov chain with state space $\{0, 1\}$. The parameters of the process are $p = P(X_i = 1)$ and λ, which measures the degree of persistence in the chain. Klotz was unable to solve the equations arising from the full likelihood for the M.L.E.'s of p and λ, so proposed and investigated an ad hoc procedure. Here explicit solutions are obtained for M.L.E.'s based on a modified likelihood function, where the modification consists of neglecting the first term of the full likelihood. In addition it is observed that Klotz's equations can in fact be solved explicitly.

In two papers by Klotz ([2], [3]) a model for Bernoulli trials with dependence is developed in which $X_1, X_2, \cdots, X_n$ is a stationary two-state (0 and 1) Markov chain with transition matrix

$$(1) \qquad \mathbf{p} = \begin{pmatrix} \dfrac{1 - 2p + \lambda p}{1 - p} & \dfrac{(1 - \lambda)p}{1 - p} \\ 1 - \lambda & \lambda \end{pmatrix}.$$

Here $p = P(X_1 = 1)$ and the parameter λ gives a measure of the degree of clustering or persistence in the chain. If the transition matrix is parametrized as $\mathbf{P} = (p_{ij})$ for $i, j = 0, 1$, then the log likelihood for the realization $x_1, x_2, \cdots, x_n$ can be written as

$$(2) \qquad L = x_1 \log p + (1 - x_1) \log (1 - p) + L'$$

where

$$(3) \qquad L' = n_{00} \log p_{00} + n_{01} \log (1 - p_{00}) + n_{10} \log (1 - p_{11}) + n_{11} \log p_{11} .$$

Here $p = (1 - p_{00})/(2 - p_{00} - p_{11})$, x_1 denotes the state first visited by the chain, and the n_{ij}'s are the usual transition counts given by the number of indices m for which $x_m = i$ and $x_{m+1} = j$ $(m = 1, \cdots, n - 1)$, so that $n_{00} + n_{01} + n_{10} + n_{11} = n - 1$.

After taking the partial derivatives of L with respect to p and λ and equating to 0, the author in [3] states that closed form expressions are difficult to obtain. He notes, however, that the estimate $\tilde{p} = S/n$, where $S = x_1 + \cdots + x_n$, has

Received July 1974; revised February 1976.

AMS 1970 *subject classifications.* Primary 62M05; Secondary 62F10.

Key words and phrases. Markov chain, maximum likelihood estimation, dependent Bernoulli sequence.

the same asymptotic distribution as does the M.L.E. for p and suggests substituting this estimate back into one of his estimating equations to obtain an estimate $\tilde{\lambda}$ for λ; he then shows that $(\tilde{p}, \tilde{\lambda})$ is asymptotically equivalent to the M.L.E.'s.

This approach seems unsatisfactory for several reasons. First, the author of [3] uses the theory in Billingsley ([1], page 14) to show that the M.L.E.'s are asymptotically unbiased and normal with covariance matrix

$$\Lambda = \begin{pmatrix} \lambda(1-\lambda)/p & \lambda(1-p) \\ \lambda(1-p) & p(1-p)(1-2p+\lambda)/(1-\lambda) \end{pmatrix}. \tag{4}$$

Because Billingsley's development of the asymptotic theory of M.L.E.'s is based on the use of the modified log likelihood L', the application of this theory to the estimates in [3] based on the full likelihood L is inconsistent, though the conclusions are still correct. The modification consists of neglecting in the full likelihood the term $p^{x_1}(1-p)^{1-x_1}$, which represents the contribution due to the first state visited by the process. In fact, using L' from the outset, it is easily verified that the modified M.L.E.'s of p_{00} and p_{11} are given by [1] page 26, $\hat{p}_{00} = n_{00}/(n_{00}+n_{01})$ and $\hat{p}_{11} = n_{11}/(n_{10}+n_{11})$. The modified M.L.E.'s for the parameters p and λ are now explicitly obtained by equating p_{00} and p_{11} to the appropriate functions of p and λ given in (1), solving for p and λ as functions of p_{00} and p_{11}, and finally substituting $\hat{p}_{00}$ and $\hat{p}_{11}$ for p_{00} and p_{11} to yield

$$\hat{\lambda} = \hat{p}_{11}, \qquad \hat{p} = (1-\hat{p}_{00})/(2-\hat{p}_{00}-\hat{p}_{11}). \tag{5}$$

The argument of [3] applies a fortiori to yield asymptotic normality of $\hat{\lambda}$ and $\hat{p}$ with limiting covariance matrix given by (4).

Thus to prefer the estimators of [3] to $\hat{p}$ and $\hat{\lambda}$ above amounts to using ad hoc estimators in place of the usual estimators $\hat{p}_{ij} = n_{ij}/(n_{i0}+n_{i1})$. When p has a known value, however, as is assumed in [2], only the model parameter λ is free to vary. In the above derivation, both p_{00} and p_{11} (and thus λ and p) are free to vary, so that when the value of p is specified, the modified M.L.E. of λ is no longer given by (5).

Finally, it is actually not very difficult to solve the system (4.2) in [3]. The system can be written as

$$n_{00}/p_{00} = (n_{01}+x_1+p)/(1-p_{00}), \qquad n_{11}/p_{11} = (n_{10}-x_1-p)/(1-p_{11}). \tag{6}$$

Assuming for the moment that p is known, (6) can be solved to yield

$$\hat{p}_{00} = n_{00}/(n_{00}+n_{01}+x_1-p), \qquad \hat{p}_{11} = n_{11}/(n_{10}+n_{11}-x_1+p), \tag{7}$$

from which we have

$$\frac{1-p}{p} = \frac{1-\hat{p}_{11}}{1-\hat{p}_{00}} = \frac{n_{10}-x_1+p}{n_{10}+n_{11}-x_1+p} \cdot \frac{n_{00}+n_{01}+x_1-p}{n_{01}+x_1-p}. \tag{8}$$

Clearing the fractions in (8) yields

$$\begin{aligned} &p(n_{10}-x_1+p)(n_{00}+n_{01}+x_1-p) \\ &\qquad - (1-p)(n_{10}+n_{11}-x_1+p)(n_{01}+x_1-p) = 0, \end{aligned} \tag{9}$$

which reduces to a quadratic equation in p since the p^3 terms cancel. The l.h.s. of (8) decreases from ∞ to 0 in p while the r.h.s. increases, so there is a unique root in [0, 1] which give the "unmodified" M.L.E. of p.

Acknowledgment. The observation that the system in [3] can be solved explicitly is due entirely to the referee. The author wishes to express his appreciation for a careful reading of the manuscript.

REFERENCES

[1] BILLINGSLEY, PATRICK (1961). *Statistical Inference for Markov Processes.* Univ. of Chicago Press.

[2] KLOTZ, JEROME (1972). Markov chain clustering of births by sex. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **4** 173-185, Univ. of California Press.

[3] KLOTZ, JEROME (1973). Statistical inference in Bernoulli trials with dependence. *Ann. Statist.* **1** 373-379.

DEPARTMENT OF MATHEMATICS
OBERLIN COLLEGE
OBERLIN, OHIO 44074