Journal of the Korean Operations
Research Society
Vol. 6, No. 2, Oct., 1981

—Research Paper—

Estimation of Parameters of a Two-State Markov Process by Interval Sampling

Joong Soon Jang and Do Sun Bai*

Abstract

This paper develops a method of modifying the usual maximum likelihood estimators of the parameters of a two state Markov process when the trajectory of the process can only be observed at regular epochs. The method utilizes the limiting behaviors of the process and the properties of state transition counts. An efficient adaptive strategy to be used together with the modified estimator is also proposed. The properties of the new estimators and the adaptive strategy are investigated using Monte Carlo simulation.

I. Introduction

Consider an alternating renewal process $\{X(t),\ t>0\}$ with states 0 and 1. It is assumed that the process is in state 0 for a time T_0 which has an exponential distribution $F_0(t)=1-\exp(-\lambda t)$ and is then in state 1 for a time T_1 which has an exponential distribution $F_1(t)=1-\exp(-\mu t)$. It returns to state 0 for a time given by $F_0(t)$, then to state 1 for a time given by $F_1(t)$. Such a process is referred to as an alternating Poisson process [4].

The purpose of this paper is to estimate the parameters λ and μ, efficiently. If the trajectory of the process can be observed continuously, the parameter estimation is straightforward, for it is possible to write down the likelihood explicitly. In many situations, however, it is not possible to observe the process continuously; rather observations are taken at regular or irregular epochs. We are interested in the sampling at regular intervals. This might arise, for example, if the process were used to represent the use or idle time of an industrial facility; sampling may take place only once an hour or once a day. A random telegraph signal whose values are $+1$ and -1 successively, and a model for semiconductor which can be either in a trapped state or in a free state are also applications.

Suppose that $X(t)$ is the state of the process at time t, and observations are taken at equal intervals Δ. We then obtain a sample $X(0)$, $X(\Delta)$, $X(2\Delta)$, ……, which consists of a sequence of zeros and ones. This process is a two state Markov chain with transition matrix, taken in the usual sense that p_{ij} is the probability of a transition from state i to state j,

*Korea Advanced Institute of Science and Technology, Department of Industrial Science

$$P=\begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix}$$

where

$$p_{01}=\frac{\lambda}{\lambda+\mu}\{1-\exp(-(\lambda+\mu)\Delta)\},\ p_{00}=1-p_{01}, \tag{1.1}$$

$$p_{10}=\frac{\mu}{\lambda+\mu}\{1-\exp(-(\lambda+\mu)\Delta)\},\ p_{11}=1-p_{10}. \tag{1.2}$$

Note that p_{01} and p_{10} should satisfy the relationship

$$0<p_{01}+p_{10}<1 \tag{1.3}$$

for all positive values of λ and μ.

If the sampling interval Δ is given, we see that there is a one to one correspodence between (λ, μ) and (p_{01}, p_{10}). The problem of estimating the parameters (λ, μ) is then equivalent to the problem of estimating the transition probabilities (p_{01}, p_{10}) in the Markov chain given above.

Brown et al. [2] have studied this problem and constructed the maximum likelihood estimators. Their simulation results show that as the product of the parameter and the sampling interval Δ, $\lambda\Delta$ (or $\mu\Delta$), becomes large, the MLE's give little information and are likely to fail to satisfy the relationship (1.3).

They have suggested an adaptive strategy to obtain estimators satisfying the relationship (1.3) [3]. They reduce the sampling interval to one half if there exist no estimators of λ and μ, and keep on reducing it to one half of the previous one until the estimators seem to be acceptable. Using their adaptive strategy, however, has some shortcomings: First, it requires too many observations, particularly when the true parameter is relatively large. Second, if the estimates seem to be acceptable, all samples obtained are combined to get final estimates of the parameters. However, the combined estimates require a lot of computations.

In this paper, we develop estimators different from the MLE's by modifying the sample transition counts through least square adjustment methods to satisfy the relationsphip (1.3). We also propose an adaptive strategy for obtaining these estimators to overcome the shortcomings of Brown et al.'s.

Ⅱ. Maximum Likelihood Estimation

Suppose that a sample of size n, $X(0)$, $X(\Delta)$, ……, $X((n-1)\Delta)$, is observed. Let p be the probability that the initial state of the process is 1 and $q=1-p$. Then we have the likelihood

$$L=p^{X(0)}q^{1-X(0)}p_{00}^{n_{00}}p_{01}^{n_{01}}p_{10}^{n_{10}}p_{11}^{n_{11}}, \tag{2.1}$$

where n_{ij} is the number of state transitions from state i to state j, $(i, j=0, 1)$. Therefore, the maximum likelihood estimators $(\hat{p}_{01}, \hat{p}_{10})$ of (p_{01}, p_{10}) are

$$\hat{p}_{01}=\frac{n_{01}}{n_{00}+n_{01}} \tag{2.2}$$

and

$$\hat{p}_{10}=\frac{n_{10}}{n_{10}+n_{11}} \tag{2.3}$$

Example 1. Suppose that we observe the following sequence with $\Delta=1$.

1 1 0 0 0 1 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0

0 1 1 0 0 0 0 0 0 0 0 1 0 0 1 1 0 1 0 1 1 1 0 0 0

Here $n=50$, $n_{00}=17$, $n_{01}=7$, $n_{10}=8$, $n_{11}=17$. Therefore, $\hat{p}_{01}=0.2917$ and $\hat{p}_{10}=0.32$. Hence,

$$\lambda=-\frac{\hat{p}_{01}}{\hat{p}_{01}+\hat{p}_{10}}\log(1-\hat{p}_{01}-\hat{p}_{10})=0.4511,$$

$$\mu=-\frac{\hat{p}_{10}}{\hat{p}_{01}+\hat{p}_{10}}\log(1-\hat{p}_{01}-\hat{p}_{10})=0.4949.$$

Example 2. Suppose now that the following sequence is observed with $\Delta=1$.

0 0 0 0 1 0 1 0 0 0 1 0 0 1 1 1 0 0 0 1 1 0 0 0 0

0 0 1 1 0 1 1 0 1 1 0 0 1 0 1 0 1 1 0 0 0 1 0 0 1

Here $n=50$, $n_{00}=17$, $n_{01}=13$, $n_{10}=12$, $n_{11}=7$. We cannot obtain the maximum likelihood estimators of λ and μ, for $\hat{p}_{01}+\hat{p}_{10}=0.4334+0.6316>1$.

Brown et al. [2] have studied this problem by Monte Carlo simulation, which exhibited a high frequency of $\hat{p}_{01}+\hat{p}_{10}>1$. Two conclusions were drawn from their studies;

a) If the ture $\lambda^*=\lambda\Delta$ (or $\mu^*=\mu\Delta$) were any value greater than 1, the expectation of the estimate λ^* (or μ^*) would be approximately 1, thus working backward, an apparent estimate near 1 gives little information of the true value.

b) Reasonable confidence limits for λ^* could be found only when λ^* became small, say roughly less than 0.4.

Ⅲ. Modified Estimation

As in Example 2, there are cases where no estimators exist. Even if they do, their values tend to be quite different from the true parameters. The reason why estimators for λ and μ do not exist may be that the samplng interval Δ is relatively long so that we cannot observe all of the state transitions. Therefore, we modify these transition counts considering the sample transition counts and those missed. Denote the modified transition counts of n_{ij} by m_{ij}. Then the modified likelihood is given by

$$L=p^{X(0)}q^{1-X(0)}p_{00}{}^{m_{00}}p_{01}{}^{m_{01}}p_{10}{}^{m_{10}}p_{11}{}^{m_{11}}. \tag{3.1}$$

This equation has the same form as (2.1). Therefore, the modified estimators of p_{01} and p_{10} are

$$p_{01}{}^{\circ}=\frac{m_{01}}{m_{00}+m_{01}} \tag{3.2}$$

$$p_{10}{}^{\circ}=\frac{m_{10}}{m_{10}+m_{11}} \tag{3.3}$$

The problem is how to obtain those modified transition counts. We used a method of least square adjustment [5] which is useful in many situations.

$$\underset{m_{ij}(i,\ j=0,\ 1)}{\text{Minimize}}\quad S=\sum_i\sum_j(m_{ij}-n_{ij})^2/n_{ij}. \tag{3.4}$$

The following constraints are imposed on these transition counts: First, since the sum of all original sample transition counts is $n-1$, the modified transition counts should satisfy the same relationship,

$$\sum_i\sum_j m_{ij}=n-1. \tag{3.5}$$

Second, as the sample size n becomes large enough, the number of jumps from state i is approximately

$$\sum_j n_{ij}\approx(n-1)\lim_{t\to\infty} Pr\{X(t)=i\}. \tag{3.6}$$

Hence, We choose the modified transition counts to satisfy the following relationship,

$$\sum_j m_{ij}=(n-1)\pi_i, \quad (i=0,\ 1) \tag{3.7}$$

where

$$\pi_0=\mu/(\lambda+\mu)$$

$$\pi_1=\lambda/(\lambda+\mu).$$

Third, the new estimators must satisfy the relationship (1.3),

$$0<p_{01}{}^{\circ}+p_{10}{}^{\circ}<1,$$

and hence, the modified transition counts must satisfy the constraint

$$m_{01}m_{10}<m_{00}m_{11}. \tag{3.8}$$

Fourth, since we assume that the parameters λ and μ have positive values, the modified transition counts are assumed to be positive.

Consequently, we choose m_{ij} $(i,\ j=0,\ 1)$ as follows.

$$\underset{m_{ij}}{\text{Minimize}}\ \sum_i\sum_j (m_{ij}-n_{ij})^2/n_{ij}$$

subject to

$$\sum_i\sum_j m_{ij}=n-1,$$

$$\sum_j m_{ij}=(n-1)\pi_i, \quad (i=0,\ 1),$$

$$m_{01}m_{10}+K=m_{00}m_{11},$$

$$m_{ij}>0, \quad (i,\ j=0,\ 1)$$

where K is a positive slack variable.

This is a well known quadratic programming problem with one redundant constraint, $\Sigma\pi_i=1$. Deleting the redundant constraint, the optimal solutions are given by

$$m_{01}=A^{-1}\left[\frac{n-1}{1+\rho}\left(\frac{1}{n_{00}}+\frac{\rho^2}{n_{10}}\right)-\frac{\rho(1+\rho)}{(n-1)}\left(\frac{1}{n_{10}}+\frac{1}{n_{11}}\right)K\right] \tag{3.9}$$

$$m_{10}=A^{-1}\left[\rho\frac{(n-1)}{1+\rho}\left(\frac{1}{n_{01}}+\frac{\rho^2}{n_{11}}\right)-\frac{(1+\rho)}{(n-1)}\left(\frac{1}{n_{00}}+\frac{1}{n_{01}}\right)K\right] \tag{3.10}$$

where

$$A=1/n_{00}+1/n_{01}+\rho^2/n_{10}+\rho^2/n_{11}.$$

$$\rho=\lambda/\mu.$$

The optimal solutions are functions of ρ and K whose values must be determined explicitly. The values of ρ can be estimated through the method of maximum likelihood. The MLE of ρ is

$$\hat{\rho}=n_{01}(n_{10}+n_{11})/n_{10}(n_{00}+n_{01}). \tag{3.11}$$

The value of K is important, but is not easy to determine. Our simulation study, however, shows that the value of K varies with the estimates of the transition probabilities and $\hat{\rho}$. If the value of $\hat{\rho}$ is near 1, the following eguations are recommended

$$K=\max\{f(s,\ \hat{\rho},\ n),\ \varepsilon\}$$

where

$$f(s, \hat{p}, n) = (n-1)^2 (a+bs)\hat{p}/(1+\hat{p})^2,$$

$$(a, b) = \begin{cases} (0.990453, \; -0.8795) & \text{if } n=50, \\ (0.988790, \; -0.9250) & \text{if } n=100, \end{cases}$$

$$s = \hat{p}_{01} + \hat{p}_{10},$$

ε; arbitary small value smaller than 1 (10^{-5} in our simulation study).

The parameters λ and μ can now be uniquely determined. However, our simulation results show that the modified estimators are biased although they always exist. (See Fig. 1 and Fig. 2).

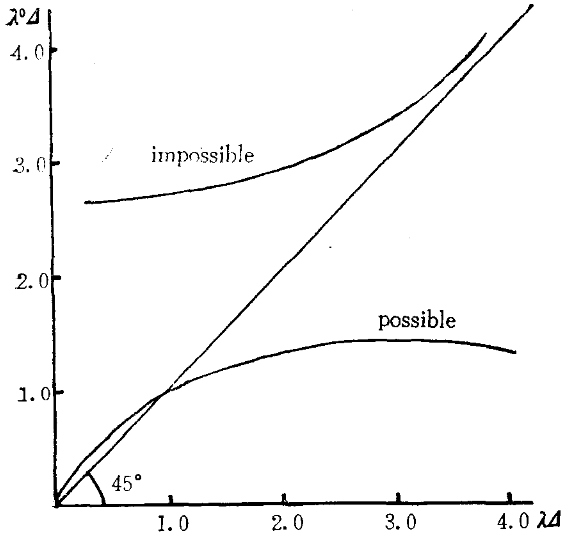

Fig. 1. Modified estimator with $n=50$

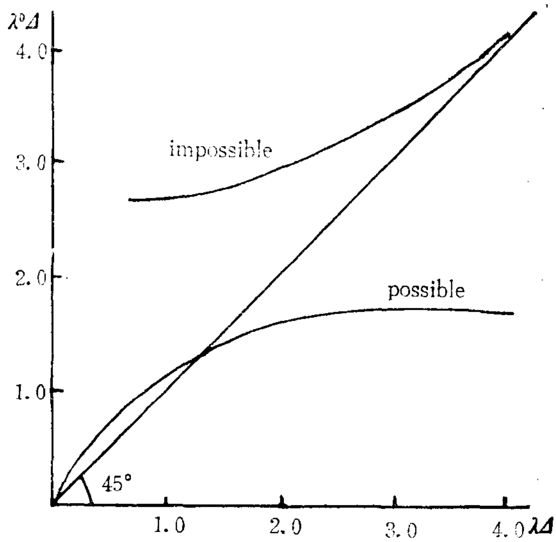

Fig. 2. Modified estimator with $n=100$

Fig. 1 and Fig. 2 show the average values of the modified estimators. Here a "possible" case represents the case where we can obtain the MLE and an "impossible" case where we cannot. Comparing these values with the M.L.E's, however, we see that the modified estimators are better than M.L.E's. If it is not possible to obtain more samples or to reduce the sampling interval Δ, the value of the modified estimator is the only information we can take from the samples, for the M.L.E's of λ and μ often do not exist.

To reduce the biases, an adaptive strategy may be used. The adaptive technique developed by Brown et al. [3], however, requires many samples and a lot of computations. Hence, we develop a new adaptive strategy to overcome such shortcomings using the modified estimators.

Ⅳ. Adaptive Strategy

In interval sampling the value of the interval length is crucial and must not be too large. If it is larger than $1/\mu$ or $1/\lambda$ which are average sojourn times in state 0 and 1, respectively, we are likely to fail to observe all state transitions. Hence, the value of the interval length should be determined to obtain accurate estimates efficiently.

Our simulation result which agrees with Brown et al.'s [2] shows that the maximum likelihood estimate of $\lambda\Delta$ (or $\mu\Delta$) is good when it is smaller than roughly 0.4. Thus, in our adaptive strategy the modified estimators are used to obtain a suitable value of the interval length in stage 1 and the parameters are estimated through the method of maximum likelihood in stage 2.

Adaptive Strategy

Stage 1. Write $\lambda_1^*=\lambda\Delta_1$ and $\mu_1^*=\mu\Delta_1$. Take a sample of size n from the process. Estimate the parameters λ_1^* and μ_1^* by the method of maximum likelihood. If these estimates are both less than 0.4, accept them. If they are too large or do not exist, obtain the modified estimates of $\lambda_1^{*\circ}$ and $\mu_1^{*\circ}$ by the method of Section Ⅲ. Determine Δ_2 so that $\lambda_1^{*\circ}\Delta_2<0.4$ and $\lambda_1^{*\circ}\Delta_2<0.4$.

Stage 2. Take a second sample from the process with interval length Δ_2. Use only stage 2 sample to estimate $\lambda_2^*=\lambda\Delta_2$ and $\mu_2^*=\mu\Delta_2$ by the method of maximum likelihood. If the MLE's exist, accept them as final estimates. If they do not, obtain the estimates of $\lambda_2^{*\circ}$ and $\mu_2^{*\circ}$ by the method of Section Ⅲ and accept them as the final estimates.

Ⅴ. Monte Carlo Study

The behaviors of the MLE and the proposed adaptive strategy are investigated using Monte Carlo simulation. Without loss of generality we take the sampling interval Δ to be 1. The sample size considered are 50 and 100.

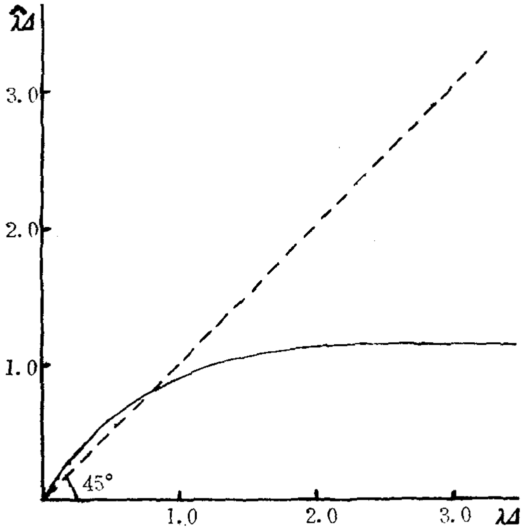

Fig. 3. MLE with $n=50$

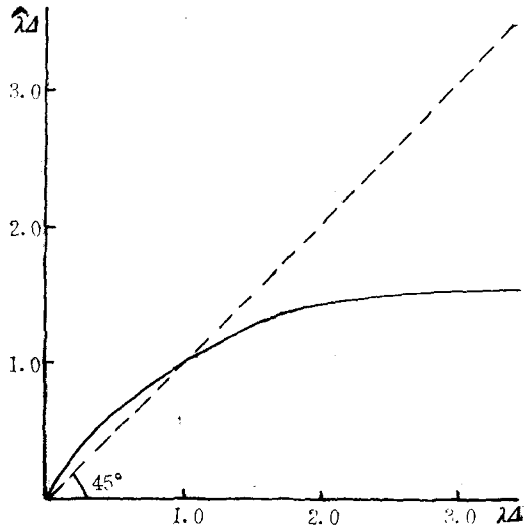

Fig. 4. MLE with $n=100$

Fig. 3 and Fig. 4. show the behaviors of the MLE's. The bold line represents the average values of the estimates.

Table 1. Number of impossible cases out of 500 experiments for $\rho=1$

n \ λ	0.2	0.5	0.8	1.0	1.2	1.5	1.8	2.0	2.2	2.5	2.8	3.0	3.2	3.5	3.8
50	0	1	46	106	140	203	236	263	274	272	289	269	289	266	273
100	0	0	6	49	107	162	218	228	261	249	262	265	277	280	276

These figures show that the MLE's give little information for large values of the true parameter.

Table 1 shows the number of cases where MLE's do not exist out of 500 experiments, which increases as the true parameter does.

Fig. 5 and Fig. 6 show the behavior of the proposed adaptive strategy. We take the initial sampling interval Δ_1 to be 1. The bold line represents the average value of the estimates as before.

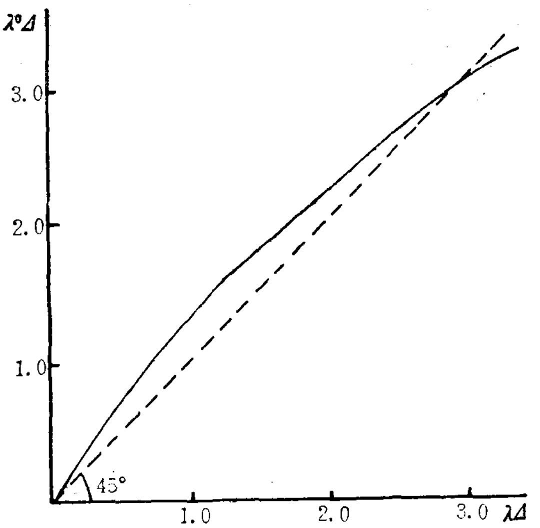

Fig. 5. Estimator using adaptive strategy with $n=50$

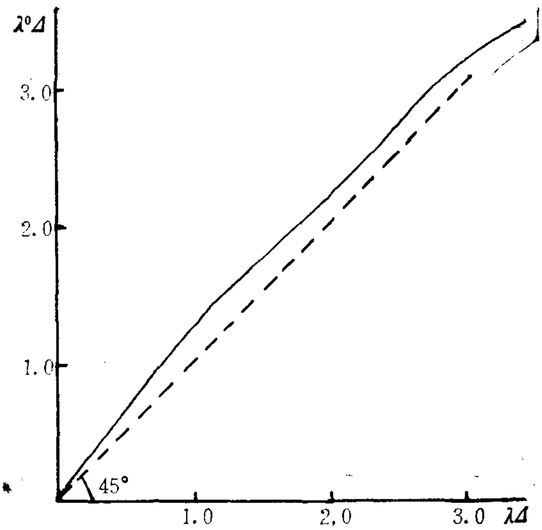

Fig. 6. Estimator using adaptive strategy with $n=100$

From the figures, we can see that the estimates are better than MLE's in average values although only two stage samples are used; the expectation line (bold line) approaches to the line which goes through the origin with slope 1.

VI. Discussion

In estimating the parameters of a two state Markov process by interval sampling, frequently MLE's cannot be obtained. To overcome such shortcomings of the classical MLE's, sample transition counts are modified to obtain new estimators that always exist. An adaptive strategy using these estimators is also proposed. These are effective when the true parameters seem to have relatively large values.

The modified estimators, however, depend largely upon the value of K, which in turn should be determined empirically for different values of ρ. The method will then also provide better estimates than the MLE's when the values of the true pameters are small.

References

1. Billingsley, P., *Statistical Inference for Markov Processes*, The University of Chicago Press, 1961.
2. Brown, M., Solomon, H. and Stephens, M.A., "Estimation of Parameters of Zero-One Process by Interval Sampling," *Operations Research*, Vol. 25, No. 3, 493—505, 1977.
3. Brown, M., Solomon, H. and Stephens, M.A., "Estimation of Parameters of Zero-One Process by Interval Sampling; An Adaptive Strategy," *Operations Research*, Vol. 27, No. 3, 606—615, 1979.
4. Cox, D.R., *Renewal Theory*, Methuen, London, 1962.
5. Deming, W.E. and Stephan, F.F., "On a Least Square Adjustment of a Sampled Frequency When the Expected Totals Are Known," *Annals of Mathematical Statistics*, Vol. 11, No. 3, 427—444, 1940.