ALMOST SURE PARAMETER ESTIMATION AND CONVERGENCE RATES FOR HIDDEN MARKOV MODELS

Robert J. Elliott[1]

and

John B. Moore[2]

(1) Department of Mathematical Sciences,
University of Alberta, Edmonton, Alberta, Canada T6G 2G1
relliott@gpu.srv.ualberta.ca
and
Department of Applied Mathematics
University of Adelaide
Adelaide, SA 5005
Australia
relliott@maths.adelaide.edu.au

(2) Department of Systems Engineering,
Australian National University, Canberra A.C.T. 0200, Australia
jmoore@alfred.anu.edu.au

Abstract: A continuous time version of Kronecker's Lemma is established and used to give rates of convergence for parameter estimates in Hidden Markov Models.

Acknowledgements: The support of NSERC grant A7964 is gratefully acknowledged. Professor Moore wishes to thank the Department of Mathematical Sciences, University of Alberta, for its hospitality in July 1996 when this work was carried out.

1. Introduction

Kronecker's Lemma is well known in discrete time; in continuous time this appears not to be the case. The continuous version is established in the Appendix. Using Kronecker's Lemma convergence results are established for parameter estimates in a Hidden Markov Model. In [1] the results are motivated by maximum likelihood arguments; here the methods are more direct. The discrete time version of Kronecker's Lemma can be found in [3]. Our convergence results are related to those of Meyer [2]. We do not discuss recursive, adaptive parameter estimation in this paper; this will be treated elsewhere.

2. Hidden Markov Model

For $t \geq 0$ consider a finite state, time homogeneous Markov chain X. Without loss of generality the state space of the chain can be taken to be the set of unit vectors $S = \{e_1, e_2, \ldots, e_N\}$ where

$$e_i = (0, \ldots, 1, 0, \ldots, 0)' \in R^N.$$

X is defined on a probability space $(\Omega, \mathcal{F}, P)$, and $\mathcal{F}_t = \sigma\{X_s : s \leq t\}$. Then it is well known (see Elliott et al. [2]), that, if A is the generator, or Q-matrix, associated with X, then

$$X_t = X_0 + \int_0^t AX_s\, ds + M_t. \tag{2.1}$$

In fact for $t \geq s$ $E[X_t \,|\, \mathcal{F}_s] = E[X_t \,|\, X_s] = e^{A(t-s)}X_s$. Therefore, with I the $N \times N$ identity matrix,

$$\begin{aligned} E[M_t - M_s \,|\, \mathcal{F}_s] &= E[X_t - X_s - \int_s^t AX_u du \,|\, \mathcal{F}_s] \\ &= (e^{A(t-s)} - I - \int_s^t Ae^{A(u-s)} du)X_s \\ &= O \in R^N. \end{aligned}$$

Here M is an $(\mathcal{F}_t, P)$ martingale. The chain X is not observed directly. Rather there is an observed process y of the form

$$y_t = \int_0^t c(X_s)\, dx + b_t, \tag{2.2}$$

where b_t is a standard Brownian motion. For simplicity we suppose here the observations y are scalar. Further, $c(X_s)$ is given by the scalar product $\langle \boldsymbol{c}, X_s \rangle$ for some vector $\boldsymbol{c} = (c_1, \ldots, c_N)'$.

As in [1], the dynamics (2.2) can be constructed by supposing that, under some "ideal" reference measure $\overline{P}$, y itself is a standard Brownian motion and X is a Markov chain with generator A.

Write:

$$G_t = \sigma\{x_s,\, y_s : s \leq t\} \quad \text{and} \quad \Lambda_t := \exp\biggl(\int_0^t c(X_s)\, db_s - \frac{1}{2} \int_0^t c(X_s)^2\, ds \biggr).$$

A probability measure P can be defined by putting

$$\left. \frac{dP}{d\overline{P}} \right|_{G_t} = \Lambda_t.$$

It is then standard, (see [1]), that under P, y has dynamics (2.2) and X remains a Markov chain with generator A.

Recall $dX_t = AX_t\, dt + dM_t$. Therefore, with prime $'$ denoting transpose,

$$\int_0^t dX_s \cdot X'_{s-} = A \int_0^t X_s X'_s\, ds + \int_0^t dM_s \cdot X'_{s-}. \tag{2.3}$$

This is a matrix valued equation. Note $X_s X_s' = \operatorname{diag} X_s$ and write

$$J_t = \int_0^t dX_s \cdot X_{s-}'$$

$$O_t = \int_0^t X_s X_s' \, ds = \int_0^t \operatorname{diag} X_s \, ds$$

$$\overline{M}_t = \int_0^t dM_s \cdot X_{s-}'.$$

Also,

$$\widehat{J}_t = E[J_t \mid Y_t], \qquad \widehat{O}_t = E[O_t \mid Y_t]$$

$$\widehat{M}_t = E[\overline{M}_t \mid Y_t].$$

Recall filtered equations for $\widehat{J}$ and $\widehat{O}$ are given in [1]. Conditioning (2.3) on Y_t we have

$$\widehat{J}_t = A\widehat{O}_t + \widehat{M}_t$$

and an estimate for A is, therefore,

$$\widehat{A}_t := \widehat{J}_t \, \widehat{O}_t^{-1}.$$

Further, the error term is

$$\widehat{A}_t - A = \widehat{M}_t \, \widehat{O}_t^{-1}.$$

3. Convergence

Suppose $\rho(t)$, $t \geq 0$, is an increasing, positive, deterministic function of t such that

$$\lim_{t \to \infty} \int_0^t \rho(s)^{-1} \, ds = \lambda < \infty.$$

For example, $\rho(t) = \max(1, t(\log t)(\log \log t)^\alpha)$ for $\alpha > 1$. Note, from the Corollary of

Kronecker's Lemma in the Appendix, $\lim_{t\to\infty} t\rho(t)^{-1} = 0$.

Theorem 3.1. $\lim_{t\to\infty} \rho(t)^{-1/2}\widehat{M}_t = 0$ a.s.

Proof: Consider $R_t := \int_0^t \rho(s)^{-1/2}\, dM_s\, X'_{s-}$.

Then R is a martingale,

$$R_t R'_t = \int_0^t R_{s-} dR'_s + \int_0^t dR_s \cdot R'_{s-} + \sum_{0<s\leq t} \Delta R_s \cdot \Delta R'_s$$

and

$$\begin{aligned} E[\mathrm{Tr}\, R_t R'_t] &= \sum_{0<s\leq t} E[\mathrm{Tr}\, \Delta R_s \cdot \Delta R'_s] \\ &= \sum_{0<s\leq t} E[\mathrm{Tr}\, \rho(s)^{-1} \Delta M_s X'_{s-} \cdot X_{s-} \Delta M'_s] \\ &= \sum_{0<s\leq t} E[\mathrm{Tr}\, \rho(s)^{-1} \Delta X_s \Delta X'_s] \\ &= \sum_{0<s\leq t} \rho(s)^{-1} E[\Delta X'_s \Delta X_s] \\ &= -2E\Big[\int_0^t \rho(s)^{-1} X'_{s-} dX_s \Big]. \end{aligned}$$

With $dX_s = AX_s ds + dM_s$, this gives

$$E[\mathrm{Tr}\, R_t R'_t] \leq \Big| E\Big[2\int_0^t \rho(s)^{-1} \langle X_s, a\rangle\, ds \Big]\Big|$$

where $\mathbf{a} = (a_{11}, a_{22}, \ldots, a_{NN})$ is the vector comprising the diagonal elements of A. Consequently,

$$\begin{aligned} E[\mathrm{Tr}\, R_t R'_t] &\leq 2 \max_i |a_{ii}| \int_0^t \rho(s)^{-1}\, ds \\ &\leq 2 \max_i |a_{ii}|\, \lambda, \end{aligned}$$

so

$$\lim_{t\to\infty} E[\operatorname{Tr} R_t R_t'] \leq 2 \max_i |a_{ii}| \, \lambda$$

and R_t is a square integrable (matrix) martingale. Therefore, from the matringale convergence theorem (see Doob [1]),

$$\lim_{t\to\infty} \int_0^t \rho(s)^{-1/2} \, dM_s \, X_{s-}'$$

exists and is finite, almost surely. From the vector corollary of the continuous time Kronecker Lemma (see Appendix)

$$\lim_{t\to\infty} \rho(t)^{-1/2} \int_0^t dM_s \, X_{s-}' = 0 \quad \text{a.s.}$$

Now these random variables $\rho(t)^{-1/2} \int_0^t dM_s \, X_{s-}'$ converge to zero in L^2 as, writing

$$\overline{R}_t = \rho(t)^{-1/2} \int_0^t dM_s \, X_{s-}'$$

$$E[\operatorname{Tr} \overline{R}_t \, \overline{R}_t'] \leq 2 \max_i |a_{ii}| t \rho(t)^{-1}.$$

This has limit zero using the Corollary of Kronecker's Lemma in the Appendix. Therefore, because trace provides a norm on the space of matrices, $\lim_{t\to\infty} \overline{R}_t = 0$ a.s. Because the $\overline{R}_t$ are bounded in L^2 we can conclude that

$$\lim_{t\to\infty} E[\overline{R}_t \mid Y_t] = 0$$

and the result follows.

Corollary 3.2. *Take $\rho(t) = \max(1, t(\log t)(\log\log t)^\alpha)$ for $\alpha > 1$. Then, as $t \to \infty$*

$$\rho(t)^{-1} \widehat{M}_t = o\,(t^{-1/2} (\log t)(\log\log t)^\alpha).$$

Corollary 3.3. *Suppose the chain X satisfies the 'excitation condition'*

$$\rho(t)^{-1}\widehat{O}_t > K > 0$$

for some matrix K, and ρ is as in Corollary 3.2. Then $\widehat{A}_t - A = o(t^{-1/2}(\log t)(\log\log t)^{\alpha})$.

Proof:

$$\begin{aligned}\widehat{A}_t - A_t &= \rho(t)^{-1}\widehat{M}_t \cdot \rho(t) \cdot \widehat{O}_t^{-1} \\ &\leq \rho(t)^{-1}\widehat{M}_t \cdot K^{-1}\end{aligned}$$

and the result follows from Corollary 3.2.

Remarks 3.4. Note that $E[\operatorname{Tr} \overline{M}_t \, \overline{M}_t'] = O(t)$ so that, if the excitation condition holds, convergence in mean square is also $O(t)$.

4. Observation Parameter

Recall $dy_s = c(X_s)\,ds + db_s$. So

$$\int_0^t dy_s\, X_s' = c\int_0^t \operatorname{diag} X_s\, ds + \int_0^t db_s\, X_s'.$$

Write

$$T_s = \int_0^t dy_s\, X_s', \qquad Q_t = \int_0^t db_s\, X_s'$$

with, again,

$$O_t = \int_0^t \operatorname{diag} X_s\, ds, \qquad \widehat{T}_t = E[T_t \mid Y_t],$$

$$\widehat{O}_t = E[O_t \mid Y_t], \qquad \widehat{Q}_t = E[Q_t \mid Y_t].$$

Filtered estimates for $\widehat{T}$ are given in [1]. An estimate for c is

$$\widehat{c}_t = \widehat{T}_t\, \widehat{O}_t^{-1}$$

and the error is $\widehat{c}_t - c = \widehat{Q}_t\,\widehat{O}_t^{-1}$. Suppose $\rho(t)$ is a function as in Section 3.

Then following result is proved similarly to Theorem 3.1.

Theorem 4.1. $\lim_{t\to\infty} \rho(t)^{-1/2}\widehat{Q}_t = 0$ a.s.

Proof: Consider $\overline{Q}_t := \int_0^t \rho(s)^{-1/2}\, db_s\, X'_s$ and write $\widehat{\overline{Q}}_t := E[\overline{Q}_t \mid Y_t]$. Then

$$E[\overline{Q}_t\,\overline{Q}'_t] = \int_0^t \rho(s)^{-1}\, ds < \lambda < \infty.$$

Consequently, $\overline{Q}_t$ is a square integrable martingale and

$$\lim_{t\to\infty} \int_0^t \rho(s)^{-1/2}\, db_s\, X'_s$$

exists almost surely. Using Kronecker's Lemma

$$\lim_{t\to\infty} \rho(t)^{-1/2} \int_0^t db_s\, X'_s = 0 \text{ a.s.}$$

These random variables $\rho(t)^{-1/2}\int_0^t db_s\, X'_s$ converge to zero in L^2, and so are uniformly integrable. Consequently, their almost sure convergence enables us to deduce that

$$\lim_{t\to\infty} \rho(t)^{-1/2}\widehat{Q}_t = 0 \quad \text{a.s.}$$

Corollary 4.2. *Suppose* $\rho(t) = \max(1, t(\log t)(\log\log t)^{\alpha})$ *for* $\alpha > 1$, *and that* X *satisfies the 'excitation condition'*

$$\lim_{t\to\infty} t\widehat{O}_t^{-1} > 0.$$

Then $\widehat{c}_t - c = o\,(t^{-1/2}(\log t)(\log\log t)^{\alpha})$.

APPENDIX:. A Version of Kronecker's Lemma in Continuous Time

We first prove the result for scalar processes.

For $0 \le r < \infty$ suppose x_r is a real valued, right continuous function with left limits and of bounded variation.

For example, $x_r = \int_0^r y_s \, ds$ where y is locally integrable. Suppose $u_r > 0$ is a right continuous nondecreasing function. Consequently, $\lim_{r\to\infty} u_r$ exists and either $\lim_{r\to\infty} u_r = u < \infty$ or $\lim_{r\to\infty} u_r = +\infty$.

Consider $z_t = \int_0^t \frac{dx_r}{u_{r-}}$.

Theorem. *If* $\lim_{t\to\infty} z_t = \xi < \infty$ *exists and* $0 \le t_0 < t$, *then*

$$\lim_{t\to\infty} \frac{1}{u_t} (x_t - x_{t_0})$$

exists.

If $\lim_{r\to\infty} u_r = +\infty$, *then this limit is* 0.

Proof: For any s, $t_0 < s < t$

$$x_t - x_s = \int_s^t u_{r-} \, dz_r = \int_s^t u_{r-} \, d(z_r - z_s)$$

$$= u_t(z_t - z_s) - \int_s^t (z_r - z_s) \, du_r.$$

a) Suppose first that $\lim_{t\to\infty} u_t = u < \infty$. Then

$$\begin{aligned} |x_t - x_s| &\le u_t |z_t - z_s| + \left| \int_s^t (z_r - z_s) \, du_r \right| \\ &\le u_t |z_t - z_s| + \sup_{s \le r \le t} |z_r - z_s| (u_t - u_s) \\ &\le 2u_t \sup_{r \ge s} |z_r - z_s| \qquad (1) \\ &\le 2u \sup_{r \ge s} |z_r - z_s| \end{aligned}$$

because u is nondecreasing. Using the Cauchy condition for z, for any $\varepsilon > 0$ there is an s_ε

such that, if $r \geq s \geq s_\varepsilon$, then $|z_r - z_s| < \frac{\varepsilon}{u}$.

Consequently, if $r \geq s \geq s_\varepsilon$

$$|x_t - x_s| < \varepsilon.$$

Therefore, x satisfies the Cauchy condition and is convergent.

b) Suppose now that $\lim_{t\to\infty} u_t = +\infty$. Give $\varepsilon > 0$. From the Cauchy condition for z there is an s'_ε such that, if $r \geq s'_\varepsilon$

$$|z_{s'_\varepsilon} - z_r| < \frac{\varepsilon}{3}.$$

Consequently, $\sup_{r\geq s'_\varepsilon} |z_r - z_{s'_\varepsilon}| \leq \frac{\varepsilon}{3}$ and, if $t \geq s'_\varepsilon$ from (1)

$$\frac{1}{u_t}\,|x_t - x_{s'_\varepsilon}| \leq \frac{2\varepsilon}{3}.$$

Now $\lim_{t\to\infty} u_t = \infty$, so there is t_ε such that, if $t > t_\varepsilon$, $u_t \geq 3|x_{s'_\varepsilon} - x_{t_0}|/\varepsilon$. That is, $\frac{1}{u_t}|x_{s'_\varepsilon} - x_{t_0}| \leq \frac{\varepsilon}{3}$. Now for $t_0 < s < t$

$$\left|\frac{1}{u_t}(x_t - x_{t_0})\right| \leq \frac{1}{u_t}|x_s - x_{t_0}| + \frac{1}{u_t}|x_t - x_s|.$$

So if $t > \max(t_\varepsilon, s'_\varepsilon)$

$$\frac{1}{u_t}\,|x_t - x_{t_0}| \leq \frac{1}{u_t}\,|x_{s'_\varepsilon} - x_{t_0}| + \frac{1}{u_t}\,|x_t - x_{s'_\varepsilon}| \leq \varepsilon$$

and the result is proved.

Corollary. *If* $\lim_{t\to\infty} \int_0^t a_r\,dr < \infty$ *where* $a_r > 0$ *is nonincreasing, then* $\lim_{t\to\infty} ta_t = 0$, *i.e.,*

$$a_t = o\left(\frac{1}{t}\right).$$

Corollary. *The result extends to vector integrands and integrators.*

References

1. J.L. Doob, *Stochastic Processes*, Wiley and Sons: New York, 1953.

2. R. J. Elliott, L. Aggoun and J. B. Moore, *Hidden Markov Models: Estimation and Control*, Applications of Mathematics Volume 29, Springer-Verlag, New York, 1995.

3. P. A. Meyer, *Quelques inégalités sur les martingales d'après Dubins et Freedman,* Lecture Notes in Math. Vol. 124 (1970), Springer-Verlag, Berlin, Heidelberg, New York, 162–169.

4. J. Neveu, *Discrete Parameter Martingales*, North Holland, Amsterdam (1975).