The Efficiency of Conditional MLE for Pure Birth Processes

Jong-ook Yoon[1)]Joo Hwan Kim[2)]

ABSTRACT

The present paper is devoted to a study of the performance, in large samples, of a conditional maximum likelihood estimator(CMLE) for the parameter λ in a pure birth processes(PBP). To conduct the conditional inference for the PBP, we drove the likelihood function of time-inhomogeneous Poisson processes. The limiting distributions of CMLE under the likelihoods L_t or $\overline{L_t}$ are investigated. We found that the CMLE is asymptotically efficient with respect to the both L_t or $\overline{L_t}$ under the efficiency criterion of Weiss & Wolfowitz(1974).

key words: pure birth process, conditional likelihood, conditional maximum likelihood estimator, efficiency

1) Lecturer, Department of Statistics, Dongguk University, Seoul, 100-715, Korea
2) Associate Professor, Department of Science and Information Science, Dongguk University, Kyongju 780-714, Korea

1. Introduction

- Notation

X_t : state of a process at time t

$p_{i,k}(s,t) = \Pr\{X_t = k \mid X_s = i\}$: transition probability

$PBP(\lambda, q)$: pure birth process with parameter λ and initial state q

$L_t(\lambda)$: likelihood function of $PBP(\lambda, q)$

$L_t(\lambda, \omega)$: likelihood of Poisson process with intensity $(q\lambda\omega e^{\lambda t}, q)$

$L_t \equiv L_t(\lambda, \eta)$: likelihood obtained from $1-1$ transformation $\eta = \log(q\lambda\omega)$

$\overline{L}_t \equiv \overline{L}_t(\lambda)$: mixture likelihood

$L_t^c \equiv L_t^c(\lambda)$: conditional likelihood

$\mathcal{L}_t \equiv \mathcal{L}_t(\lambda, \psi)$: reparameterized likelihood

$\hat{\lambda}_t^c$: conditional MLE of $PBP(\lambda, q)$

● Transition Probability

Let's denote the transition probability

$$p_{i,k}(s,t) = \Pr\{X_t = k \mid X_s = i\}, \quad t \geq s\ ;\ i, k \in I_{[0,\infty)}$$

where $I_{[0,\infty)} = \{0,1,2,\cdots\}$, and $t,\ s \in R_{[0,\infty)}$.

● Infinitesimal Condition

$$\Pr\{X_{t+h} = k \mid X_t = i\} = \begin{cases} i\lambda h + o(h), & k = i+1, \\ 1 - i\lambda h + o(h), & k = i, \\ o(h), & \text{otherwise}, \end{cases} \tag{1.1}$$

where $h \geq 0$, $i = 0, 1, 2, \cdots;\ \lambda \geq 0$, $\lim_{h \to 0} o(h)/h = 0$, and degenerate at $X_0 = q$, that is $P(X_0 = q) = 1$.

● objects of this paper

1. Derive the likelihood functions for the pure birth process and time-inhomogeneous Poisson process.

2. Investigate the limiting distribution of CMLE under two likelihood functions.

3. Check the efficiency of CMLE under the Weiss & Wolfiwitz criterion.

2. Pure Birth Process

2.1 Transition probabilities

Since $PBP(\lambda, q)$ is a time-inhomogeneous process with Markov properties, it is satisfied

$$p_{i,k}(s, t) = p_{i,k}(0, t-s), \qquad t-s \geq 0.$$

- Chapman-Kolmogorov equation

$$p_{i,k}(0,t+h) = \sum_{j=1}^{\infty} p_{i,j}(0,t)\, p_{j,k}(t,t+h)$$

$$= p_{i,k-1}(0,t)\, p_{k-1,k}(t,t+h) + p_{i,k}(0,t)\, p_{k,k}(t,t+h)$$

$$= \begin{cases} (k-1)\lambda h p_{i,k-1}(0,t) + (1-k\lambda h)\, p_{i,k}(0,t), & k>i \\ (1-i\lambda h)\, p_{i,i}(0,t), & k=i\ . \end{cases}$$

- Kolmogorov differential equation

$$\frac{d}{dt} p_{i,i}(0,t) = -i\lambda p_{i,i}(0,t)$$

$$\frac{d}{dt} p_{i,k}(0,t) = -k\lambda p_{i,k}(0,t) + (k-1)\lambda p_{i,k-1}(0,t), \quad k>i,$$

where the initial coditions are $p_{i,i}(0,0)=1$, $p_{i,k}(0,0)=0$, $(k\neq i)$.

- The transition probabilities of the $PBP(\lambda, q)$

$$p_{i,k}(0,t)=\binom{k-1}{k-i} e^{-i\lambda t}(1-e^{-\lambda t})^{k-i}, \qquad k\geq i ,$$

which is easily recognized as a negative binomial distribution.

Hence the expected value and variance of $PBP(\lambda, q)$ are, respectively,

$$E(X_t) = qe^{\lambda t}, \qquad Var(X_t)=qe^{\lambda t}(e^{\lambda t}-1).$$

- Define $B_t=X_t-q$, which is the number of birth in time intervals $[0,t)$. Then

$$E(B_t) = q\frac{1-e^{-\lambda t}}{e^{-\lambda t}}, \qquad Var(B_t) = q\frac{1-e^{-\lambda t}}{e^{-2\lambda t}}.$$

- Let's define the sojourn time Z_j of a process with Markov properties as

$$Z_j=T_j-T_{j-1},\ j=1,2,\cdots;\ T_0=0,$$

where T_j is a epoch time of j-th jump occurrence. The distribution of Z_j, $j=1,2,\cdots$ as follows;

Lemma 2.1 For $PBP(\lambda, q)$, the distributions of $Z_1, Z_2, \cdots$ are

$$\Pr(Z_j \le z) = 1 - e^{-j\lambda z}, \qquad z \ge 0 \ ; \ j = q, q+1, \cdots. \tag{2.1}$$

Proof: omitted

2.2 Likelihood function

Lemma 2.2 If a $PBP(\lambda, q)$ $\{X_t;\ t \ge 0\}$ is observed continuously over the time intervals $[0, t)$, then the likelihood is

$$L_t(\lambda) = c\lambda^{B_t} e^{-\lambda S_t}, \tag{2.2}$$

where $B_t = X_t - q$, $S_t = \int_0^t X_u du$ and $c = q(q+1)\cdots(X_t - 1)$.

Proof: omitted

- We see immediately that (B_t, S_t) is minimally sufficient for λ, and that the statistic is of dimension two.

- mixture likelihood: (2.2) can be rewritten as

$$\overline{L_t} = \overline{L}_t(\lambda) = \exp\{(tX_t - S_t)\lambda + (X_t - q)\log(\lambda e^{-\lambda t}) - \lambda t\}$$

2.3 Transformation into Poisson Process

(Harris, 1963) If $\{X_t\} \sim PBP(\lambda, q)$, then there exist a random variable W such that, as $t \to \infty$

$$W_t \equiv X_t / EX_t = X_t / (qe^{\lambda t}) \xrightarrow{a.s.} W. \tag{2.3}$$

The distribution of W is $\text{Gamma}(q, q^{-1})$, that is with density $f_W(\omega; q) = q^q \omega^{q-1} e^{-q\omega} / \Gamma(q)$, $\omega > 0$, where $\Gamma(q) = \int_0^\infty x^{(q-1)} e^{-x} dx$ and $EW = 1$.

The following result is due to Kendall(1966) when $q=1$. Keiding(1974) generalized the result for

$q > 1$.

<u>Theorem 2.1</u> Conditionally on W, X_t is a time-inhomogeneous Poisson process with intensity parameter $q\lambda W e^{\lambda t}$ and $X_0 = q$, that is, $E(X_t - q \mid W) = q W(e^{\lambda t} - 1)$.

<u>Theorem 2.2</u> Let $\{X_t, t \geq 0\}$ be a Poisson process with intensity $q w \lambda e^{\lambda t} (\lambda > 0)$ and $X_0 = q$ and define $S_t = \int_0^t X_u du$. Then, $X_t e^{-\lambda t} \xrightarrow{a.s.} q\omega$ and $S_t e^{-\lambda t} \xrightarrow{a.s.} q\omega / \lambda$, as $t \to \infty$.

[See Keiding(1974) for detail.]

<u>Lemma 2.3</u> Let $\{X_t; t \geq 0\}$ be a time-inhomogeneous Poisson process with intensity $q w \lambda e^{\lambda t} (\lambda > 0)$ and $X_0 = q$. Then the likelihood function is

$$L_t(\lambda, \omega) = \exp\left\{ (t X_t - S_t)\lambda + (X_t - q) \log(q\lambda\omega) - \left(\frac{e^{\lambda t} - 1}{\lambda} \right) q\lambda\omega \right\}, \qquad (2.4)$$

where $S_t = \int_0^t X_u du$. The statistic $(t X_t - S_t, X_t - q)$ is jointly minimal sufficient for (λ, ω).

Proof: omitted

● Denote $U_t = U(X_t) = tX_t - S_t$, and $V_t = V(X_t) = X_t - q$. (2.5)

3. Conditional MLE

3.1 Conditional Likelihood and CMLE

Condition 1 There exists a pair of minimal sufficient statistics $(U_t, V_t) = (U(X_t), V(X_t))$ for $(\lambda, \log(q\lambda\omega))$.

Condition 2 There is a one-to-one transformation $(\lambda, \log(q\lambda\omega)) \rightarrow (\lambda, \eta)$ such that the conditional likelihood for given $V(X_t)$ is free from η.

Then the likelihood $L_t(\lambda, \omega)$ can be expressed as $L_t(\lambda, \omega) \propto \exp\{\lambda U_t + \eta V_t + k_t(\lambda, \eta)\}$, where $k_t(\lambda, \eta)$ is a random function such that $|k_t(\lambda, \eta)| < \infty$. The distribution satisfying Condition 1 and 2 is two-parameter exponential family.

The Likelihood $L_t(\lambda, \omega)$ in (2.4) can be expressed by 1-1 transformation $(\lambda, \log(q\lambda\omega)) \rightarrow (\lambda, \eta)$

satisfying Condition 2.

● likelihood obtained from 1-1 transformation by condition 2

$$L_t \equiv L_t(\lambda, \eta) = \exp\{\lambda U_t + \eta V_t - k_t(\lambda, \eta)\} , \tag{3.1}$$

where U_t and V_t are defined in (2.5), $\eta = \log(q\lambda\omega)$ and,

$$k_t \equiv k_t(\lambda, \eta) = \left(\frac{e^{\lambda t}-1}{\lambda}\right) e^{\eta}.$$

Lemma 3.1 For given likelihood L_t, the conditional likelihood $L_t^c(\lambda)$ for $U_t | V_t$ is

$$L_t^c \equiv L_t^c(\lambda) = \exp\left\{\lambda U_t + V_t \log\left(\frac{\lambda}{e^{\lambda t}-1}\right)\right\}. \tag{3.2}$$

Proof: Since $X_t \sim PoiPro(q\lambda\omega e^{\lambda t}, q)$, $V_t = X_t - q \sim Poipro(q\lambda\omega e^{\lambda t}, 0)$.

The marginal distribution of V_t becomes

$$p_{V_t}(v_t;\lambda,\eta) = \frac{1}{v_t!}\exp\left\{-e^{\eta}\left(\frac{e^{\lambda t}-1}{\lambda}\right)\right\}\left(\frac{e^{\eta}(e^{\lambda t}-1)}{\lambda}\right)^{v_t}, \quad v_t=0,1,\cdots \tag{3.3}$$

From these, L_t^c is derived.

● The likelihood L_t is factorized as $L_t = L_t^c \times P_{V_t}(v_t;\lambda,\eta)$.

From Lemma 3.1, Lemma 3.2 is obtained by using L_t^c.

Lemma 3.2 The CMLE λ_t^c of λ for $PBP(\lambda,q)$ is

$$\lambda_t^c = f_t^{-1}(U_t/V_t), \tag{3.4}$$

where $f_t(\lambda) = -\frac{1}{\lambda} + \frac{te^{\lambda t}V_t}{e^{\lambda t}-1}$, and U_t, V_t are given in (2.5).

Proof: omitted

Condition 3 There exist 1-1 reparameterization $(\lambda, \eta) \to (\lambda, \phi)$ such that $p_{V_t}(v_t; \lambda, \eta)$ depends on (λ, η) only through ϕ.

- reparameterized likelihood $\mathcal{L}_t(\lambda, \phi)$

$$\mathcal{L}_t = \mathcal{L}_t(\lambda, \phi) = \exp\left\{\lambda U_t + V_t \log\left(\frac{\lambda}{e^{\lambda t} - 1}\right)\right\} \exp\{V_t \log \phi - \phi\}.$$

Let $\phi = \left(\frac{e^{\lambda t} - 1}{\lambda}\right) e^{\eta}$, and substitute $\eta = \log \phi + \log\left(\frac{\lambda}{e^{\lambda t} - 1}\right)$ into L_t in (3.1). Then $\mathcal{L}_t$ is derived.

- reparameterized likelihood $\mathcal{L}_t$ can be expressed

$$\mathcal{L}_t = L_t^c \times p_{V_t}(v_t; \phi)$$

3.2 Limiting distribution of λ_t^c

● Recall $Var(U_t) = \frac{\partial^2 k_t}{\partial \lambda^2}$, $Var(V_t) = \frac{\partial^2 k_t}{\partial \eta^2}$, $Cov(U_t, V_t) = \frac{\partial^2 k_t}{\partial \lambda \partial \eta}$ (exponential family)

Condition 4 The function $k_t(\lambda, \eta)$ in (3.1) has continuous second partial derivatives $\partial^2 k_t / \partial \lambda^2 > 0$, $\partial^2 k_t / \partial \eta^2 > 0$, both increase to infinity and

$$\left(\frac{\partial^2 k_t}{\partial \lambda \partial \eta}\right)\left\{\left(\frac{\partial^2 k_t}{\partial \lambda^2}\right)\left(\frac{\partial^2 k_t}{\partial \eta^2}\right)\right\}^{-1/2} \to \rho(\lambda, \eta),$$

where ρ is continuous in (λ, η) and $|\rho| < 1$.

Theorem 3.1 Under condition 3 and 4 model $\mathcal{L}_t$, as $t \to \infty$, λ_t^c is consistent for λ, and

$$\left(\frac{\partial^2 k_t}{\partial \lambda^2}\right)^{1/2}(\lambda_t^c - \lambda) \to N(0, (1-\rho^2)^{-1}),$$

in distribution.

<u>Theorem 3.2</u> Under condition 3 and 4, as $t \to \infty$,

$$\{ Var(U_t | V_t) \}^{1/2} (\hat{\lambda}_t^c - \lambda) \to N(0,1),$$

both under $\mathcal{L}_t$ and $\overline{L}_t$.

3.3 Information in $\overline{L}_t$, $\mathcal{L}_t$

Denote

$\overline{I}_t(\lambda) \equiv E_\lambda (- \partial^2 \log \overline{L}_t / \partial \lambda^2)$: Fisher Information in $\overline{L}_t$

$Var(U_t | V_t) = - \dfrac{\partial^2 \log L_t^c}{\partial \lambda^2}$: observed information w.r.t. L_t^c

$\mathcal{I}_t(\lambda, \psi) \equiv E_\lambda (- \partial^2 \log \mathcal{L}_t / \partial \lambda^2)$: Fisher Information in $\mathcal{L}_t$

Then,

$$\frac{Var(U_t | V_t)}{\mathcal{I}_t(\lambda, \psi)} = \frac{V_t}{\psi} \xrightarrow{a.s.} 1,$$

$$\frac{Var(U_t|V_t)}{\overline{I}_t(\lambda)} = \frac{V_t}{(e^{\lambda t}-1)} \xrightarrow{a.s.} W$$

since V_t is Poisson with mean $\psi=(e^{\lambda t}-1)e^{\eta}\lambda^{-1}$,

3.4 Asymptotic efficiency of λ_t^c

Let C be the class of all consistent estimators T_t of λ such that, under $\mathcal{I}_t$,

$$\lim_{t\to\infty}\Pr\{\mathcal{I}_t^{1/2}(\lambda,\psi)(T_t-\lambda)\le u\}$$

exist for each real u, where $\mathcal{I}^{1/2}(\lambda,\psi)$ is the Fisher information about λ contained in $\mathcal{I}_t$.

Definition 3.1(Weiss & Wolfowitz(1974)) An estimator $T_t^*\in C$ is said to be asymptotically efficient with respect to C under $\mathcal{I}_t(\lambda,\psi)$ if, for a fixed $h>0$,

$$\lim_{t\to\infty}\Pr\left\{ \mathcal{I}_t^{1/2}(\lambda,\psi)\,|T_t^* - \lambda| < h\right\}$$

$$\geq \lim_{t\to\infty}\Pr\left\{ \mathcal{I}_t^{1/2}(\lambda,\psi)\,|T_t - \lambda| < h\right\}$$

for all (λ, ψ) under $\mathcal{L}_t$, where T_t is any member of C.

Theorem 3.3 Under condition 1 and 2, λ_t^c is asymptotically efficient with respect to the class C, under $\mathcal{L}_t$, in the sense of Definition3.1.

Proof: omitted

Definition 3.2(Weiss & Wolfowitz(1974)) An estimator $T_t^* \in C$ is said to be asymptotically efficient with respect to C under $\overline{L_t}(\lambda)$ if, for a fixed $h>0$,

$$\lim_{t\to\infty} \Pr\{ \overline{I}_t^{1/2}(\lambda)\,|T_t^* - \lambda| < h\}$$
$$\geq \lim_{t\to\infty} \Pr\{ \overline{I}_t^{1/2}(\lambda)\,|T_t - \lambda| < h\}$$

for all λ under $\overline{L}_t$, where T_t is any member of C.

Theorem 3.4 Under condition 1 and 2, λ_t^c is asymptotically efficient with respect to the class C, under $\overline{L}_t$, in the sense of Definition3.2.

4. Concluding Remarks

It is shown that the CMLE is fully efficient with respect to both the reference likelihoods $\overline{L}_t$and L_t. The conditional likelihood function L_t^c is identical under both $\overline{L}_t$and L_t and it is free from the nuisance parameter. In addition to the mathematical convenience and intuitive appeal of L_t^c, we have shown that the resulting estimator λ_t^c suffers no loss of efficiency in large samples.

references

1. Andersen, E.B.(1970). Asymptotic Properties of Conditional Maximum likelihood Estimators. *Journal of the Royal Statistical Society*, Series B, 32, 283-301.
2. Basu, D.(1977). On the Elimination of Nuisance Parameters. *Journal of the American Statistical Association* 72, 355-366.
3. Efron, B.(1975) Defining the Curvature of Statistical Problem (with Applications to Second Order Efficiency). *The Annals of Statistics* 3, 1189-1242.
4. Ferguson, H.(1992). Asymptotic Properties of a Conditional Maximum Likelihood Estimator. *The Canadian Journal of Statistics* 20, 1992.
5. Godambe, V.P.(1980). On Sufficiency and Ancillarity in the Presence of a Nuisance Parameter. *Biometrika* 67, 155-162.
6. Godambe, V.P.(1984). On Ancillarity and Fisher Information in the Presence of a Nuisance Parameter. *Biometrika* 71, 626-629.
7. Harris, T.E.(1963). *The Theory of Branching Processes*. Springer-Verlag, Berlin
8. Kalbfleisch, J.D. and Sprott, D.A.(1970). Application of Likelihood Methods to Models Involving Large Numbers of Parameters. *Journal of the Royal Statistical Society,* Series B, 32, 175-208.

9. Keiding, N.(1974). Estimation in the birth process. *Biometrika* 61, 71-80.
10. Lehmann, E.L.(1983). *Theory of Point Estimation.* John Wiley & Sons, Inc. New York.
11. Severini, T.A.(1993). Local Ancillarity in the Presence of a Nuisance Parameter *Biometrika* 80, 305-320.
12. Severini, T.A.(1994). On the Approximate Elimination of Nuisance Parameters by Conditioning. *Biometrika* 81, 649-661.
13. Sprott, D.A.(1980) Maximum Likelihood in Small Samples: Estimation in the Presence of Nuisance Parameters. *Biometrika* 67, 515-523.
14. Weiss, L. and Wolfowitz, J.(1974). *Maximum Probability Estimators and Related Topics.* Lecture Notes in Mathematics 424, Springer-Verlag Berlin.