Blockwise empirical likelihood and efficiency for semi-Markov processes

Priscilla E. Greenwood
School of Mathematics and Statistics
Arizona State University
Tempe, AZ 85287-1804, USA
`pgreenw@math.la.asu.edu`

Ursula U. Müller
Department of Statistics
Texas A&M University
College Station, TX 77843-3143, USA
`uschi@stat.tamu.edu`

Wolfgang Wefelmeyer *
Mathematical Institute
University of Cologne
50931 Cologne, Germany
`wefelm@math.uni-koeln.de`

Abstract

Suppose we have linear constraints on the stationary distribution of the embedded Markov renewal process of a semi-Markov process on an arbitrary state space. Then we can improve an empirical estimator by empirical likelihood weights. Since the observations are dependent, an optimal choice of weights is determined by weighting averages over disjoint blocks of observations with slowly increasing length. The improved empirical estimator is efficient. We also introduce two additively corrected empirical estimators that are asymptotically equivalent to the weighted empirical estimator.

Keywords. Martingale approximation, perturbation expansion, local asymptotic normality, asymptotically linear estimator, improved empirical estimator, asymptotically efficient estimator.

*Presenting and corresponding author.

1 Introduction

Let $(X_0,T_0),\dots,(X_n,T_n)$ be observations of a Markov renewal process in an arbitrary state space E with σ-algebra $\mathcal{E}$. Write $V_j = T_j - T_{j-1}$ for the inter-arrival times. Let $Q(x,dy)$ denote the transition distribution of the embedded Markov chain $X_0, X_1, \dots$, and let $R(x,y,dv)$ denote the conditional distribution of the inter-arrival time V_j given $(X_{j-1},X_j) = (x,y)$. Then the observations $(X_1,V_1),\dots,(X_n,V_n)$ follow a Markov chain with transition distribution from $(X_{j-1},V_{j-1}) = (x,u)$ to $(X_j,V_j) = (y,v)$ that factors as $S(x,dy,dv) = Q(x,dy)R(x,y,dv)$ and does not depend on u.

Suppose that the embedded Markov chain is ergodic with stationary distribution π. Then (X_{j-1},X_j,V_j) also has a stationary distribution. It can be written as $P(dx,dy,dv) = \pi(dx)Q(x,dy)R(x,y,dv)$. The transition distribution $S(x,dy,dv)$, and hence the joint law of the Markov chain $(X_1,V_1),(X_2,V_2),\dots$ is determined by $P(dx,dy,dv)$. In order to estimate the joint law, it therefore suffices to estimate expectations $Pf = E[f(X_{j-1},X_j,V_j)]$ for a sufficiently large class of functions f on $E^2 \times [0,\infty)$. The natural estimator of Pf is the empirical estimator

$$\mathbb{P}f = \frac{1}{n}\sum_{j=1}^{n} f(X_{j-1},X_j,V_j).$$

Now assume that the stationary distribution P fulfills a linear constraint $Ph = 0$ for some vector-valued function h on $E^2 \times [0,\infty)$. We will show in Section 3 that we can use this constraint in three different ways to improve the empirical estimator $\mathbb{P}f$. One is the blockwise empirical likelihood, which was introduced by Kitamura (1997) to construct confidence intervals for weakly dependent (discrete-time) processes. It leads to the *blockwise weighted* empirical estimator defined in (3.2). A simpler estimator is the *blockwise additively corrected* empirical estimator defined in (3.4). A version without blocks is the *additively corrected* empirical estimator (3.1). It was first introduced for Markov chains in Müller, Schick and Wefelmeyer (2001).

All three estimators are asymptotically equivalent and asymptotically efficient in the sense of a nonparametric version of the convolution theorem of Hájek and Le Cam. In our setting, the theorem is proved in Section 2. It leads to a characterization of efficient estimators, and also suggests a construction of such an estimator, namely the additively corrected empirical estimator (3.1).

Consider the semi-Markov process Z_t, $t \ge 0$ corresponding to the above Markov renewal process. Suppose we observe a path Z_t, $0 \le t \le n$. Set $N = \max\{j : T_j \le n\}$. The results of Sections 2 and 3 carry over to the semi-Markov process by replacing the sample size n by the random sample size N. For an appropriate version of local asymptotic normality we refer to Greenwood and Wefelmeyer (1996).

2 A characterization of efficient estimators

We can "parametrize" the Markov renewal process in different ways: by the transition distribution S, by the conditional distributions Q and R, and by the stationary distribution P. What is convenient depends on the structure of the model, and also on the functional we want to estimate. Greenwood, Müller and Wefelmeyer (2004) and Müller, Schick and Wefelmeyer (2008) consider functionals of Q and R and parametrize by Q and R. For Markov chains, Greenwood and Wefelmeyer (1996) consider a nonparametric semi-Markov model and linear functionals Pf and parametrize by S. Bickel (1993) and Bickel and Kwon (2001) suggest parametrization by P. This would be convenient for the functional Pf and the constraint $Ph = 0$ considered here. However, a Markov renewal process is characterized by a property of the *transition distribution* $S(x, y, dv)$ of the Markov chain (X_j, V_j): namely that it does not depend on the previous value V_{j-1} of the inter-arrival time. This is why it is more convenient to "parametrize" with S. We will see that the property also implies that ergodicity of the Markov renewal process can be described in terms of ergodicity of the embedded Markov chain. We will use L_2-ergodicity; the results also hold under the more flexible V-ergodicity, for which we refer to Meyn and Tweedie (1993) and Schick and Wefelmeyer (2002).

We assume that the embedded chain is stationary with stationary distribution π. We want to estimate expectations Pf of unbounded functions f. The constraint $Ph = 0$ also typically involves an unbounded h, for example when we assume that the embedded Markov chain has mean zero. This is why we assume $L_2(\pi)$-ergodicity rather than uniform ergodicity. Let $\|g\| = \pi(g^2)^{1/2}$ denote the norm of a function $g \in L_2(\pi)$, and let $\|K\| = \sup\{\|Kg\| : \|g\| = 1\}$ denote the corresponding operator norm of a kernel K on $E \times \mathcal{E}$. Let $\Pi(x, dy) = \pi(dy)$ denote the *stationary projection* of Q. Exponential $L_2(\pi)$-ergodicity of the embedded Markov chain is implied by the condition $\|Q - \Pi\| < 1$. If the embedded Markov chain X_j is stationary with stationary distribution π, then the Markov chain (X_j, Y_j) is stationary with stationary distribution $\kappa = \int \pi(dx)S(x, \cdot, \cdot)$, and $\|Q - \Pi\| < 1$ implies $\|S - \Pi_\kappa\|_\kappa < 1$, where $\|\cdot\|_\kappa$ denotes the $L_2(\kappa)$-norm and Π_κ denotes the stationary projection of S.

In order to characterize efficient estimators, we show that the distribution of the observations $X_0, (X_1, V_1), \dots, (X_n, V_n)$ is *locally asymptotically normal* in the following nonparametric sense. Let

$$U = \{u \in L_2(P) : \ Su = 0\}.$$

For each $u \in U$ we can construct a perturbation S_{nu} of S that is *Hellinger differentiable* with derivative u,

$$P\Big(\frac{dS_{nu}}{dS} - 1 - \frac{1}{2}n^{-1/2}u\Big)^2 = o(n^{-1}).$$

Write $P^{(n)}$ and $P^{(n)}_{nu}$ for the joint law of the observations under S and S_{nu}, respectively. Let

N denote a standard normal random variable. The following theorem shows nonparametric local asymptotic normality. For Markov chains, different proofs are in Penev (1991), Bickel (1993), and Greenwood and Wefelmeyer (1995). For Markov step processes see Höpfner, Jacod and Ladelli (1990) and Höpfner (1993).

Theorem 1. *Assume that* $\|Q - \Pi\| < 1$*. Let* $u \in U$*. Then*

$$\log \frac{dP_{nu}^{(n)}}{dP^{(n)}} = n^{-1/2} \sum_{j=1}^{n} u(X_{j-1}, X_j, V_j) - \frac{1}{2} Pu^2 + o_P(n^{-1/2})$$

and

$$n^{-1/2} \sum_{j=1}^{n} u(X_{j-1}, X_j, V_j) \Rightarrow (Pu^2)^{1/2} N.$$

From Kartashov (1985a, 1985b, 1996) we obtain the following *perturbation expansion* for $g \in L_2(P)$:

$$n^{1/2}(P_{nu}g - Pg) \to P(uAg) \quad \text{for } u \in U \tag{2.1}$$

with

$$Ag(x, y, v) = g(x, y, v) - S_x g + \sum_{t=0}^{\infty} (Q_y^t Sg - Q_x^{t+1} Sg).$$

It is a feature of Markov renewal processes that powers of S do not appear in the operator A. The reason is that the transition distribution S does not depend on the previous value of the inter-arrival time.

For $g \in L_2(P)$, the *martingale approximation* of Gordin (1969) and Gordin and Lifšic (1978) is

$$n^{-1/2} \sum_{j=1}^{n} (g(X_{j-1}, X_j, V_j) - Pg) = n^{-1/2} \sum_{j=1}^{n} Ag(X_{j-1}, X_j, V_j) + o_P(1), \tag{2.2}$$

By a central limit theorem for martingales, the right side is asymptotically normal with variance $P[(Ag)^2]$.

Now suppose that the stationary distribution P fulfills the constraint $Ph = 0$ for some d-dimensional vector of functions $h \in L_2^d(P)$. The constraint also holds for the perturbed stationary distribution P_{nu}, and we obtain from (2.1), applied to the components of h, that $P(uAh) = 0$. The perturbations u are therefore constrained to

$$U_h = \{u \in U : \ P(uAh) = 0\}.$$

the stochastic expansion in Theorem 1 involves a norm $(Pu^2)^{1/2}$ on U. By the perturbation expansion (2.1), applied to f, we see that the functional Pf is *differentiable* at P with *gradient* $Af \in U$ in the sense that

$$n^{1/2}(P_{nu}f - Pf) \to P(uAf) \quad \text{for } u \in U.$$

The *canonical gradient* under the constraint $Ph = 0$ is the projection of Af onto U_h. By definition of U_h, the space U has the orthogonal decomposition $U = U_h \oplus [Ah]$, where $[Ah]$ is the linear span of the components of Ah. Hence the canonical gradient of Pf can be written $u_h = Af - u_h^\perp$, where $u_h^\perp$ is the projection of u_h onto $[Ah] = U_h^\perp$. This projection is of the form $u_h^\perp = c_h^\perp Ah$ with

$$c_h^\perp = P(Ah \cdot Ah^\top)^{-1} P(Ah \cdot Af).$$

We obtain the canonical gradient

$$u_h = Af - P(Af \cdot Ah^\top) P(Ah \cdot Ah^\top)^{-1} Ah.$$

An estimator $\hat{\vartheta}$ is called *asymptotically linear* for Pf at P with *influence function* v if $v \in U$ and

$$n^{1/2}(\hat{\vartheta} - Pf) = n^{-1/2} \sum_{j=1}^n v(X_{j-1}, X_j, V_j) + o_P(1).$$

Given the constraint $Ph = 0$, an estimator $\hat{\vartheta}$ is called *regular* for Pf at P with *limit* L if L is a random variable such that

$$n^{1/2}(\hat{\vartheta} - P_{nu} f) \Rightarrow L \quad \text{for } u \in U_h.$$

The convolution theorem of Hájek (1970) then says that $L = P(u_h^2)^{1/2} N + M$ with M independent of N. This justifies calling $\hat{\vartheta}$ *efficient* if $L = P(u_h^2)^{1/2} N$. The convolution theorem also implies the following characterization of efficient estimators. We refer to Bickel, Klaassen, Ritov and Wellner (1998).

Theorem 2. *Assume that $\|Q - \Pi\| < 1$. Let $f \in L_2(P)$ and $h \in L_2(P)$ with $Ph = 0$. Under the constraint $Ph = 0$, an estimator $\hat{\vartheta}$ for Pf is efficient at P if and only if $\hat{\vartheta}$ is asymptotically linear for Pf at P with influence function equal to the canonical gradient u_h,*

$$n^{1/2}(\hat{\vartheta} - Pf) = n^{-1/2} \sum_{j=1}^n u_h(X_{j-1}, X_j, V_j) + o_P(1). \tag{2.3}$$

The asymptotic variance of such an estimator $\hat{\vartheta}$ is

$$P(u_h^2) = P((Af)^2) - P(Af \cdot Ah^\top) P(Ah \cdot Ah^\top)^{-1} P(Ah \cdot Af).$$

This should be compared with the asymptotic variance of the empirical estimator $\mathbb{P}f = (1/n) \sum_{j=1}^n f(X_{j-1}, X_j, V_j)$, which is $P((Af)^2)$. In the next section we construct three efficient estimators for Pf under the constraint $Ph = 0$.

3 Three efficient estimators

The characterization of efficient estimators for Pf under the constraint $Ph = 0$, given in Theorem 2, immediately suggests a construction that replaces the unknown expectations in the canonical gradient by empirical estimators. Define the *additively corrected empirical estimator* of Pf as

$$\mathbb{P}_a f = \mathbb{P}f - \hat{\gamma}^\top \hat{\Sigma}^{-1} \frac{1}{n} \sum_{j=1}^{n} h(X_{j-1}, X_j, V_j), \tag{3.1}$$

where $\hat{\gamma}$ is an empirical estimator for $P(Ah \cdot Af)$,

$$\begin{aligned}
\hat{\gamma} = {} & \frac{1}{n} \sum_{j=1}^{n} h(X_{j-1}, X_j, V_j) f(X_{j-1}, X_j, V_j) \\
& + \sum_{k=1}^{m} \frac{1}{n-k} \sum_{j=1}^{n-k} \big(h(X_{j-1}, X_j, V_j) f(X_{j+k-1}, X_{j+k}, V_{j+k}) \\
& \qquad + h(X_{j+k-1}, X_{j+k}, V_{j+k}) f(X_{j-1}, X_j, V_j) \big),
\end{aligned}$$

and $\hat{\Sigma}$ is an empirical estimator for $P(Ah \cdot A^\top h)$,

$$\begin{aligned}
\hat{\Sigma} = {} & \frac{1}{n} \sum_{j=1}^{n} h(X_{j-1}, X_j, V_j) h^\top (X_{j-1}, X_j, V_j) \\
& + 2 \sum_{k=1}^{m} \frac{1}{n-k} \sum_{j=1}^{n-k} h(X_{j-1}, X_j, V_j) h(X_{j+k-1}, X_{j+k}, V_{j+k}).
\end{aligned}$$

Theorem 3. *Assume that $\|Q - \Pi\| < 1$. Let $f \in L_2(P)$ and $h \in L_2^d(P)$ with $Ph = 0$. Under the constraint $Ph = 0$, the estimator $\mathbb{P}_a f$ is asymptotically linear in the sense of (2.3) for Pf at P with influence function u_h, and therefore regular and efficient.*

The proof is similar to that of the corresponding result for Markov chains in Müller, Schick and Wefelmeyer (2001), and we omit it.

A different improvement of the empirical estimator $\mathbb{P}f$ consists in weighting it appropriately. The empirical likelihood of Owen (1988, 2001) works only for independent observations. For (weakly) dependent observations we can use the blockwise empirical likelihood introduced by Kitamura (1997). Decompose (the initial section of) the time points $1, \dots, n$ into $\nu = [n/m]$ disjoint blocks of length m, where m tends slowly to infinity with the sample size n. Set

$$F_i = \frac{1}{m} \sum_{k=1}^{m} f(X_{(i-1)m+k-1}, X_{(i-1)m+k}, V_{(i-1)m+k}),$$

$$H_i = \frac{1}{m} \sum_{k=1}^{m} h(X_{(i-1)m+k-1}, X_{(i-1)m+k}, V_{(i-1)m+k}).$$

Then, for n a multiple of m, the empirical estimator for Pf can be written

$$\mathbb{P}f = \frac{1}{\nu}\sum_{i=1}^{\nu} F_i;$$

otherwise this holds only up to $o_P(n^{-1/2})$. The *blockwise weighted empirical estimator* for Pf is

$$\mathbb{P}_w f = \frac{1}{\nu}\sum_{i=1}^{\nu} \frac{F_i}{1+\lambda^\top H_i} \tag{3.2}$$

with random vector λ chosen such that $1+\lambda^\top H_i > 0$ and

$$\sum_{i=1}^{\nu} \frac{H_i}{1+\lambda^\top H_i} = 0.$$

Theorem 4. *Assume that $\|Q-\Pi\| < 1$. Let $f \in L_2(P)$ and $h \in L_2^d(P)$ with $Ph = 0$. Under the constraint $Ph = 0$, the estimator $\mathbb{P}_w f$ admits the stochastic expansion*

$$\mathbb{P}_w f = \mathbb{P}f - \frac{1}{\nu}\sum_{i=1}^{\nu} F_i H_i^\top \Big(\frac{1}{\nu}\sum_{i=1}^{\nu} H_i H_i^\top\Big)^{-1} \frac{1}{\nu}\sum_{i=1}^{\nu} H_i + o_P(n^{-1/2}) \tag{3.3}$$

and is asymptotically linear in the sense of (2.3) for Pf at P with influence function u_h. In particular, $\mathbb{P}_w f$ is regular and efficient.

Proof. Similarly as in Owen (1988, 2001) we show that

$$\lambda = \frac{1}{\nu}\sum_{i=1}^{\nu} \Big(\frac{1}{\nu}\sum_{i=1}^{\nu} H_i H_i^\top\Big)^{-1} \frac{1}{\nu}\sum_{i=1}^{\nu} H_i + o_P(n^{-1/2}).$$

An expansion of $\mathbb{P}_w f$ then leads to (3.3). □

An immediate consequence of (3.3) and Theorem 4 is that the *blockwise additively corrected empirical estimator*

$$\mathbb{P}_b f = \mathbb{P}f - \frac{1}{\nu}\sum_{i=1}^{\nu} F_i H_i^\top \Big(\frac{1}{\nu}\sum_{i=1}^{\nu} H_i H_i^\top\Big)^{-1} \frac{1}{\nu}\sum_{i=1}^{\nu} H_i \tag{3.4}$$

of Pf is regular and efficient under the constraint $Ph = 0$.

References

[1] Bickel, P. J. (1993). Estimation in semiparametric models. In: *Multivariate Analysis: Future Directions* (C. R. Rao, ed.), 55–73, North-Holland, Amsterdam.

[2] Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models.* Springer, New York.

[3] Bickel, P. J. and Kwon, J. (2001). Inference for semiparametric models: Some questions and an answer (with discussion). *Statist. Sinica* **11**, 863–960.

[4] Gordin, M. I. (1969). The central limit theorem for stationary processes. *Soviet Math. Dokl.* **10**, 1174–1176.

[5] Gordin, M. I. and Lifšic, B. A. (1978). The central limit theorem for stationary Markov processes. *Soviet Math. Dokl.* **19**, 392–394.

[6] Greenwood, P. E., Müller, U. U. and Wefelmeyer, W. (2004). Efficient estimation for semiparametric semi-Markov processes. *Comm. Statist. Theory Methods* **33**, 419-435.

[7] Greenwood, P. E. and Wefelmeyer, W. (1995). Efficiency of empirical estimators for Markov chains, *Ann. Statist.*, **23**, 132–143.

[8] Greenwood, P. E. and Wefelmeyer, W. (1996). Empirical estimators for semi-Markov processes. *Math. Meth. Statist.* **5**, 299–315.

[9] Hájek, J. (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. verw. Gebiete* **14**, 323–330.

[10] Höpfner, R. (1993). On statistics of Markov step processes: representation of log-likelihood ratio processes in filtered local models. *Probab. Theory Related Fields* **94**, 375–398.

[11] Höpfner, R., Jacod, J. and Ladelli, L. (1990). Local asymptotic normality and mixed normality for Markov statistical models. *Probab. Theory Related Fields* **86**, 105–129.

[12] Kartashov, N. V. (1985a). Criteria for uniform ergodicity and strong stability of Markov chains with a common phase space. *Theory Probab. Math. Statist.* **30**, 71–89.

[13] Kartashov, N. V. (1985b). Inequalities in theorems of ergodicity and stability for Markov chains with common phase space. I. *Theory Probab. Appl.* **30**, 247–259.

[14] Kartashov, N. V. (1996). *Strong Stable Markov Chains.* VSP, Utrecht.

[15] Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Ann. Statist.* **25**, 2084–2102.

[16] Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability.* Springer, London.

[17] Müller, U. U., Schick, A. and Wefelmeyer, W. (2001). Improved estimators for constrained Markov chain models. *Statist. Probab. Lett.* **54**, 427–435.

[18] Müller, U. U., Schick, A. and Wefelmeyer, W. (2008). Optimality of estimators for misspecified semi-Markov models. *Stochastics* **80**, 181–196.

[19] Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–249.

[20] Owen, A. B. (2001). *Empirical Likelihood.* Monographs on Statistics and Applied Probability 92, Chapman & Hall/CRC, London.

[21] Penev, S. (1991). Efficient estimation of the stationary distribution for exponentially ergodic Markov chains. *J. Statist. Plann. Inference* **27**, 105–123.

[22] Schick, A. and Wefelmeyer, W. (2002). Estimating joint distributions of Markov chains. *Stat. Inference Stoch. Process.* **5**, 1–22.