Journal of the Operations Research
Society of Japan
Vol. 25, No. 2, June 1982

MAXIMUM LIKELIHOOD ESTIMATION OF PARAMETERS IN BIRTH AND DEATH PROCESSES BY POISSON SAMPLING

Yutaka Baba
Tokyo Institute of Technology

(Received May 24, 1980; Final October 5, 1981)

Abstract Maximum likelihood estimates of parameters in continuous time Markov chains are obtained when the observation plan is Poisson sampling. Furthermore, for birth and death processes, variance-covariance matrix of parameters is obtained. Especially, for queueing model M/M/1, the variance-covariance matrix by Poisson sampling is compared with the variance-covariance matrix by complete observation in detail.

1. Introduction

Let $\{x(t),\ t \geqq 0\}$ be a continuous time Markov chain with transition intensity matrix $Q = (q_{ij}(\theta))$, where θ is a m-dimensional column vector which represents an unknown parameter ranging over a set Θ in m-dimensional Euclidean space. The problem which we shall consider is that of estimating from observations on the process.

If $x(t)$ is observed continuously over the time $[0,T]$, we say that this observation plan is complete observation.

In this paper, we shall consider a different observation plan named Poisson sampling. We shall observe the process at a random observation points $\{v_0 = 0^+ < v_1 < v_2 < \ldots\}$, and count the number of transitions of states occuring in the random observation intervals $(v_{k-1}, v_k]$ $(k = 1,2,\ldots)$. Now, assume that $\zeta_k = v_k - v_{k-1}$ $(k = 1,2,\ldots)$ are independent identically distributed exponential variates with parameter ν. This sampling is then called Poisson sampling derived from $\{x(t),\ t \geqq 0\}$. This method was first considered by Kingman [4]. But an application to a statistical problem using Poisson sampling is only seen in Basawa [1].

Here we shall consider the estimation of the parameters of a continuous time Markov chain based on Poisson sampling. We often encounter the case that the between-event intervals are unobservable. Such situations are not uncommon in practice, and occurs particularly in certain queueing and other stochastic models. In such a case, Poisson sampling is a powerful method, because Kingman [4] showed that it determines the original process completely.

In Section 2, we shall construct the likelihood function of a continuous time Markov chain of general type when the observation is done by Poisson sampling. And solving the maximum likelihood equations, we shall obtain the maximum likelihood estimates of parameters. Furthermore we want to get the variance-covariance matrix of parameters and compare two matrices in the complete observation case and Poisson sampling case. However, it is very difficult to obtain the variance-covariance matrix in a general type Markov chain. So, in Section 3, we shall only consider the birth and death process which is a very useful but simple structured model. For the birth and death process, we have the variance-covariance matrices of parameters for two simple models given by Wolff [6]. But even for the birth and death process, the variance-covariance matrix by Poisson sampling is not yet represented by explicit forms. Hence, in Section 4, we shall deal with in detail a simple but useful queueing model $M/M/1$ which is a special case of birth and death processes. For this model, we can compare the variance-covariance matrices of complete observation case over $[0,T]$ and Poisson sampling case of w observation points. And the determination of w is done in order to obtain smaller variance of parameters than the variance of complete observation case.

2. Maximum Likelihood Estimation of Parameters in Continuous Time Markov Chains by Poisson Sampling

We observe a continuous time Markov chain with transition intensity matrix $Q = (q_{ij}(\theta))$ by Poisson sampling until the number of transitions of states amounts to n. So the data obtained by this sampling are successive states $(x_0, x_1, \ldots, x_n)$ and l_i $(i = 1, \ldots, n)$ which are the number of observation points between x_{i-1} and x_i.

Let $q_i(\theta) = \sum_{j \neq i} q_{ij}(\theta)$ and $\pi_{ij}(\theta) = q_{ij}(\theta)/q_i(\theta)$. Then, from the properties of a continuous time Markov chain, the following equations hold clearly:

$$\text{(2.1)} \qquad \Pr(l_{i+1} = k \mid x_i = j) = \int_0^\infty \frac{exp(-\nu t)(\nu t)^k}{k!} q_j(\theta) exp(-q_j(\theta)t)dt$$

$$= \frac{q_j(\theta)\nu^k}{\{\nu + q_j(\theta)\}^{k+1}} \qquad (k = 0,1,\ldots)$$

(2.2) $$E(l_{i+1} \mid x_i = j) = \frac{\nu}{q_j(\theta)} .$$

Thus, ignoring the distribution of initial state x_0, the likelihood function based on these data is

(2.3) $$L_n(\theta) = \prod_{i=0}^{n-1} \pi_{x_i x_{i+1}}(\theta) \frac{q_{x_i}(\theta)\nu^{l_{i+1}}}{\{\nu + q_{x_i}(\theta)\}^{l_{i+1}+1}}$$

$$= \prod_{i=0}^{n-1} \frac{q_{x_i x_{i+1}}(\theta)\nu^{l_{i+1}}}{\{\nu + q_{x_i}(\theta)\}^{l_{i+1}+1}} .$$

Therefore, from the following maximum likelihood equations,

(2.4) $$\frac{\partial lnL_n(\theta)}{\partial \theta_i} = 0 \qquad (i = 1,\ldots,m)$$

we can get the maximum likelihood estimates (MLE) of θ.

Especially, if the state space is finite, that is, $S = \{1,\ldots,s\}$ and $q_{ij}(\theta) = q_{ij}$, then the log-likelihood function with Lagrange multiplier be-becomes

(2.5)
$$\begin{aligned} lnL_n &= \sum_{k=0}^{n-1} lnq_{x_k x_{k+1}} + \sum_{k=0}^{n-1} l_{k+1} ln\nu \\ &\quad - \sum_{k=0}^{n-1} (l_{k+1} + 1) ln(\nu + q_{x_k}) + \sum_{i=1}^{s} \lambda_i (q_i - \sum_{j\neq i} q_{ij}) \\ &= \sum_{j\neq i} n_{ij} lnq_{ij} + \sum_{k=0}^{n-1} l_{k+1} ln\nu \\ &\quad - \sum_{i=1}^{s} r_i ln(\nu + q_i) + \sum_{i=1}^{s} \lambda_i (q_i - \sum_{j\neq i} q_{ij}) \end{aligned}$$

where $r_i = \sum_{k=0}^{n-1} \delta_{x_k,i}(l_{k+1} + 1)$, ($\delta_{x_k,i}$ is the Kronecker delta) and n_{ij} is the number of transitions such that $x_k = i$ and $x_{k+1} = j$.

Hence the maximum likelihood equations are

$$\frac{\partial lnL_n}{\partial q_{ij}} = \frac{n_{ij}}{q_{ij}} - \lambda_i = 0 \qquad (i = 1,\ldots,s;\ j = 1,\ldots,s;\ j \neq i)$$

$$(2.6)\qquad \frac{\partial \ln L_n}{\partial q_i} = -\frac{r_i}{\nu + q_i} + \lambda_i = 0 \qquad (i = 1,\dots,s)$$

$$\frac{\partial \ln L_n}{\partial \lambda_i} = q_i - \sum_{j\neq i} q_{ij} = 0 \qquad (i = 1,\dots,s).$$

Solving equations (2.6), we obtain

$$(2.7)\qquad \hat{q}_{ij} = \frac{\nu n_{ij}}{r_i - \sum_{j\neq i} n_{ij}} \qquad (i = 1,\dots,s;\ j = 1,\dots,s;\ j \neq i).$$

This MLE has a simple form. But the denominator $r_i - \sum_{j\neq i} n_{ij}$ of (2.7) may be zero. For such i, we cannot estimate q_{ij}.

3. Determination of Variance-Covariance Matrices for Birth and Death Processes by Poisson Sampling

In Section 2, we have obtained the MLE of $q_{ij}(\theta)$. It is, however, very difficult for a Markov chain in general type to obtain the variance-covariance matrix of $q_{ij}(\theta)$. So, in this Section, we shall consider only for ergodic birth and death processes.

We denote by $\lambda_i(\theta)$ and $\mu_i(\theta)$ $(i = 0,1,\dots)$ the parameters of birth and death processes. In our notaions defined in Section 2, $\lambda_i(\theta) = q_{i,i+1}(\theta)$ and $\mu_i(\theta) = q_{i,i-1}(\theta)$.

In constructing the likelihood function, it will be useful to consider the portion of it, L_i, which represents a single transition $x_i \to x_{i+1}$. Given

$$(3.1)\qquad x_i = j \ \text{ and } \ l_{i+1} = k,$$

it is easy to show the following expression

$$(3.2)\qquad \begin{aligned} \ln L_i(\theta) &= \ln\lambda_j(\theta) + k\ln\nu - (k+1)\ln\{\nu + \lambda_j(\theta) + \mu_j(\theta)\} \\ &\qquad\qquad \text{if } x_{i+1} = j+1 \\ &= \ln\mu_j(\theta) + k\ln\nu - (k+1)\ln\{\nu + \lambda_j(\theta) + \mu_j(\theta)\} \\ &\qquad\qquad \text{if } x_{i+1} = j-1. \end{aligned}$$

Define

$$(3.3)\qquad \begin{aligned} G_u &= (\partial/\partial\theta_u)\ln L_i(\theta) && (u = 1,\dots,m) \\ G &= (G_1,\dots,G_m)^T \\ G_{uv} &= (\partial^2/\partial\theta_u\partial\theta_v)\ln L_i(\theta) && (u = 1,\dots,m;\ v = 1,\dots,m). \end{aligned}$$

Then, from Theorem 7.3 of Billingsley [2], $\sqrt{n}(\hat{\theta} - \theta)$ is asymptotically normal (it converges in law to the normally distributed random variable) with mean 0 and variance-covariance matrix $\sigma^{-1}(\theta)$. Besides under regularity conditions the following facts are well known

$$(3.4)\qquad E(G_u) = 0, \quad E(G_{uv}) = -\sigma_{uv}.$$

From (2.4), the variance-covariance matrix of G_u, $E(GG^T)$, is

(3.5) $$E(GG^T) = \sigma(\theta) = (\sigma_{uv}) = (-E(G_{uv})).$$

We shall calculate the asymptotic variance-covariance matrix for two simple models given by Wolff [6].

Model 1:

$\lambda_j > 0, \quad j = 0,1,\dots,M-1, \quad \lambda_M = 0$

$\mu_j > 0, \quad j = 1,2,\dots,M, \quad \mu_0 = 0.$

There are $2M$ unknown parameters $\theta^T = (\lambda_0,\dots,\lambda_{M-1},\mu_1,\dots,\mu_M)$, which are required only to be finite and positive. λ_j is the $j+1$ st component and μ_j is the $M+j$ th component of θ. From equation (3.2), under the condition (3.1), we have

(3.6)
$$G_{j+1} = \frac{1}{\lambda_j} - \frac{k+1}{\nu+\lambda_j+\mu_j}, \quad G_{M+j} = -\frac{k+1}{\nu+\lambda_j+\mu_j} \quad \text{for } x_{i+1} = j+1$$
$$G_{j+1} = -\frac{k+1}{\nu+\lambda_j+\mu_j}, \quad G_{M+j} = \frac{1}{\mu_j} - \frac{k+1}{\nu+\lambda_j+\mu_j} \quad \text{for } x_{i+1} = j-1$$

otherwise $G_u = 0$.

We continue to condition on (3.1)

(3.7)
$$G_{j+1,j+1} = -\frac{1}{\lambda_j^2} + \frac{k+1}{(\nu+\lambda_j+\mu_j)^2},$$
$$G_{M+j,M+j} = \frac{k+1}{(\nu+\lambda_j+\mu_j)^2}, \quad G_{j+1,M+j} = \frac{k+1}{(\nu+\lambda_j+\mu_j)^2} \quad \text{for } x_{i+1} = j+1$$
$$G_{j+1,j+1} = \frac{k+1}{(\nu+\lambda_j+\mu_j)^2}, \quad G_{j+1,M+j} = \frac{k+1}{(\nu+\lambda_j+\mu_j)^2}$$
$$G_{M+j,M+j} = -\frac{1}{\mu_j^2} + \frac{k+1}{(\nu+\lambda_j+\mu_j)^2} \quad \text{for } x_{i+1} = j-1$$

otherwise $G_{uv} = 0$.

Thus we have

(3.8)
$$-E(G_{j+1,j+1} \mid x_i = j,\ l_{i+1} = k)$$
$$= \frac{\lambda_j}{\lambda_j+\mu_j}\left\{\frac{1}{\lambda_j^2} - \frac{k+1}{(\nu+\lambda_j+\mu_j)^2}\right\} - \frac{\mu_j}{\lambda_j+\mu_j}\cdot\frac{k+1}{(\nu+\lambda_j+\mu_j)^2}$$
$$= \frac{1}{\lambda_j(\lambda_j+\mu_j)} - \frac{k+1}{(\nu+\lambda_j+\mu_j)^2}.$$

And similarly

(3.9)
$$-E(G_{M+j,M+j} \mid x_i = j,\ l_{i+1} = k) = \frac{1}{\mu_j(\lambda_j+\mu_j)} - \frac{k+1}{(\nu+\lambda_j+\mu_j)^2}$$

$$-E(G_{j+1,M+j} \mid x_i = j,\ l_{i+1} = k) = \frac{k+1}{(\nu + \lambda_j + \mu_j)^2}\,.$$

From equations (3.8), (3.9), and (2.2), we have

$$-E(G_{j+1,j+1} \mid x_i = j) = \frac{1}{\lambda_j(\lambda_j + \mu_j)} - \frac{1}{(\lambda_j + \mu_j)(\nu + \lambda_j + \mu_j)}$$

$$= \frac{\nu + \mu_j}{\lambda_j(\lambda_j + \mu_j)(\nu + \lambda_j + \mu_j)}$$

(3.10)

$$-E(G_{M+j,M+j} \mid x_i = j) = \frac{\nu + \lambda_j}{\mu_j(\lambda_j + \mu_j)(\nu + \lambda_j + \mu_j)}$$

$$-E(G_{j+1,M+j} \mid x_i = j) = \frac{1}{(\lambda_j + \mu_j)(\nu + \lambda_j + \mu_j)}$$

Here we define some notations. Let P_j $(j = 0,1,\ldots,M)$ be the stationary distribution and y_j $(j = 0,1,\ldots,M)$ the probability that a randomly selected transition was out of state j. Then, using the fact by Wolff [6], it is easily shown that

(3.11) $$y_j = (\lambda_j + \mu_j)P_j/2R \qquad (j = 0,1,\ldots,M)$$

where $R = \sum_{j=0}^{M-1} \lambda_j P_j = \sum_{j=0}^{M} \mu_j P_j$ (R is called the average throughput).

Hence, from (3.10), we have

$$V(G_{j+1}) = E[-E(G_{j+1,j+1} \mid x_i = k)] = -\sum_{k=0}^{M} E(G_{j+1,j+1} \mid x_i = k)y_k$$

$$= \frac{y_j(\nu + \mu_j)}{\lambda_j(\lambda_j + \mu_j)(\nu + \lambda_j + \mu_j)}$$

(3.12) $$V(G_{M+j}) = \frac{y_j(\nu + \lambda_j)}{\mu_j(\lambda_j + \mu_j)(\nu + \lambda_j + \mu_j)}$$

$$\mathrm{Cov}(G_{j+1,M+j}) = -\frac{y_j\nu}{(\lambda_j + \mu_j)(\nu + \lambda_j + \mu_j)}$$

otherwise $\mathrm{Cov}(G_u, G_v) = 0$.

Hence, using equations (3.11) and (3.12), we have

$$V(G_{j+1}) = \frac{P_j(\nu + \mu_j)}{2R\lambda_j(\nu + \lambda_j + \mu_j)} \qquad (j = 0,1,\ldots,M-1)$$

(3.13) $$V(G_{M+j}) = \frac{P_j(\nu + \lambda_j)}{2R\mu_j(\nu + \lambda_j + \mu_j)} \qquad (j = 1,2,\ldots,M)$$

$$\mathrm{Cov}(G_{j+1,M+j}) = -\frac{P_j\nu}{2R(\nu + \lambda_j + \mu_j)} \qquad (j = 1,2,\ldots,M-1)$$

otherwise $\mathrm{Cov}(G_u, G_v) = 0$.

Model 2:

$$\lambda_j = \lambda h(j) , \quad \mu_j = \mu g(j) , \quad j = 0,1,\dots, \text{ and } g(0) = 0$$

where $\theta^T = (\lambda, \mu)$, and $h(j)$, $g(j)$ are assumed to be arbitrary but known function of j. Moreover, for ergodicity, we assume that $\sum_{n=1}^{\infty} (\lambda/\mu)^n \sum_{k=0}^{n-1} \frac{h(k)}{g(k+1)} < \infty$. Given (3.1), we have from (3.2)

$$G_1 = \frac{1}{\lambda} - \frac{(k+1)h(j)}{\nu + \lambda h(j) + \mu g(j)} , \quad G_2 = - \frac{(k+1)g(j)}{\nu + \lambda h(j) + \mu g(j)} \quad \text{if } x_{i+1} = j + 1$$

$$G_1 = - \frac{(k+1)h(j)}{\nu + \lambda h(j) + \mu g(j)} , \quad G_2 = \frac{1}{\mu} - \frac{(k+1)g(j)}{\nu + \lambda h(j) + \mu g(j)} \quad \text{if } x_{i+1} = j - 1$$

(3.14)
$$\begin{aligned} G_{11} &= - \frac{1}{\lambda^2} + \frac{(k+1)\{h(j)\}^2}{\{\nu + \lambda h(j) + \mu g(j)\}^2} \\ G_{22} &= \frac{(k+1)\{g(j)\}^2}{\{\nu + \lambda h(j) + \mu g(j)\}^2} \\ G_{12} &= \frac{(k+1)h(j)g(j)}{\{\nu + \lambda h(j) + \mu g(j)\}^2} \qquad \text{if } x_{i+1} = j + 1 \\ G_{11} &= \frac{(k+1)\{h(j)\}^2}{\{\nu + \lambda h(j) + \mu g(j)\}^2} \\ G_{22} &= - \frac{1}{\mu^2} + \frac{(k+1)\{g(j)\}^2}{\{\nu + \lambda h(j) + \mu g(j)\}^2} \\ G_{12} &= \frac{(k+1)h(j)g(j)}{\{\nu + \lambda h(j) + \mu g(j)\}^2} \qquad \text{if } x_{i+1} = j - 1 . \end{aligned}$$

By similar calculations as Model 1, we have

(3.15)
$$\begin{aligned} V(G_1) &= \frac{1}{2\lambda^2} - \sum_{j=0}^{\infty} \frac{\{h(j)\}^2 P_j}{2R\{\nu + \lambda h(j) + \mu g(j)\}} \\ V(G_2) &= \frac{1}{2\mu^2} - \sum_{j=0}^{\infty} \frac{\{g(j)\}^2 P_j}{2R\{\nu + \lambda h(j) + \mu g(j)\}} \\ \mathrm{Cov}(G_1, G_2) &= - \sum_{j=0}^{\infty} \frac{h(j)g(j)P_j}{2R\{\nu + \lambda h(j) + \mu g(j)\}} . \end{aligned}$$

In both Model 1 and Model 2, we have obtained the asymptotic variance-covariance matrices of parameters. But, in general, it is difficult to get the explicit forms of P_j and R. So, $\sigma(\theta)$ or its inverse $\sigma^{-1}(\theta)$ is not represented here in explicit form. In next Section, we shall deal with in detail a simple but useful model *M/M/1* which is a special case of Model 2.

4. Further Investigation of Queueing Model M/M/1

Let us define

$$(4.1)\qquad \begin{aligned} \lambda_i &= \lambda && \text{for } i = 0,1,\ldots \\ \mu_i &= \mu && \text{for } i = 1,2,\ldots \\ &= 0 && \text{for } i = 0 \end{aligned}$$

and for ergodicity assume that $\lambda < \mu$. So that, in our notation in Section 3,

$$(4.2)\qquad \begin{aligned} h(i) &= 1 && \text{for } i = 0,1,\ldots \\ g(i) &= 1 && \text{for } i = 1,2,\ldots \\ &= 0 && \text{for } i = 0 . \end{aligned}$$

On this model, we shall consider the variance-covariance matrix of parameters both when the observation is complete over $[0,T]$, and when the observation is made by Poisson sampling of w observation points.

In Section 3, we have fixed the number of transitions of states. But in this Section, we fix the number of observation points. So the number of transitions of states, n, becomes a random variable and depends on w. We denote it as $n(w)$.

We shall also consider the determination of w so that we get smaller variance than that in the case of complete observation over the interval $[0,T]$.

When the data were obtained by complete observation over $[0,T]$, we define the following notations.

$$(4.3)\qquad \begin{aligned} \gamma_i &= \textit{total soujourn time in state } i \textit{ over } [0,T] \\ a_i &= \textit{the number of transitions } i \to i+1 \textit{ over } [0,T] \\ b_i &= \textit{the number of transitions } i \to i-1 \textit{ over } [0,T] \end{aligned}$$

$$a = \sum_{i=0}^{\infty} a_i , \quad b = \sum_{i=0}^{\infty} b_i .$$

Then, by using the notations (4.3), it is easily derived that the maximum likelihood estimates of λ and μ are

$$(4.4)\qquad \hat{\lambda} = \frac{a}{T} , \quad \hat{\mu} = \frac{b}{T - \gamma_0} .$$

Furthermore, from Billingsley [2] and Reynolds [5], it is easily shown that, for large T,

$$(4.5) \qquad \hat{\lambda} \sim N(\lambda, \frac{\lambda}{T}), \qquad \hat{\mu} \sim N(\mu, \frac{\mu^2}{\lambda T})$$

Next we consider the Poisson sampling case. We construct the likelihood function by Poisson sampling when the observation intervals are independently and identically distributed exponential variates with parameter ν, and the number of observation points is w. We define some notations similarly to (4.3).

(4.6)
u_i = *the number of transitions* $i \to i+1$ *until* w*-th Poisson observation point*

d_i = *the number of transitions* $i \to i-1$ *until* w*-th Poisson observation point*

$$u = \sum_{i=0}^{\infty} u_i, \qquad d = \sum_{i=0}^{\infty} d_i.$$

Then, similarly to (2.5) in Section 2, let

$$(4.7) \qquad r_i = \sum_{k=0}^{n(w)-1} \delta_{x_k,i}(l_{k+1}+1) \quad \text{and} \quad r = \sum_{i=0}^{\infty} r_i.$$

Hence, analogously to (2.3), ignoring the initial distribution and end effect, that is, the information contained in the process from the last transition point to the last observation point, it is easily shown that the likelihood function based on these data is

$$(4.8) \qquad L = \frac{\lambda^u \mu^d}{(\nu+\lambda)^{r_0}(\nu+\lambda+\mu)^{r-r_0}} \cdot K$$

where K is a constant free from λ and μ. Taking the natural logarithm of both sides, we have

$$(4.9) \qquad lnL = u ln\lambda + d ln\mu + lnK - r_0 ln(\nu+\lambda) - (r-r_0) ln(\nu+\lambda+\mu) \ .$$

Hence, from the following maximum likelihood equations,

$$(4.10) \qquad \begin{aligned} \frac{\partial lnL}{\partial \lambda} &= \frac{u}{\lambda} - \frac{r_0}{\nu+\lambda} - \frac{r-r_0}{\nu+\lambda+\mu} = 0 \\ \frac{\partial lnL}{\partial \mu} &= \frac{d}{\mu} - \frac{r-r_0}{\nu+\lambda+\mu} = 0 \end{aligned}$$

we obtain the maximum likelihood estimates

(4.11) $$\hat{\lambda} = \frac{u\nu}{r - u - d} \ , \qquad \hat{\mu} = \frac{d(r-d)\nu}{(r - r_0 - d)(r - u - d)} \ .$$

Besides, from equations (4.10), we have

$$\frac{\partial^2 lnL}{\partial\lambda^2} = -\frac{u}{\lambda^2} + \frac{r_0}{(\nu+\lambda)} + \frac{r - r_0}{(\nu+\lambda+\mu)^2}$$

(4.12) $$\frac{\partial^2 lnL}{\partial\mu^2} = -\frac{d}{\mu^2} + \frac{r - r_0}{(\nu+\lambda+\mu)^2}$$

$$\frac{\partial^2 lnL}{\partial\lambda\partial\mu} = \frac{r - r_0}{(\nu+\lambda+\mu)^2} \ .$$

When the number of observation points is w, actual w-th observation time, T_w, obeys w-th Erlang distribution with parameter ν. Therefore the probability density function of T_w, $f(t)$, is

(4.13) $$f(t) = \frac{\nu^w t^{w-1}}{\Gamma(w)} exp(-\nu t) \ .$$

Since it is well known that both u and d obey Poisson distribution with parameter λT_w, $E(u \mid T_w) = E(d \mid T_w) = \lambda T_w$. Hence, we have

(4.14) $$E(u) = E(d) = \int_0^\infty \lambda t f(t)dt = \frac{\lambda\nu^w}{\Gamma(w)}\int_0^\infty t^w exp(-\nu t)dt = \frac{\lambda w}{\nu} \ .$$

Next let z be the number of observation points from the last transition point to the last observation point. Then it is easily shown that

(4.15) $$r = u + d + w - z \ .$$

Taking expectations of both sides, we have

(4.16) $$E(r) = E(u) + E(d) + w - E(z) = \frac{2\lambda w}{\nu} + w - E(z) \ .$$

Since, for large w, it is easily verified that $E(z)$ is negligible, we have

(4.17) $$E(r) \approx \frac{2\lambda w}{\nu} + w \ .$$

For $M/M/1$, stationary distribution P_j and average throughput R defined in Section 3 have simple forms and are represented by

(4.18) $$P_j = (1 - \frac{\lambda}{\mu})(\frac{\lambda}{\mu})^j \qquad (j = 0,1,\ldots) \qquad R = \lambda.$$

And from (3.11), we have

(4.19) $$y_0 = \frac{1}{2}(1 - \frac{\lambda}{\mu}) \ .$$

Hence we have

(4.20) $$E(u_0) = E(u + d)y_0 = \frac{2\lambda w}{\nu}\cdot\frac{1}{2}(1 - \frac{\lambda}{\mu}) = \frac{w\lambda(\mu - \lambda)}{\mu\nu} .$$

Furthermore by taking the expectation of both sides in (4.7), we have

(4.21) $$E(r_0) = E(u_0)\cdot\frac{\nu + \lambda}{\lambda} = \frac{w(\mu - \lambda)(\nu + \lambda)}{\mu\nu}$$

where $\frac{\nu + \lambda}{\lambda}$ is the expectation of the number of observation points + 1 when a transition is from 0 to 1. From equations (4.12), (4.14), (4.17), (4.20) and (4.21) we obtain

$$-E\left(\frac{\partial^2 lnL}{\partial\lambda^2}\right) \approx \frac{w(\lambda^2 + \nu^2 + \lambda\nu + \mu\nu)}{\lambda\nu(\nu + \lambda)(\nu + \lambda + \mu)}$$

(4.22) $$-E\left(\frac{\partial^2 lnL}{\partial\mu^2}\right) \approx \frac{w\lambda(\nu + \lambda)}{\mu^2\nu(\nu + \lambda + \mu)}$$

$$-E\left(\frac{\partial^2 lnL}{\partial\lambda\partial\mu}\right) \approx -\frac{w\lambda}{\mu\nu(\nu + \lambda + \mu)} .$$

Therefore asymptotic variance-covariance matrix of λ and μ are

(4.23) $$\begin{pmatrix} -E\left(\frac{\partial^2 lnL}{\partial\lambda^2}\right) & -E\left(\frac{\partial^2 lnL}{\partial\lambda\partial\mu}\right) \\ -E\left(\frac{\partial^2 lnL}{\partial\lambda\partial\mu}\right) & -E\left(\frac{\partial^2 lnL}{\partial\mu^2}\right) \end{pmatrix}^{-1} \approx \begin{pmatrix} \frac{\lambda(\nu + \lambda)}{w} & \frac{\lambda}{w} \\ \frac{\lambda}{w} & \frac{\mu^2(\lambda^2 + \nu^2 + \lambda\nu + \mu\nu)}{w\lambda(\nu + \lambda)} \end{pmatrix}$$

Comparing with the complete observation case, in order to obtain smaller variances of λ and μ than those of complete observation over $[0,T]$, the number of observation points by Poisson sampling, w, will be needed as follows by using (4.5) and (4.23)

(4.24) $$\frac{\lambda(\nu + \lambda)}{w} \leqq \frac{\lambda}{T}, \quad \frac{\mu^2(\lambda^2 + \nu^2 + \lambda\nu + \mu\nu)}{w\lambda(\nu + \lambda)} \leqq \frac{\mu^2}{\lambda T} .$$

Solving the inequalities (4.24), we obtain

(4.25) $$w \geqq \{\nu + \lambda + \frac{\nu(\mu - \lambda)}{\nu + \lambda}\}T .$$

Now, we consider the problem which is better a complete observation or a Poisson sampling when we need the observation cost. Let C_0, C_1 be the observation cost per unit time by complete observation and the observation cost per an observation by Poisson sampling, respectively. Then the cost of observing the process over $[0,T]$ by complete observation is given by C_0T and the cost of observing the process until the w-th observation point by Poisson sampling is given by C_1w. From the inequality (4.25), if

$$(4.26) \qquad C_1 \leqq \frac{C_0}{\nu + \lambda + \dfrac{\nu(\mu - \lambda)}{\nu + \lambda}} ,$$

then the cost for Poisson sampling is less than that for complete observation, in order to get the same variances. In such a case Poisson sampling is better than the complete observation. In Fig. 1, we showed the maximum value of C_1/C_0 in the case where the cost for Poisson sampling with parameter ν is less than that for complete observation, when $\mu = 1$ and $\rho = \lambda/\mu$ is 0.5, 0.7 and 0.9.

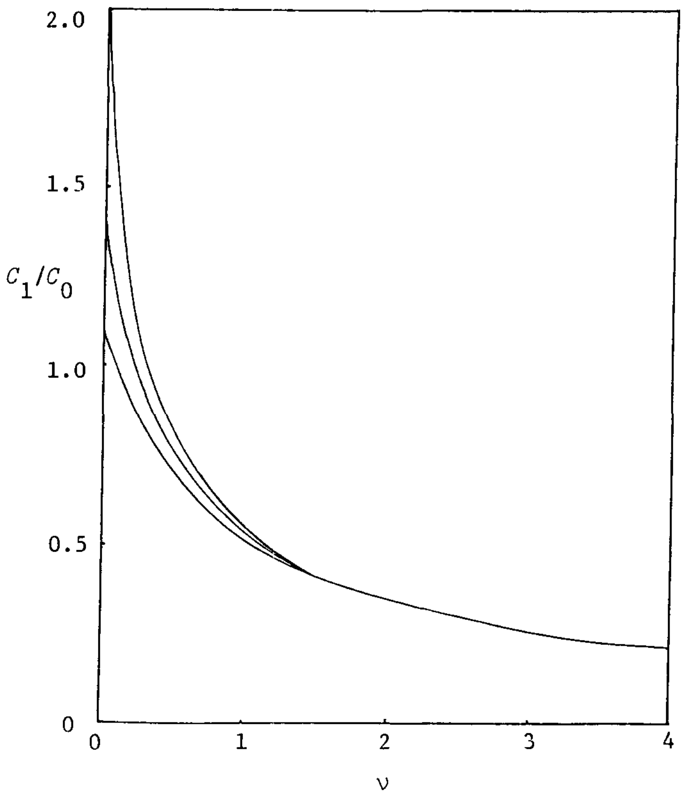

Fig. 1

Acknowledgement

I wish to thank Professor Hidenori Morimura for his invaluable suggestions and helpful comments.

References

[1] Basawa, I. V.: Maximum Likelihood Estimation of Parameters in Renewal and Markov-Renewal Processes. *Austral. J. Statist.* Vol. 16 (1974), 33-43.

[2] Billingsley, P: *Statistical Inference for Markov Processes.* Univ. Chicago Press, 1961.

[3] Clarke, A. B.: Maximum Likelihood Estimates in a Simple Queue. *Anal. Math. Statist.* Vol. 28 (1957), 1036-1040.

[4] Kingman, J. F. C.: Poisson Counts for Random Sequence of Events. *Anal. Math. Statist.* Vol. 34 (1963), 1217-1232.

[5] Reynolds, J. F.: On Estimating the Parameters of a Birth-Death Process. *Austral. J. Statist.* Vol. 15 (1973), 35-43.

[6] Wolff, R. W.: Problems of Statistical Inference for Birth and Death Queueing Models. *Opns. Res.* Vol. 13 (1965), 343-357.

Yutaka BABA: Department of Information Sciences, Tokyo Institute of Technology Oh-okayama, Meguro-ku, Tokyo 152, Japan.