Supplementary Material :
Priors over Recurrent Continuous Time Processes

Ardavan Saeedi **Alexandre Bouchard-Côté**
Department of Statistics
University of British Columbia

A Transience and recurrence

In this section, we describe the recurrent/transient behavior of samples from GEPs and DDPs (we assume the construction of [7] in this section, specialized to Dirichlet Process marginals for simplicity). Both processes can be viewed as time varying mixture models where a state θ_t at each observed time parameterizes a likelihood model L_{θ_t}. We assume that the observed times are the natural numbers, but this condition can be relaxed while preserving the results below.[1]

We assume in this section that Ω is countable, but the same theorem can be extended to Ω uncountable by using the HGEP.

We look at the 1-skeleton for the parameters sampled from these processes and used to sample the observations. Note that these chains (θ_n) are not Markovian, but the concepts of recurrence and transience are still applicable.

Proposition 8. *For the DDP construction of [7] specialized to Dirichlet Process marginals, the chain (θ_n) is transient.*

Proof. Let N_i denote the number of observations between the ordered Poisson subordinator events i and $i+1$, and $V_n \overset{\text{iid}}{\sim} \text{Beta}(1, \alpha_0)$. By construction, the claim is equivalent to:

$$\mathbb{P}\left(\sum_{n=1}^{\infty} N_n V_1 \prod_{j<n}(1-V_j) < \infty\right) = 1,$$

so it is enough to show that:

$$\mathbb{E}\left[\sum_{n=1}^{\infty} N_n V_1 \prod_{j<n}(1-V_j)\right] < \infty.$$

By using independence of all the V_j, N_n, and the fact that the expectations of non-degenerate beta distributions are in $(0,1)$, we have that the left hand side is equal to:

$$\sum_{n=1}^{\infty} \mathbb{E}[N_n]\,\mathbb{E}[V_1] \prod_{j<n} \mathbb{E}[(1-V_j)] = c\sum_{n=1}^{\infty} \rho^n$$
$$< \infty,$$

for some $c > 0, \rho \in (0,1)$, which proves the claim. □

Proposition 9. *For the GEP, the chain (θ_n) is recurrent.*

Proof. We first prove the jump process is recurrent; next, we prove the result for the 1-skeleton by showing, with probability one, that the waiting times are greater than one infinitely often.

[1] For example, by assuming the observations are separated by at least ϵ, for some fixed $\epsilon > 0$.

Consider the subsequence $\theta_{S(m)}$ such that $S(m)$ is the minimum integer with $\theta_{S(m-1)} \neq \theta_{S(m)}$. Note that $\mathbb{P}(S(m) < \infty) = 1$ as long as H_0 is not a single atom, in which case the claim holds trivially.

We will use the Chinese Restaurant Process (CRP) to represent this subsequence: let $U_1, U_2, \ldots$ be iid uniform, and let

$$E_m = \left(U_m < \frac{1}{m + \alpha_0} \right).$$

Given $\theta_{S(m)}$, the next distinct state $\theta_{S(m+1)}$ can be obtained from one iid draw from the U_i's and at most one draw from H_0. By the form of the CRP mixtures, we can set without loss of generality $\theta_{S(m+1)} = \theta_1$ whenever E_m holds. Since the E_m are independent and

$$\sum_{m=1}^{\infty} \mathbb{P}(E_m) = \infty,$$

by the second Borel-Cantelli lemma we have $\mathbb{P}(E_m \text{ i.o.}) = 1$.

This argument applies to the jump process; in order to prove the result for the 1-skeleton we need to show the waiting times at θ are greater than 1 infinitely often. To do this, we simply use Proposition 5, which implies that the sequence of jumps $J_{j(\theta,1)}, J_{j(\theta,2)}, \ldots$ introduced in Proposition 5 can equivalently be obtained by first sampling Γ from a gamma distribution, and then by conditionally independent sampling from an exponential with parameter Γ. Therefore, as long as $(\Gamma > 0)$, an event of probability one, waiting times greater than one will be sampled infinitely often. □

By Theorem 2.7.1 of [24], we get the following corollary directly (recall that if ζ is equal to the sum of waiting times $\zeta = J_1 + J_2 + J_3 + \ldots$, a process is *explosive* if there is a positive probability that ζ is finite):

Corollary 10. *GEPs are explosion-free.*

B Proofs

Proof of Proposition 6:

Proof. From Proposition 5, we have:

$$\begin{aligned} p(j_1, j_2, \ldots, j_K) &= \mathbf{1}[j_k > 0, k \in \{1, \ldots, K\}] \frac{\alpha_0 \beta_0^{\alpha_0}}{(\beta_0 + j_1)^{\alpha_0 + 1}} \frac{(\alpha_0 + 1)(\beta_0 + j_1)^{\alpha_0 + 1}}{(\beta_0 + j_1 + j_2)^{\alpha_0 + 2}} \\ &\quad \times \cdots \times \frac{(\alpha_0 + K - 1)(\beta_0 + j_1 + \cdots + j_{K-1})^{\alpha_0 + K - 1}}{(\beta_0 + j_1 + \cdots + j_K)^{\alpha_0 + K}} \\ &\propto \mathbf{1}[j_k > 0, k \in \{1, \ldots, K\}] (\beta_0 + j_1 + \cdots + j_K)^{-\alpha_0 - K}, \end{aligned}$$

and the normalization of this expression is indeed equal to $1/((\alpha_0)_K \beta_0^{\alpha_0})$. □

Proof of Proposition 7:

Proof. Conditioning on $\|\mu_0\|$, we have that θ_{N+1} and J_{N+1} are independent. Therefore $(\theta_{N+1} \big| X, \{A_n\}_{n=1}^N, \|\mu_0\|)$ can be viewed as a hierarchical Dirichlet process with concentrations $\|H_0\|$ for the top level DP, and $\|\mu_0\|$ for the lower level DPs. From [20], the distribution $\bar{\mu}'^{(\mathrm{H})}_{\theta_N}$ is then the predictive distribution for θ_{N+1}. For $J_{N+1} \big| X, \|\mu_0\|$, we get from Proposition 5 that the predictive is $\mathrm{TP}(\|\mu'^{(\mathrm{H})}_{\theta_N}\|, \beta'_{\theta_N})$. □

C Comparison to subordination of infinite HMMs

In addition to DDPs and GEPs, another way of constructing a non-parametric prior over continuous time processes is via Subordination of Infinite HMMs (SIHMM). In other words, SIHMM is obtained by first simulating a Poisson process with rate λ, and conditionally on the sampled locations,

simulating an infinite HMM. In this section, we show that this approach is not equivalent to the GEP prior, and give some of the advantages of GEPs over the infinite HMM subordination approach.

To show that SIHMMs and GEPs are different, it is enough to show that with probability one, a sample from a GEP cannot be uniformized (see [25] for background on uniformization). Informally, the additional flexibility of GEPs comes from each row i in the infinite rate matrix having a different total rate $-q_{i,i}$. Since there is an infinite number of rows, the maximum rate goes to infinity, making standard uniformization impossible. Other types of uniformization have been proposed (dynamic and adaptive uniformization), but they require truncations [26], while our model does not. More formally:

Proposition 11. *GEPs cannot be uniformized.*

Proof. Let Q denote a random rate matrix sampled from the GEP distribution, and $\lambda_i = -q_{i,i}$. Uniformization requires the simulation of a Poisson process with rate $\lambda < \infty$ satisfying $\lambda \geq \lambda_i$ for all i. Let

$$\begin{aligned} E_{N,n} &= \left(\max\{\lambda_i : 1 \leq i \leq n\} > N\right), \\ E_N &= \bigcup_n E_{N,n}, \\ E &= \bigcap_N E_N, \end{aligned}$$

and note that since the distribution of the $|q_{i,i}|$ has support on $(0, \infty)$, $\mathbb{P}(E^c_{N,n}) = (1 - \epsilon_N)^n$ for some $\epsilon_N > 0$.

It follows that for all N, $\mathbb{P}(E_N) = 1$, and hence that $P(E) = 1$, contradicting $\lambda < \infty$. □

D Chinese Restaurant Franchise (CRF) auxiliary variables

In this section, we review and formalize the table creation auxiliary variables A_n used in Section 4.

The idea is to view the predictive distribution $\bar{\mu}'^{(\mathrm{H})}_{\theta_N}$ for the next state θ_{N+1} in the hierarchical model as a mixture of two possibilities. The two possibilities are (1) to sample from the empirical distribution over the transitions starting at θ_N, $\bar{F}_{\theta_N}$ (this is called "joining one of the existing tables in the current restaurant (i.e. current state θ_N)" in the CRF analogy), or (2) to sample from a back-off distribution, $\bar{\mu}''$ ("creating a new table in the current restaurant"). One of these two events is selected with probability proportional to $(\|F_{\theta_N}\|, \|\mu_0\|)$. When alternative (1) is selected, the successor state θ_{N+1} is determined by $\bar{F}_{\theta_N}$ ("the new customer picks the dish of the selected existing table"), when alternative (2) is selected, the new table picks a dish. This is done recursively using the same process, except that the empirical distribution is now over the dishes picked by tables across all restaurants, G, and the back-off distribution becomes the normalized base measure, $\bar{H}_0$. Note that for the model of Section 4, indicators for the higher-level dish selection process need not be represented, but they are needed in higher hierarchies.

By augmenting the state-space of the sampler with an indicator over which of the two alternatives (1,2) is selected at each transition, the predictive distribution takes a tractable form.

Formally, the definition of the table creation auxiliary variables is therefore as follows:

$$\begin{aligned} \mathbb{P}(A_{N+1} = a | \|\mu_0\|, X) &\propto \|\mu_0\|^a \, \|F_{\theta_N}\|^{1-a} \, \mathbf{1}[a \in \{0, 1\}] \\ \theta_{N+1} \big| A_{N+1}, X &\sim (1 - A_{N+1}) \bar{F}_{\theta_N} + A_{N+1} \bar{\mu}''. \end{aligned}$$

E Resampling top-level normalization auxiliary variables in HGEPs

The next result shows that a Gibbs kernel can be used to resample the auxiliary variable $\|\mu_0\|$ used in HGEP posterior inference.

SMC algorithm to construct a proposal for the PMCMC sampler

G: the number of measurements in the current sequence (k)
g: particles generation $g \in \{0, 1, \dots, G\}$
M: number of particles
m: particle number $m \in \{1, \dots, M\}$
k: sequence number
$(F_\theta^{(\backslash k)}, T_\theta^{(\backslash k)})$: sufficient statistics from sequences other than k which are held fixed
$(F_{\theta,m,n}^{(k)}, T_{\theta,m,n}^{(k)})$: sufficient statistics from the mth particle of sequence k up to its nth event
$\mathrm{d}y$: an observation $\mathrm{d}y \in \mathcal{F}_\mathcal{X}$
S_θ: sufficient statistics for the likelihood model $\mathcal{P}$
$L(\mathrm{d}y|S_\theta)$: predictive likelihood given S_θ
(We omit writing $\forall m \in \{1, \dots, M\}$ to avoid excessive notation)

- Set $X_{m,0} = (\theta_{\text{beg}}, 0)$
For $g = 0$ to $g = G - 1$ do
  - Extend $X_{m,g}$ to a new particle $X'_{m,g+1}$:
    - Copy the events of $X_{m,g}$ into $X'_{m,g+1}$
    Loop over n until covering the first g measurements, $t_1^{(k)}, t_2^{(k)}, \dots, t_g^{(k)}$ (i.e. $t_{g+1}^{(k)} \leq \sum_{n=1}^{N_{m,g+1}} J_{m,n}$)
      - Sample $(\theta_{m,n+1}, J_{m,n+1})$ from Equation (2) given $\mu'_\theta = F_\theta^{(\backslash k)} + F_{\theta,m,n}^{(k)} + H_0$ and $\beta'_\theta = T_\theta^{(\backslash k)} + T_{\theta,m,n}^{(k)} + \beta_0$
      - Update, $S_{\theta,m,n}^{(k)}$, sufficient statistics for the likelihood model from the mth particle of sequence k up to its nth event
    End Loop
  - Compute the weight of the particles (we assume sufficient statistics are additive) $W_{m,g+1} = L(\mathrm{d}y|S_\theta^{(\backslash k)} + S_{\theta,m,n}^{(k)})$
  - Generate the new population of particles $X_{m,g+1}$ by resampling M times from $\bar{\pi}_{g+1}$. π_{g+1} is a weighted sum of Dirac delta functions, $\pi_{g+1} = \sum_{m=1}^{M} W_{m,g+1} \delta_{X'_{m,g+1}}$.
End For
- Sample $X_*^{(k)}$ from $\bar{\pi}_G$

Figure 4: Pseudocode for the SMC step of the PMCMC algorithm

Proposition 12. *The conditional distribution of $\|\mu_0\|$ given the other variables is a gamma distribution,*

$$\|\mu_0\| \Big| X, \{A_n\}_{n=1}^N \sim \mathrm{Gamma}(a, b),$$

with the following parameters: $a = \|H_0 + G\|, b = \gamma_0 + \sum_{\theta \in \Omega} \log(\beta'_\theta / \beta_0)$.

Proof. Using Equation (3) and standard CRP computations, we get that the conditional has density $p(x)$ proportional to:

$$\begin{aligned} p(x) &\propto \left(x^{\alpha_0 - 1} \exp(-\gamma_0 x)\right) \left(\prod_{\theta \in \Omega} \frac{(x)_{\|F_\theta\|} \beta_0^x}{(T_\theta + \beta_0)^{x + \|F_\theta\|}} \right) \times \left(x^{\|G\|} \prod_{\theta \in \Omega} \left((x)_{\|F_\theta\|}\right)^{-1} \right) \\ &\propto \left(x^{\alpha_0 + \|G\| - 1} \right) \exp\left(-x \left(\gamma_0 + \sum_{\theta \in \Omega} \log\left(\beta'_\theta / \beta_0\right) \right) \right) \end{aligned}$$

□

F Properties of translated Pareto distributions

In this section, we show two useful basic properties of the translated Pareto distribution of Section 3.

We start by showing how to sample from $\mathrm{TP}(\alpha, \beta)$ using the inverse cdf method:

Proposition 13. *Let $U \sim \mathrm{Unif}$, and define:*

$$T = \frac{\beta(1 - U^{1/\alpha})}{U^{1/\alpha}}, \tag{4}$$

then $T \sim \mathrm{TP}(\alpha, \beta)$.

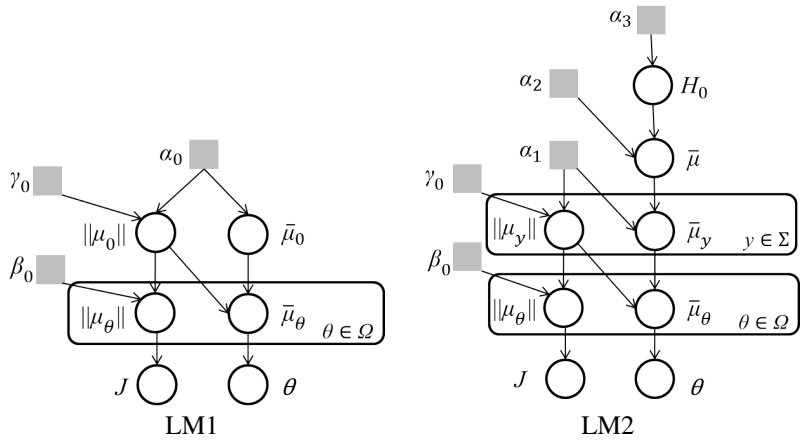


Figure 5: Comparison of two likelihood models. Note the slightly different convention for the α variables for LM2.

Proof. The cdf is given by:

$$F(T) = \int_0^T \frac{\alpha\beta^\alpha \, \mathrm{d}t}{(t+\beta)^{\alpha+1}}$$
$$= 1 - \left(\frac{\beta}{T+\beta}\right)^\alpha,$$

and solving $F(T) = 1 - U$ yields Equation (4). □

Next, we give an expression for the first moments:

Proposition 14. *Let $T \sim \mathrm{TP}(\alpha, \beta)$. We have:*

$$\mathbb{E}[T] = \begin{cases} \frac{\beta}{\alpha-1} & \text{if } \alpha > 1 \\ \infty & \text{o.w.} \end{cases}$$

G More information on the experiments

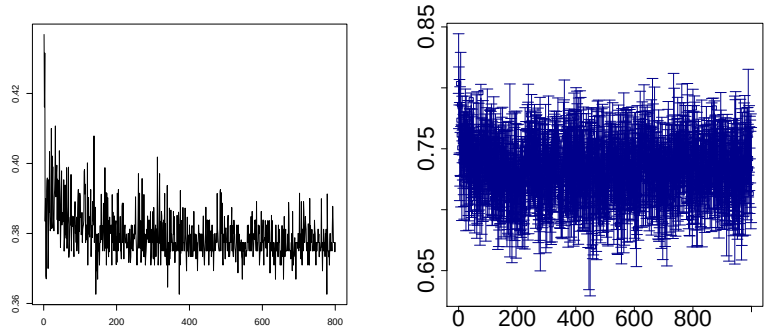
In this section, we give more information on the experiments of Section 6. We start by describing the likelihood models in more details.

G.1 Likelihood model

We tried two likelihood models: one Dirichlet-multinomial (LM1) model, and one multinomial-Dirac (LM2) model (see Figure 5).

LM1 uses a (finite) Dirichlet distribution for H_0, and a multinomial distribution for L_θ. More precisely, we assume that the observations y takes one of the K discrete values in a set Σ and given the parameter θ, follow a multinomial distribution with the K-dimensional parameter vector θ, where each entry is in $[0, 1]$. The base measure on the random variables θ has a Dirichlet prior; that is, $H_0 = \mathrm{Dir}(\alpha)$ where α is a K-dimensional positive parameter vector. Thus, the predictive likelihood is given by $L(\{y\}|S_\theta) = \frac{\alpha_y + S_{\theta,y}}{\sum_{i=1}^K (\alpha_i + S_{\theta,i})}$, in which S_θ is the sufficient statistic for state θ. In other words, $S_{\theta,k}$ is the empirical count of the number of times that we have observed category k when we were at state θ and created a new table.

LM2 uses a product of multinomial and uniform distributions for H_0, i.e. $H_0 = \mathrm{Mult} \times \mathrm{Unif}$, and a Dirac delta for L_θ. In LM2, if $\theta = (y, u)$ is known at a time step t, where y is a symbol in the observed alphabet Σ and $u \in [0, 1]$, the observation at that time step is a deterministic function of θ: $L_{(y,u)} = \delta_y$. The uniform distribution can be thought as being responsible for generating unique identifiers for hidden states (it could be replaced by any other non-atomic distribution without any



(a) (b)

Figure 6: (a) Error rate using LM1. (b) Mean acceptance rate per scan for a proposal using 100 particles as a function of MCMC iteration using LM2.

effect on the predictive distributions). The form of LM2 allows for adding useful structure in the hierarchy. Instead of having only one back-off model over all previous θ_n, we add an additional, intermediate back-off model over the previous hidden states that emitted the same observation $y \in \Sigma$. Concretely, this simply means adding a level μ_y in the hierarchy between μ_0 and μ_θ. In addition, we added another finite Dirichlet level to the hierarchy on top of H_0, to learn an overall frequency over the symbols $y \in \Sigma$. LM2 can therefore be thought as a Dirichlet-Multinomial-Dirac model. While LM2 may seem complex at first glance, by using a recursive implementation of the HGEP, this does not significantly increase the complexity of the code.

Figure 6 shows the mean error on the MS dataset of LM1 as a function of the number of Gibbs scans, averaged across 5 runs. In this experiment, we found that LM2 works significantly better than LM1, thus, the results justify the need for the more sophisticated approach for the likelihood model. The form of LM2 is also closer to the competing EM model, so we used LM2 for the experiments in Section 6.

G.2 Data

The RNA data [4] is publicly available at

`http://www.rna.ccbb.utexas.edu/DAT/3C/Alignment/Files/16S/16S.3.alnfasta.zip`

A tree was constructed on a random subset of 30 species using PhyML [27], and the nucleotides at speciation events were reconstructed using a K2P rate matrix and the sum-product algorithm on trees. We then considered the time series consisting of paths from one modern leaf to the root.

For the synthetic data, we first generated random parameters as follows: we used an Erdös-Rényi model with probability parameter $1/5$ to generate a random sparse matrix of size 10×10. The non-diagonal zeros in this matrix correspond to entries with $\mathrm{Unif}(0, 1/100)$ rate, and the non-diagonal ones in this matrix correspond to entries with $\mathrm{Unif}(0, 1/2 + 1/100)$ rate. The diagonal entries were filled with minus the value of the non-diagonal ones, and each row as set to deterministically emit one of the symbols in a finite alphabet Σ at random, $|\Sigma| = 4$.

The sampled data used in our experiments is available in the file data.txt in the supplementary material. For the MS data, we only include the time steps held-out (anon.txt), the data itself is confidential for anonymity and license reasons.

G.3 Miscellanea

We used 100 particles for the proposal distributions, and found that using 1000 particles did not change the results significantly, but using 10 particles degraded performance because most proposals

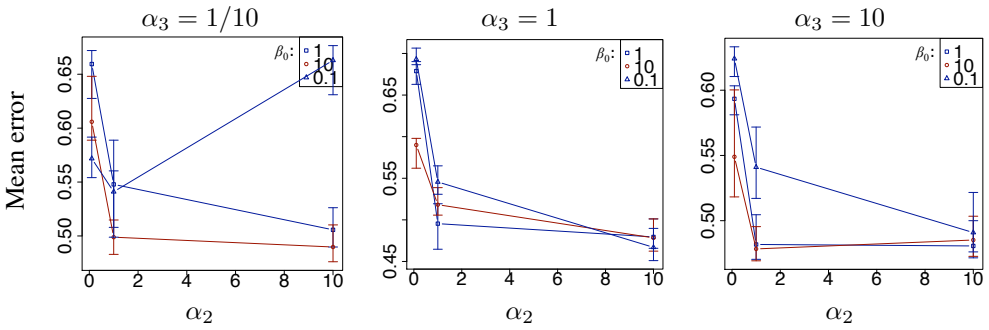


Figure 7: Effects of the hyper-parameters on mean error on synthetic data. Results are averaged over 5 runs with different random seeds.

were rejected in this regime. We show the mean acceptance rate per scan for 100 particles as a function of MCMC iteration in Figure 6.

To initialize the MCMC chain, we used a first scan where the moves are always accepted.

H Effect of hyper-parameters

In this section, we study the effects of the hyper-parameters. Qualitatively, the measure normalizations $\|\mu\|$, or concentration parameters (labeled α_i in the graphs in this section), have the same interpretation as the concentration parameters of HDPs. In addition, they control in conjunction with the rate parameters β the waiting times. From the results of Section F,the relation between the time scale (time between events) and the parameters should roughly looks like $\beta/(\alpha - 1)$ when $\alpha > 1$. When $\alpha \leq 1$, the mean of the predictive jump time distribution is infinite only when there was no waiting time observed from that state. Intuitively, this makes sense, as one might prefer a fat tail distribution for the waiting time when no apriori information is available.

Quantitatively, we observed that the hyper-parameters of LM2 did not have a large effect with the exception of small values for α_2 (see Figure 7). Note that we did not use the results from these experiments to tune the hyper-parameters for the experiments in Section 6, we kept default values of 1 for all hyper-parameters in these experiments.

References

[1] C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte Carlo methods. *Journal Of The Royal Statistical Society Series B*, 2010.

[2] J. Van Gael, Y. Saatci, Y. W. Teh, and Z. Ghahramani. Beam sampling for the infinite hidden Markov model. In *ICML*, 2008.

[3] M. Mandel. Estimating disease progression using panel data. *Biostatistics*, 2010.

[4] J.J. Cannone, S. Subramanian, M.N. Schnare, J.R. Collett, L.M. D'Souza, Y. Du, B. Feng, N. Lin, L.V. Madabusi, K.M. Muller, N. Pande, Z. Shang, N. Yu, and R.R. Gutell. The comparative RNA web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BioMed Central Bioinformatics*, 2002.

[5] S.N. MacEachern. Dependent nonparametric processes. In *Section on Bayesian Statistical Science, American Statistical Association*, 1999.

[6] J.E. Griffin. The Ornstein-Uhlenbeck Dirichlet process and other time-varying processes for Bayesian nonparametric inference. *Journal of Statistical Planning and Inference*, 2008.

[7] J.E. Griffin and M.F.J. Steel. Stick-breaking autoregressive processes. *Journal of Econometrics*, 2011.

[8] M. F. J. Steel. *The New Palgrave Dictionary of Economics*, chapter Bayesian time series analysis. Palgrave Macmillan, 2008.

[9] S. Heiler. A survey on nonparametric time series analysis. CoFE Discussion Paper 99-05, Center of Finance and Econometrics, University of Konstanz, 1999.

[10] M. J. Beal, Z. Ghahramani, and C. E. Rasmussen. The infinite hidden Markov model. In *Machine Learning*. MIT Press, 2002.

[11] E.B. Fox, E.B. Sudderth, M.I. Jordan, and A.S. Willsky. An hdp-hmm for systems with state persistence. In *Proceedings of the International Conference on Machine Learning*, 2008.

[12] J. Van Gael, Y. W. Teh, and Z. Ghahramani. The infinite factorial hidden Markov model. In *NIPS'08*, 2008.

[13] P.A.P. Moran. *The Theory of Storage*. Methuen, 1959.

[14] M. Friesl. Estimation in the Koziol-Green model using a gamma process prior. *Austrian Journal of Statistics*, 2008.

[15] V. Rao and Y. W. Teh. Spatial normalized gamma processes. In *Advances in Neural Information Processing Systems*, 2009.

[16] L. Kuo and S. K. Ghosh. Bayesian nonparametric inference for nonhomogeneous Poisson processes. Technical report, University of Connecticut, Department of Statistics, 1997.

[17] J. F. C. Kingman. *Poisson Processes*. The Clarendon Press Oxford University Press, 1993.

[18] M. Schroder. Risk-neutral parameter shifts and derivatives pricing in discrete time. *The Journal of Finance*, 2004.

[19] D. Dufresne. G distributions and the beta-gamma algebra. *Electronic Journal of Probability*, 2010.

[20] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 2004.

[21] R. Neal. Markov chain sampling methods for Dirichlet process mixture models. Technical report, U of T, 2000.

[22] P. Liang, S. Petrov, M. I. Jordan, and D. Klein. The infinite PCFG using hierarchical Dirichlet processes. In *Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP/CoNLL)*, 2007.

[23] A. Hobolth and J.L. Jensen. Statistical inference in evolutionary models of DNA sequences via the EM algorithm. *Statistical applications in Genetics and Molecular Biology*, 2005.

[24] J.R. Norris. *Markov chains*. Cambridge Series in Statistical and Probabilistic Mathematics, 1998.

[25] L. Mateiu and B. Rannala. Inferring complex DNA substitution processes on phylogenies using uniformization and data augmentation. *Syst. Biol.*, 2006.

[26] L. Zhang, H. Hermanns, E. M. Hahn, and B. Wachter. Time-bounded model checking of infinite-state continuous-time Markov chains. In *Int. Conf. on Application of Concurrency to System Design*, 2008.

[27] S. Guindon and O. Gascuel. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, 2004.