BAYESIAN STATISTICS 7, pp. 000–000
J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid,
D. Heckerman, A. F. M. Smith and M. West (Eds.)

The Markov Modulated Poisson Process and Markov Poisson Cascade with Applications to Web Traffic Modeling

STEVEN L. SCOTT
University of Southern California, USA
sls@usc.edu

PADHRAIC SMYTH
University of California, Irvine, USA
smyth@ics.uci.edu

SUMMARY

A Markov modulated Poisson Process (MMPP) is a Poisson process whose rate varies according to a Markov process. The nonhomogeneous MMPP developed in this article is a natural model for point processes whose events combine irregular bursts of activity with predictable (e.g. daily and hourly) patterns. We show how the MMPP may be viewed as a superposition of unobserved Poisson processes that are activated and deactivated by an unobserved Markov process. The MMPP is a continuous time model which may also be viewed as a discretely indexed nonstationary hidden Markov model by viewing intervals between events as a sequence of dependent random variables. The HMM representation allows one to probabilistically reconstruct the latent Markov and Poisson processes using a set of forward-backward recursions. The recursions allow MMPP parameters to be estimated either by an EM algorithm or by a rapidly mixing Markov chain Monte Carlo algorithm which uses the recursions for data augmentation. The Markov-Poisson cascade (MPC) is an MMPP whose underlying Markov process obeys certain restrictions which uniquely order the event rates for the observed process. The ordering avoids a possible label switching issue without slowing down the rapidly mixing algorithms we use to implement the model. We apply the MPC to a data set containing click rate data for individual computer users browsing through the World Wide Web. Because the complete data posterior distribution for the MPC is a product of exponential family distributions we are able to incorporate data from multiple users into a hierarchical model using existing methods from hierarchical Poisson regression.

Keywords: HIDDEN MARKOV MODEL; POINT PROCESS; BURST; FORWARD-BACKWARD RECURSIONS; CONTINUOUS TIME; NONHOMOGENEOUS POISSON PROCESS.

1. INTRODUCTION

The Markov modulated Poisson process (MMPP) is a doubly stochastic Poisson process (Cox 1955; Gutíerrez-Peña and Nieto-Barajas 2003) whose rate varies according to a Markov process. This article decomposes the MMPP into a superposition of latent Poisson processes which are activated and deactivated by a latent Markov process. The result is a natural model for point process data where events combine predictable patterns with irregular bursts of activity. The MMPP is most frequently seen in queuing theory (Du 1995; Olivier and Walrand 1994) but it has other interesting applications. Davison and Ramesh (1996) applied a discretized MMPP to a binary time series of precipitation data by numerically optimizing the discretized MMPP likelihood. Scott (1998) used the MMPP to model criminal intrusions on a telephone network. Other uses of the MMPP exist in environmental, medical, industrial, and sociological research. Inference for MMPP parameters has received little attention because most applications of the MMPP assume known model parameters. Turin (1996) proposed an EM algorithm for finding maximum likelihood estimates of MMPP parameters. Scott(1999) provided a Bayesian method for inferring the parameters of a stationary two state MMPP. This article extends Scott (1999) to the nonstationary case with an arbitrary number of states. To our knowledge this is the first treatment of inference for the nonhomogeneous MMPP. We show how the MMPP can be viewed as a superposition of latent Poisson processes, which in turn may be expressed as a nonhomogeneous, discretely indexed hidden Markov model (HMM) by partitioning time into intervals between observed events. Expressing the MMPP as an HMM allows one to probabilistically reconstruct the latent Markov and Poisson processes using a set of forward-backward recursions. The recursions allow MMPP parameters to be estimated through familiar latent variable methods such as the EM algorithm or MCMC data augmentation. The Markov-Poisson cascade (MPC) is a special case of MMPP that enforces an ordering of the state space of the underlying Markov process. The MPC maintains the ordering in a natural way, so that a potential label switching issue is avoided without slowing down our rapidly mixing MCMC algorithms or modifying the specification of model parameters as in Robert and Titterington (1998).

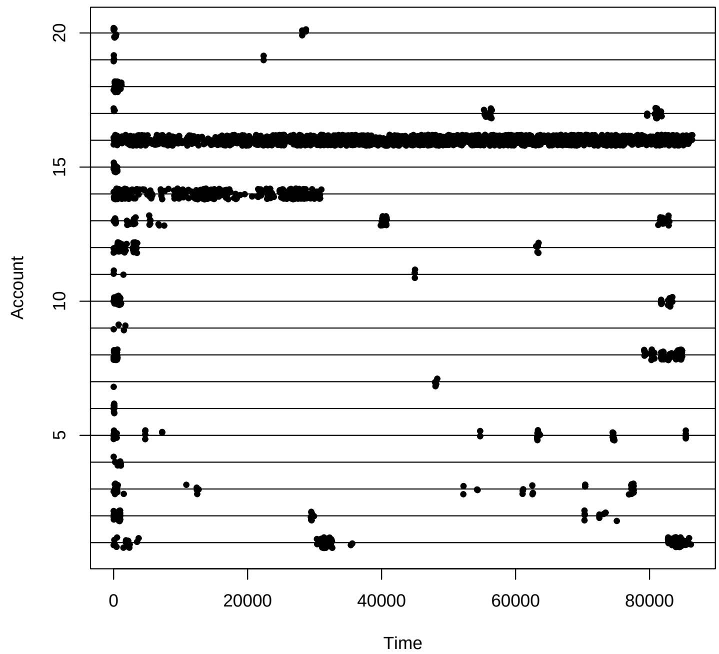


Figure 1. *Mouse click events for the first 20 computers in our data set. Time is measured in seconds from midnight.*

Our motivating example is a data set containing click rate (page request) data produced by 1025 computer users as they navigated through the World Wide Web over a 24 hour period. Figure 1 plots page request times for the first twenty users in our data set. The data are a convenience sample from a much larger data set collected by a software tool that automatically logs all page requests within each user's Web browser (with the user's permission). The data are unusual because they were collected on the "client side" at the individuals' computers. It is much easier, and thus more common, for Web traffic studies to be conducted using data obtained from Web servers configured to record the request time, the requested page, and the IP address of the requesting computer ("server side data"). Collecting data on individual users' computers eliminates several technical issues associated with server side data (Cooley *et al.* 1999). Among the difficulties are that an IP address does not uniquely identify a particular individual, the Web server can only see a user's behavior at one particular site, and some requested pages may not be recorded at the server side because they are provided by the user's memory cache instead of by the Web server.

Computer scientists are interested in page request click rates for academic and commercial reasons. Academically, it is hoped that a better understanding of how individuals use the Web will lead to improved design of Web sites, Web interfaces, and internet traffic management schemes. Commercially, online merchants are interested in click rates because they contain information about an individual's interest in the content of a Web page. Very rapid click rates suggest that an individual is searching for something which he has not found on the current page. Moderate click rates suggest the individual is absorbing content (e.g. reading or listening to music). When a user's click rate falls to zero his session has ended, so resources devoted to him should be directed elsewhere. Click rates can be combined with other information about the content which is being browsed in order to make inference about a user's state of mind. The hidden Markov model representation developed in Section 2 provides a straightforward method of incorporating additional page characteristics into the MMPP, leading to a more elaborate biometric model. For a more detailed discussion of client side Web traffic modeling consult Catledge and Pitkow (1995), Cunha, *et al.* (1995), Tauscher and Greenberg (1997), and Cockburn and McKenzie (2001).

The remainder of the article proceeds as follows. Section 2 presents a theory of inference for the MMPP and the MPC. Section 3 applies the model to the mouse click data. Section 4 provides a concluding discussion.

2. MODEL

2.1. *Decomposing the MMPP into Latent Poisson Processes*

Let $N(t)$ be a point process formed by the superposition of unobserved components $N_0(t), \dots,$ $N_{M-1}(t)$, each of which may sometimes be inactive. Each $N_m(t)$ is a Poisson process whose rate is $\lambda_m(t)$ when $N_m(t)$ is active and 0 when $N_m(t)$ is inactive. We assume that $\lambda_m(t)$ is a parametric or slowly varying nonparametric function supplied by the modeler. An unobserved continuous-time Markov process $A(t)$ determines which component processes are active at time t. Active component processes are independent of one another, and depend on $A(t)$ only through activation/deactivation. Because a variety of models could be used for $A(t)$, depending on physical considerations, we defer discussion of our preferred model until Section 2.4, following material which is independent of the choice of A. For now we assume only that if $A(t) \neq A(t')$ then different subsets of $N_0, \dots, N_{M-1}$ are active at t and t'.

Heuristically, one imagines that $A(t)$ governs the beginnings and endings of "activity bursts" with intensity $\sum_m \lambda_m(t) I_A^m(t)$, where $I_A^m(t) = 1$ if $A(t)$ indicates process m is active and $I_A^m(t) = 0$ otherwise. Define an activity burst of level m to be an interval (t_0, t_1) such that $N_m(t)$ is active for $t_0 < t < t_1$, $N_m(t)$ is inactive just before t_0 and just after t_1, and events are generated at t_0 and t_1. By defining an activity burst as beginning and ending with its first and last events, we force $A(t)$ to remain constant between events, which eliminates the possibility of pathological "bursts" containing no events. Forcing activity bursts to begin and end with events is mathematically equivalent to assuming that each $A(t)$ transition produces an event. One may think of events produced by $A(t)$ as coming from Poisson processes $B_m(t)$ and $D_m(t)$, which produce the events corresponding to the birth and death of N_m. The current state of $A(t)$ determines whether $B_m(t)$ or $D_m(t)$ are inactive at time t. In particular, N_m cannot be born if it is already alive, and it cannot die if it is already dead. For most applications interest will focus on the case where $B_m(t)$ and $D_m(t)$ have rate functions, denoted $\beta_m(t)$ and $\delta_m(t)$, which are much smaller than $\lambda_m(t)$. The rates $\beta_m(t)$ and $\delta_m(t)$ derive from the "infinitesimal rates"

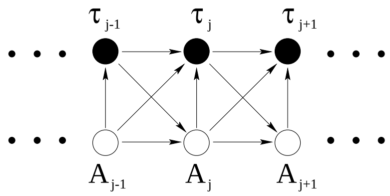


Figure 2. *DAG for the MMPP. Closed and open circles represent observed and unobserved quantities, respectively.*

of the generator matrix for $A(t)$. The MMPP is $N = \sum_m (B_m + N_m + D_m)$. Conditional on $A(t) = A$, $N(t)$ is a Poisson process with rate

$$\theta_A(t) = \sum_{m=0}^{M-1} I_A^m(t)\lambda_m(t) + I_B^m(t)\beta_m(t) + I_D^m(t)\delta_m(t) \tag{1}$$

where $I_B^m(t)$ and $I_D^m(t)$ indicate whether $B_m(t)$ and $D_m(t)$ are active at time t.

2.2. *Expressing the MMPP as a Hidden Markov Model*

The MMPP may be expressed as a nonstationary hidden Markov model where the observed data are the event times $\boldsymbol{\tau} = (\tau_1 < \cdots < \tau_n)$ produced by N, and the hidden Markov chain is the sequence $\{A_j = A(\tau_j)\}$. Expressing the MMPP as an HMM allows recursive procedures for HMM's to be used for calculating likelihood, calculating the marginal posterior distribution of each A_j given $\boldsymbol{\tau}$, and estimating model parameters using latent variable methods.

One intuitively understands the MMPP as an HMM because $\{A_j\}$ is a discretization of a continuous-time Markov process, and each τ_j is produced by a Poisson process whose rate is known over the interval (τ_{j-1}, τ_j) given (A_{j-1}, A_j). The formal proof that the MMPP is an HMM follows immediately from a well known fact about Poisson processes presented below as Theorem 1. Note that dependence on model parameters is notationally suppressed for probability calculations in this section.

Theorem 1. *Let $N_0, \ldots, N_M$ be independent Poisson processes with rate functions $\lambda_0(t)$, $\ldots$, $\lambda_M(t)$. Let $\Lambda_m^{(t_0)}(t) = \int_{t_0}^t \lambda_m(u)\, du$. Let T represent the time of the first event generated by any of $N_0, \ldots, N_M$ after time t_0, and let Y denote the index of the first process to produce an event. Then*

$$Pr(T > t, Y = m) = \int_t^\infty \lambda_m(u) \exp\left(-\sum_{r=0}^{M} \Lambda_r^{(t_0)}(u)\right) du.$$

Write τ_j^k for $(\tau_j, \ldots, \tau_k)$ and similarly for other vectors. Theorem 1 implies that $p(\tau_j, A_j \mid \tau_1^{j-1}, A_1^{j-1})$ depends only on (τ_{j-1}, A_{j-1}), a relationship illustrated in Figure 2. Figure 2 is not the traditional DAG associated with a hidden Markov model, but it is sufficiently close that the standard HMM recursions need only be slightly modified. The most important recursive procedure for HMM's is the forward-backward recursion, which calculates the marginal posterior distribution of each (A_{j-1}, A_j) transition conditional on $\boldsymbol{\tau}$. The forward recursion calculates $p_{jrs} = p(A_{j-1} = r, A_j = s \mid \tau_1^j)$. The backward recursion updates these distributions so that they condition on all $\boldsymbol{\tau}$. Theorem 1 implies the forward recursion for the MMPP is

$$p_{jrs} \propto p(\tau_j, A_{j-1} = r, A_j = s \mid \tau_1^{j-1}) = \theta_{rs}^*(\tau_j) \exp\left[-\Theta_r^{(\tau_{j-1})}(\tau_j)\right] \pi_{j-1}(r). \tag{2}$$

The proportionality in (2) is reconciled by $\sum_r \sum_s p_{jrs} = 1$. One recursively computes $\pi_{j-1}(r) = p(A_{j-1} = r \mid \tau_1^{j-1})$ as $\pi_j(s) = \sum_r p_{jrs}$. The symbol $\Theta_r^{(\tau_{j-1})}(\tau_j) = \int_{\tau_{j-1}}^{\tau_j} \theta_r(t)\, dt$ is the expected number of events in (τ_{j-1}, τ_j) given $A(\tau_{j-1}) = r$. If the (r, s) transition implies the birth or death of process m then $\theta^*_{rs}(\tau_j) = \beta_m(\tau_j)$ or $\delta_m(\tau_j)$, respectively. If no births or deaths took place then $\theta^*_{rs}(\tau_j) = \sum_m I_A^m(\tau_{j-1})\lambda_m(\tau_j)$, which is the sum of all λ_m's from processes which were active during (τ_{j-1}, τ_j).

The standard HMM backward recursion applies to the MMPP because the dependence among the A_j's in Figure 2, given $\boldsymbol{\tau}$, is first order Markov. Let $p'_{jrs} = p(A_{j-1} = r, A_j = s \mid \boldsymbol{\tau})$ and $\pi'_j(s) = p(A_j = s \mid \boldsymbol{\tau})$ denote the updated probabilities. The backward recursion begins with $\pi_n(s) = \pi'_n(s)$. Then $p'_{jrs} = p_{jrs}\pi'_j(s)/\pi_j(s)$ where $\pi'_j(s)$ is computed from p'_{j+1rs} in the previous step of the recursion.

2.3 *Parameter Estimation and Posterior Sampling*

The traditional role of the forward-backward recursion is to implement the E-step of an EM algorithm (Baum *et al.* 1970, Dempster *et al.* 1977). The complete data likelihood for the MMPP is

$$L_{com} = \prod_{j=1}^{n} \prod_{m=0}^{M-1} \lambda_m(\tau_j)^{y_{jm}} \beta_m(\tau_j)^{b_{jm}} \delta_m(\tau_j)^{d_{jm}} \exp\left(-\int_{\tau_{j-1}}^{\tau_j} \theta_A(t)\, dt\right) \tag{3}$$

where y_{jm}, b_{jm} and d_{jm} are 0/1 indicators revealing which of N_m, B_m, or D_m produced the event at τ_j. Equation (3) is log-linear in the missing indicators $(y_{jm}, b_{jm}, d_{jm}, I_A^m(\tau_{j-1}), I_B^m(\tau_{j-1}), I_D^m(\tau_{j-1}))$ so the E-step of EM simply replaces each indicator with its conditional probability given $\boldsymbol{\tau}$. All required probabilities are available from the forward backward-recursions. The only probability which requires any calculation to extract is $p(y_{jm} = 1 \mid \boldsymbol{\tau}) \propto \sum_r p(A_{j-1} = r, A_j = r \mid \tau) I^m_{A_j=r}(\tau_j)\lambda_m(\tau_j)$.

The forward-backward recursions may also be used to sample the missing data directly from its conditional distribution given $\boldsymbol{\tau}$ in a single Gibbs step (Scott 2002, Chib 1996). First draw A_n from $\pi_n(s)$. Then draw each A_j given $A_{j+1} = s$ from the distribution proportional to either p_{jrs} or p'_{jrs}. If $A_{j-1} \neq A_j$ then event j is either a birth or a death. If $A_{j-1} = A_j$ then draw $(y_{j0}, \ldots, y_{jM-1})$ from the distribution proportional to $I^m_{A_j}(\tau_j)\lambda_m(\tau_j)$.

Equation (3) says that if $\lambda_m(t)$, $\beta_m(t)$ and $\delta_m(t)$ have distinct parameters, then those parameters are independent in the complete data likelihood. Thus, the parameter estimation or simulation step for an EM or data augmentation algorithm simply involves repeated inference for the parameters of an ordinary Poisson process.

2.4. *Identifiability and The Markov Poisson Cascade*

Depending on physical considerations, there are several state spaces and transition probabilities one could use for $A(t)$. We propose the following model as a default, as it seems to balance physical, computational, and identifiability considerations.

Let $\mathcal{S} = \{0, \ldots, M-1\}$ be the state space, where $A(t) = m$ implies $N_0, \ldots, N_m$ are active and $N_{m+1}, \ldots, N_{M-1}$ are inactive. If $A(t) = m$ then all possible death processes $D_1, \ldots, D_m$ are active, but B_{m+1} is the only active birth process. We call an MMPP with this $A(t)$ a Markov-Poisson cascade (MPC) because the death of a process which is low in the hierarchy immediately deactivates all processes above it. Processes which are terminated by the death of a "lower" process do not generate events. If $A(t) = s$ then equation (1) for the MPC becomes

$$\theta_s(t) = \beta_{s+1}(t) + \sum_{m=0}^{s} (\lambda_m(t) + \delta_m(t)) \, .$$

When paired with a prior distribution enforcing $\lambda_m(t) > \beta_m(t)$ (which simply means that events inside a burst occur more frequently than the bursts themselves) the MPC avoids three identifiability issues associated with the MMPP: label switching, state collapsing, and role reversal. Label switching is an intrinsic feature of all finite mixture models. It occurs because the complete data likelihood is invariant to a permuation of the state labels. The MPC prevents label switching by enforcing $\theta_s(t) < \theta_{s+1}(t)$, provided $\beta_{s+1}(t) + \lambda_s(t) + \delta_s(t) > \beta_s(t)$. This is a very mild restriction which obtains with high posterior probability even under a weak prior. Note that one need not order the individual λ_m's, β_m's or δ_m's. The ordering is automatically enforced on θ_m.

State collapsing and role reversal occur when M is too large to be supported by the data. State collapsing occurs when the model eliminates a redundant state by setting $\lambda_m \approx 0$, which allows N_m to be active without producing any events. Parameters are more interpretable if unnecessary states are left unused. By preventing N_{m+1} from activating unless N_m is active, but allowing lower states to die, the MPC diminishes the prior activation probability of states at the top of the hierarchy. This causes unneccessary states to filter to the top where they remain unused.

Role reversal is a consequence of the symmetry between $\lambda_m(t)$ and $(\beta_m(t), \delta_m(t))$ in (3). Role reversal occurs because a long sequence of events produced by N_m has essentially the same likelihood as an alternating sequence of N_m births and deaths. Scott (1999) observed that a prior forcing $\beta_m < \lambda_m$ prevents role reversal.

3. APPLICATION

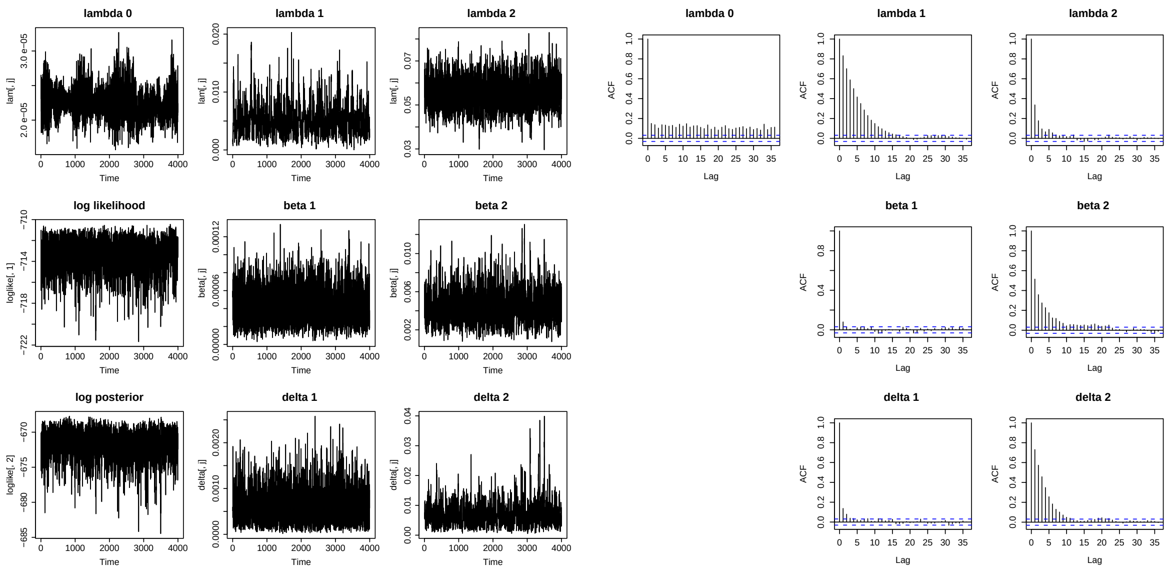
We assume user i in the click rate dataset follows an MPC with parameters $\phi_i = \{\lambda_{im}(t), \beta_{im}(t), \delta_{im}(t) : m = 0, \ldots, M-1\}$. For convenience we set $M = 3$, where the three states indicate respectively the absence of a Web session, a session with a slow click rate, and rapid clicking. In a serious application we would allow M to depend on i using any of several Bayesian methods for model selection or model averaging. We force $\beta_0(t) = \delta_0(t) = 0$ so that N_0 remains active as a baseline to catch isolated events. Each user's click stream almost certainly contains strong daily and hourly patterns, but these patterns are inestimable because only a single day has been observed. Therefore we assume all rates are constant, e.g. $\lambda_{im}(t) = \lambda_{im}$. Had a much longer time window been observed we could incorporate daily and hourly patterns into the click rates using the timing model proposed by Lambert *et al.* (2001), for example.

The familiar exponential family distributions underlying the MPC make it easy to embed MPC parameters in a hierarchical model. We assume the prior distribution

$$p(\phi_i, \ldots, \phi_n) = \prod_i \prod_m \text{Ga}(\lambda_{im} \mid a_{\lambda m}, b_{\lambda m}) \text{Ga}(\beta_{im} \mid a_{\beta m}, b_{\beta m}) \text{Ga}(\delta_{im} \mid a_{\delta m}, b_{\delta m}), \qquad (4)$$

where $\text{Ga}(\cdot \mid a, b)$ is the gamma distribution with mean a/b and variance a/b^2. The hyperparameters in (4) are interpretable as prior event counts and observation times. For example $a_{\beta m}$ is a prior number of births for N_m, and $b_{\beta m}$ is a prior amount of time spent waiting for N_m to be born. Following Christiansen and Morris (1997), we assume an improper uniform prior on each a/b and assume each $p(a) = z_0/(z_0 + a)^2$, which is a proper normalized distribution with no moments. Christiansen and Morris show this prior has good frequency properties in the context of Poisson regression. The only "tuning parameter" is z_0, which we set to the relatively uninformative value of 0.10.

Let $\boldsymbol{z}_i$ denote the missing indicators required to compute L_{com} for user i and let $\boldsymbol{\alpha}$ denote the set of $(a_{..}, b_{..})$ pairs in (4). We used an MCMC algorithm which cycles between sampling from $p(\boldsymbol{z}_i \,|\, \phi_i)$, $p(\phi_i \,|\, \boldsymbol{z}_i, \boldsymbol{\alpha})$ for each i, and from $p(\boldsymbol{\alpha}|\phi_1, \ldots, \phi_n)$ with $n = 1025$. We ran the algorithm for 5000 iterations and removed the first 1000 as burn-in. Figure 3 shows the remaining 4000 draws of ϕ_i for a sample account. The time series plots and autocorrelation functions in Figure 3 indicate rapid mixing attributable to the forward-backward recursions used in the data augmentation step.

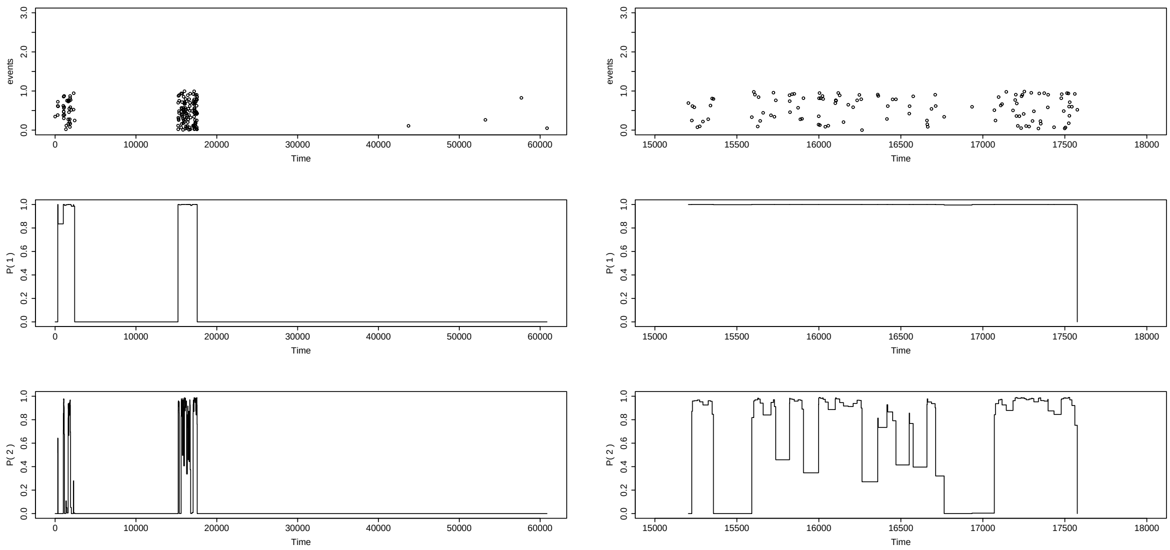


(a) *(b)*

Figure 3. *(a) Time series plots and (b) autocorrelations of 4000 MCMC draws from the posterior distribution of MPC parameters for a sample user.*

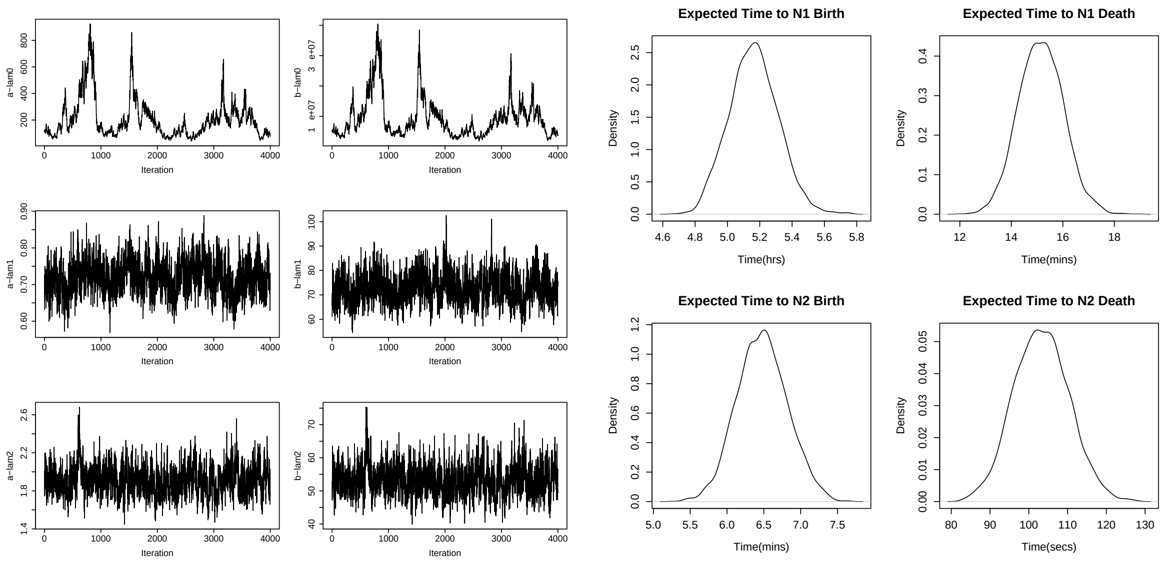
Figure 4 plots the data for the sample account, along with the posterior probability that N_1 and N_2 are active each event time. This account was selected because it showed several stray events in addition to the more typical bursts corresponding to Web sessions. We were concerned that the stray events might be interpreted as low intensity Web sessions. Instead, the posterior distribution of $A(t)$ in Figure 4 behaves as hoped. Stray events are attributed to N_0, while N_1 bursts persist across moderate gaps. Level 2 bursts switch more rapidly. The smallest definitive gap between level 2 bursts for this account is the four minute interval containing time 15500. The probabilities in Figure 4 are Bayesian in the sense that they average over the posterior distribution of ϕ_i. A corresponding plot created by running the forward-backward recursions conditional on a point estimate of ϕ_i is not shown because it is nearly identical. This suggests that an online merchant who lacks the time to implement an MCMC algorithm may safely base predictions on quicker empirical Bayes calculations.

The hierarchical model specification allows us to coherently estimate aggregate summaries of Web sessions by examining the posterior distributions of hyperparameters. Figure 5 shows MCMC sample paths for the prior parameters of $\{(a_{\lambda m}, b_{\lambda_m}) : m = 0, 1, 2\}$. All three (a, b) pairs have moved from their initial values of (1,1000), suggesting that there is information about rate parameters across several accounts. The only parameters that had trouble mixing were $(a_{\lambda 0}, b_{\lambda 0})$, because of the considerable probability mass close to $\lambda_0 = 0$. The posterior medians for $a_{\lambda 0}$ and $b_{\lambda 0}$ are 148.5 and 6,660,925, or about 2 clicks per day over 78 days. This



(a) *(b)*

Figure 4. *Data and estimated hidden Markov process for the account in Figure 3. The top row is jittered event times. The middle and bottom rows plot the respective probabilities that N_1 and N_2 are active at time t. Panel (a) plots the full observation window. Panel (b) is a close-up of panel (a).*



(a) *(b)*

Figure 5. *(a) Sample paths for $a_{\lambda m}$ and $b_{\lambda m}$. (b) Posterior distribution of overall expected waiting time until births or deaths of sessions of different levels. Note the different time scales.*

means the "borrowed information" for λ_0 is 78 times the information in an individual account. Posterior shrinkage for λ_1 and λ_2 is much less, with prior exposure times of about 80 and 50 seconds, respectively.

Figure 5 also shows the posterior distribution of $b_{\beta m}/a_{\beta m}$ and $b_{\delta m}/a_{\delta m}$ for $m = 1, 2$, which are the prior expected waiting times until the birth or death of process m. The top row of Figure 5(b) indicates that the average time between Web sessions is slightly over five hours, while the average duration of a session is about fifteen minutes. The bottom row suggests that within a Web session, a burst of rapid clicking can be expected about every six minutes, and the burst is expected to last about a minute and a half. The inferences about the average session duration are interesting becasue little seems to be known about distributions of session lengths from past studies (e.g., Catledge and Pitkow 1995; Cockburn and MacKenzie 2001). These inferences are apparently the first to come from a model based method of determining Web session boundaries.

4. DISCUSSION

The MPC is a flexible model for point processes subject to irregular bursts of activity. This article has shown how to estimate MPC parameters using either a rapidly mixing MCMC sampler or an EM algorithm. The key to either approach is a set of forward-backward recursions for probabilistically restoring the latent Markov process. The same recursions allow for the filtering of future observations once MPC parameters have been estimated. The computational speed offered by the forward-backward recursions means that the MPC can be a viable model even in applications where MCMC is not feasible due to time constraints. The MPC is a well defined model, free of label switching issues. Finally, because the full conditional distributions underlying the MPC are all familiar models from the exponential family, it is straightforward to incorporate MPC parameters into a hierarchical model for a collection of users.

ACKNOWLEDGEMENTS

The work of P. Smyth was supported in part by the National Science Foundataion under grants IIS-9703120 and IIS-0083489, by a gift from Microsoft Research, and by an IBM Faculty Partnership award.

REFERENCES

Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Statist.* **41**, 164–171.

Catledge, L. D., and Pitkow, J. E. (1995). Characterizing browsing strategies in the World-Wide Web. *Computer Networks and ISDN Systems* **27**, 1065–1073.

Chib, S. (1996). Calculating posterior distributions and modal estimates in Markov mixture models. *J. Econometrics* **75**, 79–97.

Christiansen, C. L. and Morris, C. N. (1997). Hierarchical Poisson regression modeling. *J. Amer. Statist. Assoc.* **92**, 618–632.

Cockburn, A. and McKenzie, B. J. (2001). What do Web Users Do? An Empirical Analysis of Web Use. *International Journal of Human-Computer Studies* **56**, 903–922.

Cooley, R., Mobasher, B., and Srivastava, J. (1999). Data preparation for mining World Wide Web browsing patterns. *Knowledge and Information Systems* **1**, 5–32.

Cox, D. R. (1955). Some statistical methods connected with series of events. *J. Roy. Statist. Soc. B* **17**, 129–164, (with discussion).

Cunha, C., Bestavros, A., and Crovella, M. (1995). Characteristics of WWW Client-based Traces. *Tech. Rep.*, NEC, USA. Available from `citeseer.nj.nec.com/cunha95characteristics.html`

Davison, A. C. and Ramesh, N. I. (1996). Some models for discretized series of events. *J. Amer. Statist. Assoc.* **91**, 601–609.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. B* **39**, 1–22.

Du, Q. (1995). A monotonicity result for a single-server queue subject to a Markov-modulated Poisson process. *J. Appl. Probability* **32**, 1103–1111.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732.

Gutíerrez-Peña, E. and Nieto-Barajas, L. E. (2003). Bayesian Nonparametric Inference for Mixed Poisson Processes. *In this volume*.

Kolaczyk, E. D. (1999) Bayesian multi-scale models for Poisson processes. *J. Amer. Statist. Assoc.* **94**, 920–933.

Lambert, D., Pinheiro, J., and Sun, D. (2001). Estimating millions of dynamic timing patterns in real-time. *J. Amer. Statist. Assoc.* **96**, 316–330.

Olivier, C. and Walrand, J. (1994). On the existence of finite-dimensional filters for Markov-modulated traffic. *J. Appl. Probability* **31**, 515–525.

Robert, C. P. and Titterington, D. M. (1998). Reparameterization strategies for hidden Markov models and Bayesian approaches to maximum likelihood estimation. *Statist. Computing* **8**, 145–158.

Scott, S. L. (1998). *Bayesian Methods and Extensions to the Two State Markov Modulated Poisson Process*. Ph.D. Thesis, Harvard University, USA.

Scott, S. L. (1999). Bayesian analysis of a two state Markov modulated Poisson process. *J. Comp. Graph. Statist.* **8**, 662–670.

Scott, S. L. (2002). Bayesian methods for hidden Markov models: Recursive computing in the 21st century. *J. Amer. Statist. Assoc.* **97**, 337–351.

Tauscher, L. and Greenberg, S. (1997) Revisitation Patterns in World Wide Web Navigation. *Human Factors in Computing Systems: Proceedings of the CHI ’97 Conference*. New York: ACM, 399–406.

Turin, W. (1996). Fitting probabilistic automata via the EM algorithm. *Comm. Statist. Stochastic Models* **12**, 405–424.