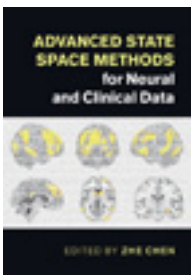


Advanced State Space Methods for Neural and Clinical Data

Edited by Zhe Chen

Book DOI: http://dx.doi.org/10.1017/CBO9781139941433

Online ISBN: 9781139941433

Hardback ISBN: 9781107079199

Chapter

Bayesian nonparametric learning of switching dynamics in cohort physio

logical time series: application in critical care patient monitoring p

p. 257-282

Chapter DOI: http://dx.doi.org/10.1017/CBO9781139941433.012

Cambridge University Press

11 Bayesian nonparametric learning of switching dynamics in cohort physiological time series: application in critical care patient monitoring

L. H. Lehman, M. J. Johnson, S. Nemati, R. P. Adams and R. G. Mark

11.1 Introduction

The time series of vital signs, such as heart rate (HR) and blood pressure (BP), can exhibit complex dynamic behaviors as a result of internally and externally induced changes in the state of the underlying control systems (Peng *et al.* 1995; Ivanov *et al.* 1999; Costa *et al.* 2002). For instance, time series of BP can exhibit oscillations on the order of seconds (e.g., due to the variations in sympathovagal balance), to minutes (e.g., as a consequence of fever, blood loss, or behavioral factors), to hours (e.g., due to humoral variations, sleep-wake cycle, or circadian effects) (Mancia 2012; Parati *et al.* 2013). A question of interest is whether "similar" dynamical patterns can be automatically identified across a heterogeneous patient cohort, and be used for prognosis of patients' health and progress.

In this work, we present a Bayesian nonparametric switching Markov processes framework with conditionally linear dynamics to learn phenotypic dynamic behaviors from vital sign time series of a patient cohort, and use the learned dynamics to characterize the changing physiological states of patients for critical-care bed-side monitoring (Lehman *et al.* 2012, 2013, 2014a; Nemati 2012). We assume that although the underlying dynamical system may be nonlinear and nonstationary and the stochastic noise components can be non-Gaussian, the dynamics can be approximated by a collection of linear dynamical systems (Nemati 2012; Nemati *et al.* 2012). Each such linear "dynamic" (or mode) is a time-dependent rule that describes how the future state of the system evolves from its current state, centered around a given system equilibrium point. Therefore, an ideal algorithm would be able to identify time series segments that follow a "similar" dynamic, and would switch to a different mode upon a change in the state of the underlying system.

We explore several variants of the Bayesian nonparametric approach to discovery of shared dynamics among patients via switching Markov processes: hierarchical Dirichlet process (HDP) autoregressive hidden Markov model (HDP-AR-HMM) (Teh *et al.* 2006; Fox *et al.* 2008), an explicit-duration HDP-based hidden semi-Markov model (HDP-AR-HSMM) (Johnson & Willsky 2013a), and the beta process autoregressive HMM (BP-AR-HMM) (Fox 2009; Fox *et al.* 2009, 2014). Given a collection of time

series from a cohort, these techniques allow for simultaneous learning of the underlying dynamic modes, and segmentation of the time series in terms of the most likely dynamic describing the time series evolution at any given point in time. The Bayesian nonparametric framework provides a mechanism to infer the number of dynamical modes from the data. Each such dynamical mode is possibly recurrent within the same time series and shared across multiple patients. The proposed framework allows for defining a notion of "similarity" among physiological time series based on their underlying shared dynamics. Therefore, one may consider two subjects to be similar if their underlying vital signs time series exhibit similar dynamics in response to external (e.g., tilting of body) or internal perturbations (e.g., onset of blood infection).

The rest of the chapter is organized as follows. We first present an overview of Markov switching processes and their Bayesian nonparametric variants. We review the sticky HDP-AR-HMM, HDP-AR-HSMM and BP-AR-HMM models, and describe the inference algorithms for these models. We validate the proposed techniques using HR and BP time series from a human laboratory study of subjects undergoing a tilt-table test, where the timing of the occurrence of the different dynamics and the sharing of the dynamics across multiple time series or subjects were known a priori. We present performance of these techniques in discriminating between two different postural positions in the tilt data set.

We test the prognostic value of the discovered vital sign dynamic behaviors. We apply a variant of the HDP-AR-HSMM approach to the HR and BP dynamics of an intensive care unit (ICU) cohort from the MIMIC II database (Saeed *et al.* 2011) during the first 24 hours of their ICU stays, and test whether cardiovascular dynamics during the first 24 hours of ICU admission are predictive of survival and mortality after adjusting for existing acuity scores.

11.2 Bayesian nonparametric switching Markov modeling of cohort time series

11.2.1 Overview of Bayesian nonparametric learning of switching Markov processes

Markov switching processes, such as the HMM, the switching vector autoregressive process (SVAR), and the switching linear dynamical system (SLDS), characterize complex dynamical phenomena as repeated returns to a set of simpler models (Fox 2009; Fox *et al.* 2010; Nemati 2012; Johnson 2014). In this chapter, we consider several Bayesian nonparametric variants of the SVAR to model physiological time series via Markov transitions among an unbounded collection of simpler linear dynamical systems. Two types of stochastic processes, the beta process and the HDP, are used as a priors on SVAR models to allow new states to be generated as more observations are made, thus allowing the data to drive the complexity of the learned model. In particular, our approach to discovery of shared dynamics among patients is based on variants of the HDP-AR-HMM, HDP-AR-HSMM (Teh *et al.* 2006; Johnson & Willsky 2013a; Johnson 2014; Fox *et al.* 2008; Fox 2009) and the BP-AR-HMM (Fox *et al.* 2009, 2014).

11.2.1.1 The AR-HMM

The AR-HMM, or switching vector AR (SVAR) process, models dynamics as switching among a set of simpler linear dynamical modes or behaviors. We assume that there exists a library of possible dynamic behaviors, with the k-th behavior parameterized by $\theta_k = \{A_k, \Sigma_k\}$ (i.e., a set of AR coefficients and the associated noise covariance).

Let $y_t^{(i)}$ represent the observation vector of the i-th time series at time t, and $z_t^{(i)}$ the state of the corresponding Markov chain at time t. Let π_k be the state-specific transition distribution for mode k. Due to the Markovian structure on the state sequence, $z_t^{(i)} \sim \pi_{z_{t-1}^{(i)}}$, for all $t > 1$. An order r switching VAR process, denoted by VAR(r), is defined as follows:

$$z_t^{(i)} \sim \pi_{z_{t-1}^{(i)}}, \tag{11.1}$$

$$y_t^{(i)} = \sum_{l=1}^{r} A_l^{z_t^{(i)}} y_{t-l}^{(i)} + e_t^{(i)}(z_t^{(i)}) \triangleq A_{z_t^{(i)}} \tilde{y}_t^{(i)} + e_t^{(i)}(z_t^{(i)}), \tag{11.2}$$

where mode-specific process noise $e_t^{(i)}(z_t^{(i)}) \sim \mathcal{N}(0, \Sigma_{(Z_t)})$, $A_k = [A_1^k ... A_r^k]$ define the set of lag matrices, and $\tilde{y}_t^{(i)} = [{y_{t-1}^{(i)}}^\top ... {y_{t-r}^{(i)}}^\top]^\top$.

11.2.1.2 The MNIW prior for the VAR dynamic parameters

The Matrix-Normal-Inverse-Wishart (MNIW) prior is the natural conjugate prior on the shared dynamic parameters $\theta_k = \{A_k, \Sigma_k\}$. We write

$$A_k, \Sigma_k \sim \texttt{MNIW}(S_0, \nu_0, M_0, K_0),$$

where $\lambda = (S_0, \nu_0, M_0, K_0)$ are hyperparameters. The MNIW prior consists of two parts: an inverse Wishart prior on Σ_k and a matrix normal prior on A_k (conditional on Σ_k):

$$\Sigma_k \mid S_0, \nu_0 \sim \texttt{InvWishart}(S_0, \nu_0), \tag{11.3}$$

$$A_k \mid \Sigma_k, M_0, K_0 \sim \texttt{MN}(A_k; M_0, \Sigma_k, K_0), \tag{11.4}$$

where ν_0 is the degrees of freedom, S_0 is the scale matrix, M_0 is the prior mean dynamic matrix for A_k, Σ_k describes the column covariance of A_k, and K_0 describes the row covariance of A_k. Thus, M_0 is the prior mean value for A_k (i.e., the expected value $\mathbb{E}[A_k] = M_0$), and Σ_k and K_0 together control the covariance of A_k.

11.2.1.3 The HDP prior for the HMM

HMMs are often used to model sequential or temporal data, where each time step is associated with a state, and observations are independent given the states. Classical approaches to the HMM assume a fixed, prespecified number of Markov modes. HDPs have been used as a prior to model HMMs with unbounded numbers of hidden states, and they provide a way in which the number of modes present in any finite amount of data can be inferred (Teh *et al.* 2006).

In the HDP-HMM, a stick-breaking distribution is first sampled to generate an infinite sequence of average transition probabilities β. For each state k, a sequence of outgoing transition probabilities π_k is generated by independently sampling a Dirichlet process

(DP) with β as its base probability distribution. With this hierarchical construction, the resulting set of transition distributions tend to favor transitioning to the same popular states. However, because the model includes an infinite number of states, new states can be visited as more observations are made and thus the model complexity can adapt to the complexity of the dataset.

We denote this hierarchical prior by writing

$$\beta \sim \texttt{GEM}(\gamma), \tag{11.5}$$

$$\pi_k \sim \texttt{DP}(\alpha, \beta), \tag{11.6}$$

where GEM denotes the stick-breaking process (Sethuraman 1994), and γ and α are concentration parameters that control the allocation of probability mass in β and the diversity among the π_k, respectively. For the HDP-HMM, this process generates the rows of the infinite transition matrix, where each state-specific transition distribution π_k is the row corresponding to state k. In the HDP-AR-HMM, each HMM state, or mode, is associated with a linear dynamical process in order to capture more complex temporal dependencies in the observed data sequence.

11.2.1.4 The sticky HDP-AR-HMM and HDP-AR-HSMM

Prior work has shown that HDP-AR-HMM inadequately captures temporal state persistence (Fox 2009). The sticky HDP-AR-HMM augments the HDP-AR-HMM with an extra parameter κ that biases the process towards self-transitions to capture temporal state persistence (Fox *et al.* 2008; Fox 2009). The sticky variant of HDP-AR-HMM model (see Figure 11.1) is defined as the generative process

$$\beta \sim \texttt{GEM}(\gamma) \tag{11.7}$$

$$\pi_k \sim \texttt{DP}\left(\alpha + \kappa, \frac{\alpha\beta + \kappa\delta_k}{\alpha + \kappa}\right) \tag{11.8}$$

$$z_t^{(i)} \sim \pi_{z_{t-1}^{(i)}} \tag{11.9}$$

$$y_t^{(i)} = \sum_{l=1}^{r} A_l^{z_t^{(i)}} y_{t-l}^{(i)} + e_t^{(i)}(z_t^{(i)}) \tag{11.10}$$

where $\alpha\beta + \kappa\delta_k$ indicates that an amount κ is added to the k-th component of $\alpha\beta$, increasing the expected probability of self-transition by an amount proportional to κ.

An alternative way to control self-transition probabilities is to use the explicit-duration HDP hidden semi-Markov model (HDP-HSMM) (Johnson & Willsky 2013a; Johnson 2014). The HDP-HSMM allows for explicit modeling of state-specific duration distributions and thus provides a more general approach to encouraging state persistence.The model augments a standard HMM with a random state duration; it is semi-Markov, as the transition to the next state depends not only on the current state, but also on how long the observations have stayed in that state. Although the duration distributions can be chosen to take any form, here we choose geometric duration distributions so as to create a model similar to the sticky HDP-AR-HMM. Specifically, we model observations via conditionally linear dynamics as in HDP-AR-HMM, but

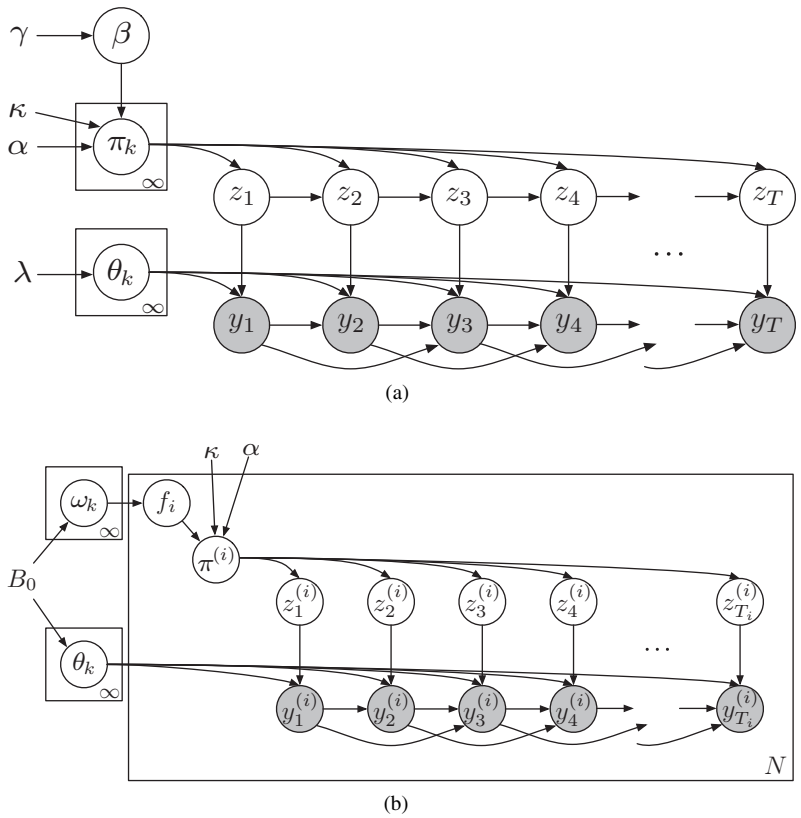


Figure 11.1 Graphical models for (a) sticky HDP-AR-HMM and (b) BP-AR-HMM.

explicitly models the state duration with geometric state duration distributions. The model emulates a standard HMM (since the duration has a geometric distribution), while providing a Bayesian prior to model self-transition probability that directly controls the duration distribution. We place a $\texttt{Beta}(\alpha_0, \beta_0)$ prior on each state's geometric duration parameter (i.e., self-transition probability) and thus control state persistence in a more interpretable way.

11.2.1.5 The BP-AR-HMM

In the BP-AR-HMM, the beta process implements a feature-based Bayesian nonparametric approach to discover the dynamic behaviors shared by a collection of related time series (Fox 2009; Fox *et al.* 2009, 2014). It explicitly models the dynamic behaviors using feature vectors and determine both *shared* and *unique* dynamic behaviors in a collection of time series. Briefly, the BP-AR-HMM technique assumes that there exists a library of possible dynamic behaviors, where the k-th behavior is parameterized by $\theta_k = \{A_k, \Sigma_k\}$ (i.e., a set of AR coefficients and the associated noise covariance), and each time series i can take on a subset of these behaviors. A binary feature vector

$f_i = [f_{i1}, f_{i2}, ...]$ is used to denote the subset of behaviors exhibited by time series i, where $f_{ik} = 1$ indicates that time series i exhibits the k-th behavior.

The beta process has been used to construct latent feature models with an unknown number of latent features (Ghahramani *et al.* 2006). Draws from beta processes are random discrete measures; each atom corresponds to a feature, with the mass corresponding to the probability that the feature is present for an object (Ghahramani *et al.* 2006; Fox *et al.* 2009). Let B be a random measure drawn from the beta process (parameterized by a base measure B_0), then $B \mid B_0 \sim \texttt{BP}(1, B_0)$ is defined by its masses ω_k, and locations θ_k (see Figure 11.1). The BP-AR-HMM model (Fox 2009; Fox *et al.* 2009, 2014) is defined as follows:

$$B \sim \texttt{BP}(1, B_0), \tag{11.11}$$

$$X_i \sim \texttt{BeP}(B),\ i = 1, \ldots, N, \tag{11.12}$$

$$\pi_k^{(i)} \mid f_i \sim \texttt{Dir}([\alpha, \ldots, \alpha + \delta(k,j)\kappa, \ldots, \alpha] \odot f_i), \tag{11.13}$$

$$z_t^{(i)} \sim \pi_{z_{t-1}^i}^{(i)}, \tag{11.14}$$

$$y_t^{(i)} = \sum_{l=1}^{r} A_l^{z_t^i} y_{t-l}^{(i)} + e_t^{(i)}(z_t^i), \tag{11.15}$$

where X_i is a Bernoulli process (denoted by $\texttt{BeP}(B)$) realization that determines the subset of features allocated to the i'th time series. $\pi_k^{(i)}$ represents the time-series-specific, feature-constrained distribution from dynamic mode k for time series i, restricting the time series i to transition among behaviors available in its feature vector f_i. A Dirichlet prior, parameterized by α and κ, is placed on the Markovian state switching probabilities $\pi_k^{(i)}$. Conditioned on f_i, $\pi_k^{(i)} \sim \texttt{Dir}([\ldots \alpha + \delta(k,j)\kappa, \ldots] \odot f_i)$, where $\texttt{Dir}(\cdot)$ denotes the Dirichlet distribution, $\odot$ denotes the element-wise vector product, and the Kronecker delta function defined by $\delta(k,j)$ is 0 when $k \neq j$ and 1 otherwise.

11.2.2 Inference algorithms and implementations

The goal of inference is to represent the posterior $p(\theta, \pi, z_{1:T} \mid y_{1:T})$, which is intractable to compute exactly. In particular, we need a representation that enables us to estimate posterior expectations of the form

$$\mathbb{E}\left[f(\theta, \pi, z_{1:T}, y_{1:T}) \mid y_{1:T}\right] = \int \sum_{z_{1:T}} f(\theta, \pi, z_{1:T}, y_{1:T}) p(\theta, \pi, z_{1:T} \mid y_{1:T}) d\theta d\pi, \tag{11.16}$$

for some function f. In this chapter, we use Markov chain Monte Carlo (MCMC) algorithms for approximate inference.

In MCMC inference, we construct an iterative process that generates samples from the posterior asymptotically in the sense that the process forms an ergodic Markov chain on the state space $(\theta, \pi, z_{1:T})$ and the stationary distribution of that Markov chain is the posterior distribution $p(\theta, \pi, z_{1:T} \mid y_{1:T})$. By running the chain and collecting a set of

samples $\mathcal{S}$ from the chain's trajectory, we can approximate the posterior expectation of f by

$$\mathbb{E}\left[f(\theta, \pi, z_{1:T}, y_{1:T}) | \, y_{1:T}\right] \approx \frac{1}{|\mathcal{S}|} \sum_{(\hat{\theta}, \hat{\pi}, \hat{z}_{1:T}) \in \mathcal{S}} f(\hat{\theta}, \hat{\pi}, \hat{z}_{1:T}, y_{1:T}). \tag{11.17}$$

The approximation error, or sample variance, in equation (11.17) is determined by the number of samples collected (i.e., for how many iterations the MCMC algorithm is run) and by the mixing rate of the Markov chain. However, the mixing rate of the chain is typically unknown and difficult to estimate. See Robert & Casella (2004) and Gelman *et al.* (2013) for more information on approximate inference using MCMC, and Bishop (2006), Wainwright & Jordan (2008) and Murphy (2012) for more information on approximate inference and alternative algorithms.

For the remainder of this section we describe MCMC algorithms for the models discussed in this chapter. These algorithms share the same general structure across all of the models: they alternate between block resampling of the hidden state sequences and resampling the model parameters. The hidden state sequences are resampled using message passing and the parameters are resampled using conjugate priors. First, we detail these steps for a Gibbs sampler for the sticky HDP-AR-HMM model. Next, we overview the differences in the MCMC sampling algorithms for the HDP-AR-HSMM and the BP-AR-HMM. Finally, we discuss the computational complexity and scalability of these algorithms.

11.2.2.1 A Gibbs sampler for the sticky HDP-AR-HMM

Algorithms for approximate inference in Bayesian nonparametric models based on the HDP need to handle draws from the HDP, which are infinite and cannot be instantiated completely (Teh *et al.* 2006; Fox *et al.* 2008; Van Gael *et al.* 2008). One approach common in MCMC algorithms is to use the convenient properties of the HDP to marginalize the HDP draws, eliminating π and β from the sampler's state (Teh *et al.* 2006; Fox *et al.* 2008). However, in the HDP-AR-HMM, once π and β are collapsed the state sequence $z_{1:T}$ loses its Markov chain structure and is Markov only with respect to the complete graph. As a result, in collapsed samplers each z_t must be drawn sequentially, resulting in very slow mixing times because the states are likely to be highly correlated. Thus it is preferable to instantiate some finite approximation to β and π so that the Markov structure among the hidden states can be exploited by dynamic programming algorithms.

Another approach to inference in HDP models is known as the weak limit approximation (Fox *et al.* 2008; Fox 2009; Johnson & Willsky 2013a; Johnson 2014), in which the infinite model is approximated by a finite one. That is, choosing some finite approximation parameter K we model β and π using finite Dirichlet distributions of size K

$$\beta \sim \texttt{Dir}(\gamma/K, \ldots, \gamma/K), \tag{11.18}$$

$$\pi_k \sim \texttt{Dir}(\alpha\beta_1, \ldots, \alpha\beta_j + \kappa\delta_{kj}, \ldots, \alpha\beta_K). \tag{11.19}$$

In this setting, K is an algorithm parameter rather than a model parameter, and one can achieve any desired approximation quality for large enough values of K

(Ishwaran & Zarepour 2002). Furthermore, this finite representation of the transition matrix allows the state sequence $z_{1:T}$ to be resampled as a block. Thus the weak limit approximation provides a natural way to trade off computational efficiency and posterior approximation, which naturally corresponds to trading off the variance error and bias error in the finite-sample estimate equation (11.17).

Using a weak limit approximation, we can construct a Gibbs sampler for the HDP-AR-HMM that cycles through updating components of the model in turn. That is, we iterate the process of resampling

$$z_{1:T} \mid \pi, \theta, y_{1:T}, \qquad \theta \mid z_{1:T}, y_{1:T} \qquad \text{and} \qquad \beta, \pi \mid z_{1:T}. \tag{11.20}$$

For simplicity, throughout this section we suppress notation for conditioning on hyperparameters and the superscript notation for multiple observation sequences.

11.2.2.2 Sampling $z_{1:T} \mid \pi, \theta, y_{1:T}$

Given the observation parameters θ and the transition parameters π, the hidden state sequence $z_{1:T}$ is Markov with respect to a chain graph. Therefore even though the density $p(z_{1:T} \mid \pi, \theta, y_{1:T})$ is supported on K^T values, it can be sampled in time $\mathcal{O}(TN^2)$ with dynamic programming. In particular, we exploit HMM message passing algorithms to marginalize in one direction along the chain and then sample in the other.

The standard HMM backward message passing recursions are

$$B_t(k) = p(y_{t+1:T} \mid \theta, \pi, z_t = k) \tag{11.21}$$

$$= \sum_{j=1}^{K} p(z_{t+1} = j \mid z_t = k, \pi) p(y_{t+1} \mid z_{t+1} = j, \theta) B_{t+1}(j), \tag{11.22}$$

for $t = 1, 2, \ldots, T-1$ and $k = 1, 2, \ldots, K$, where $B_T(k) = 1$ and where $y_{t+1:T} = (y_{t+1}, y_{t+2}, \ldots, y_T)$. Using these messages, we can write the conditional distribution of the first state z_1, marginalizing over all the future states $z_{2:T}$, as

$$p(z_1 = k \mid \pi, \theta, y_{1:T}) \propto p(z_1 = k \mid \pi) p(y_1 \mid z_1 = k, \theta) B_1(k), \tag{11.23}$$

which can be sampled efficiently. Given a sampled value $\bar{z}_1$, we can write the conditional distribution of the second state z_2 as

$$p(z_2 = k \mid \pi, \theta, y_{1:T}, z_1 = \bar{z}_1) \propto p(z_2 = k \mid z_1 = \bar{z}_1, \pi) p(y_2 \mid z_2 = k, \theta) B_2(k). \tag{11.24}$$

Therefore after passing HMM messages backward we can recursively sample forwards to construct a joint sample of the entire state sequence.

11.2.2.3 Sampling $\theta \mid z_{1:T}, y_{1:T}$

To resample the observation parameters θ conditioned on a fixed sample of the state sequence $z_{1:T}$ and the observations $y_{1:T}$, i.e., to sample from $p(\theta \mid z_{1:T}, y_{1:T}) \propto p(y \mid \theta, \tilde{y}) p(\theta \mid \lambda)$, we exploit conjugacy (Bernardo & Smith 2009) between the AR likelihood $p(y \mid \tilde{y}, \theta)$ and the MNIW prior $p(\theta \mid \lambda)$. Recall that $\theta = \{(A_k, \Sigma_k)\}$. Expanding the MNIW parameters as $\lambda = (S_0, \nu_0, M_0, K_0)$, we write the prior on the

k-th observation parameter as $p(A_k, \Sigma_k \mid S_0, \nu_0, M_0, K_0)$. Because of conjugacy between the prior and the likelihood, the posterior also follows the MNIW distribution, we have

$$p(A_k, \Sigma_k \mid z_{1:T}, y_{1:T}, S_0, \nu_0, M_0, K_0) = p(A_k, \Sigma_k \mid S_n, \nu_n, M_n, K_n), \tag{11.25}$$

where (S_n, ν_n, M_n, K_n) are posterior hyperparameters that are functions of the elements of $y_{1:T}$ assigned to state k as well as the preceding lagged observations:

$$S_n = S_0 + S_{yy^\top} + (M_0 K_0^{-1} M_0^\top - M_n K_n^{-1} M_n^\top), \tag{11.26}$$

$$M_n = (M_0 K_0^{-1} + S_{y\tilde{y}^\top}) K_n, \tag{11.27}$$

$$K_n = (K_0^{-1} + S_{\tilde{y}\tilde{y}^\top})^{-1}, \tag{11.28}$$

$$\nu_n = \nu_0 + n, \tag{11.29}$$

where

$$S_{yy^\top} = \sum_{t:z_t=k} y_t y_t^\top \qquad S_{\tilde{y}\tilde{y}^\top} = \sum_{t:z_t=k} \tilde{y}_t \tilde{y}_t^\top, \tag{11.30}$$

$$S_{y\tilde{y}^\top} = \sum_{t:z_t=k} y_t \tilde{y}_t^\top \qquad n = \#\{t : z_t = k\}. \tag{11.31}$$

That is, the posterior distribution has the same form as the prior distribution but with new hyperparameters that include statistics of the data.

Therefore resampling $\theta \mid z_{1:T}, y_{1:T}$ includes three steps: collecting statistics from the data assigned to each state, forming each state's posterior hyperparameters, and updating each state's observation parameter by simulating a draw from the appropriate MNIW. To simulate $(A, \Sigma) \sim \texttt{MNIW}(S_n, \nu_n, M_n, K_n)$ we sample

$$\Sigma \sim \texttt{InvWishart}(S_n, \nu_n), \tag{11.32}$$

$$A = M_n + \Sigma^{\frac{1}{2}} G K_n^{-\frac{1}{2}} \quad \text{where} \quad G_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0, 1). \tag{11.33}$$

11.2.2.4 Sampling $\beta, \pi \mid z_{1:T}$

To resample the transition parameters β and π, which are draws from the weak limit approximation to the (sticky) HDP, we employ an auxiliary variable sampling scheme (Teh *et al.* 2006; Fox *et al.* 2008). This auxiliary variable scheme simplifies the update to β.

We resample $\beta, \pi \mid z_{1:T}$ by first sampling auxiliary variables $m \mid \beta, z_{1:T}$. We then sample $\beta, \pi \mid z_{1:T}, m$ by first sampling from the marginal $\beta \mid m$ and then the conditional $\pi \mid \beta, z_{1:T}$. The dependence among the variables with the introduction of the auxiliary variables m is shown in Figure 11.2.

We write the transition counts in the sampled state sequence $z_{1:T}$ as

$$n_{kj} = \#\{t : z_t = k,\ z_{t+1} = j,\ t = 1, 2, \ldots, T\}. \tag{11.34}$$

Suppressing conditioning notation for simplicity, the auxiliary variables $m = \{m_{kj} : k, j = 1, 2, \ldots, K\}$ are sampled via

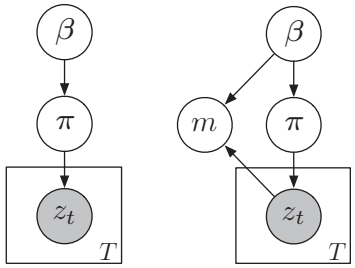


Figure 11.2 A simplified graphical model to show the auxiliary variables m used in sampling $\beta, \pi \mid z_{1:T}$.

$$m_{kj} = \sum_{l=1}^{n_{kj}} b_{kjl} \quad \text{where} \quad b_{kjl} \overset{\text{iid}}{\sim} \texttt{Bern}\left(\frac{\alpha\beta_j}{\alpha\beta_j + \kappa} \frac{\alpha\beta_j + \kappa\delta_{kj}}{\alpha\beta_j + l + \kappa\delta_{kj}}\right), \tag{11.35}$$

where $\texttt{Bern}(p)$ denotes a Bernoulli random variable that takes value 1 with probability p and takes value 0 otherwise. Note that the update for the HDP-HMM without a sticky bias corresponds to setting $\kappa = 0$ in these updates. See Teh *et al.* (2006) and Fox (2009) for details.

Given the auxiliary variables, the update to β is a Dirichlet-multinomial conjugate one, where

$$\beta \mid m \sim \texttt{Dir}(\gamma/K + m_{\cdot 1}, \gamma/K + m_{\cdot 2}, \ldots, \gamma/K + m_{\cdot K}), \tag{11.36}$$

where $m_{\cdot j} = \sum_{k=1}^{K} m_{kj}$ for $j = 1, 2, \ldots, K$. The update to $\pi \mid \beta, z_{1:T}$ is similar, with

$$\pi_k \mid \beta, z_{1:T} \sim \texttt{Dir}(\alpha\beta_1 + n_{k1}, \ldots, \alpha\beta_j + n_{kj} + \kappa\delta_{kj}, \ldots, \alpha\beta_K + n_{kK}). \tag{11.37}$$

It is also possible to treat the concentration parameters α and γ as well as the sticky bias parameter κ as random variables which are included in the sampling inference. For details on resampling α, γ, and κ, see Teh *et al.* (2006) and Fox (2009).

11.2.2.5 MCMC for the HDP-AR-HSMM and BP-AR-HMM

MCMC algorithms for the HDP-AR-HSMM and BP-AR-HMM based on the weak limit approximation are broadly similar to that of the HDP-AR-HMM and so we do not detail them here. Instead, we highlight some key differences and provide references.

The HDP-HSMM admits a similar Gibbs sampling strategy to that of the (sticky) HDP-HMM (Johnson & Willsky 2013b). Because HDP-HSMM allows arbitrary state-specific duration distributions, in its corresponding Gibbs sampling algorithm the parameters of those duration distributions are resampled in a step analogous to that for the observation parameters. One can choose duration models so that the update is efficient, including duration distributions with conjugate priors. One significant difference from the (sticky) HDP-HMM Gibbs sampler is that resampling the state sequence is much more expensive for general duration distributions. However, with

geometric duration distributions there is essentially no additional computational cost to the HDP-HSMM Gibbs sampling algorithm.

In the BP-AR-HMM, because the dynamical modes are shared as features among a set of observation sequences according to a beta process, the MCMC updates are more complex than those of the Gibbs samplers for the (sticky) HDP-HMM and HDP-HSMM. In particular, the algorithm must sample over which modes are expressed in which sequences. One method (Fox *et al.* 2014) uses a Metropolis birth–death proposal over the modes used in each observation sequence.

11.2.2.6 Computational complexity and scalability

The computational complexity of inference with MCMC algorithms ultimately depends on the mixing rate of the constructed Markov chain, and this mixing rate is difficult to analyze or estimate for complex models such as those considered in this chapter. However, while it is hard to evaluate how many iterations of the sampler are necessary for good performance, it is straightforward to analyze the computational costs of each iteration of the sampler. Here we briefly review how the computational complexity of the sampler iterations scale with the length of an observation sequence T and the size of the weak limit approximation K.

The computational cost of an iteration of these samplers is typically dominated by the cost of block resampling the state sequence; that is, of sampling $z_{1:T} \mid \theta, \pi, y_{1:T}$. The cost of computing the HMM backward messages scales as $\mathcal{O}(TK^2)$, since the algorithm computes a matrix-vector product requiring K^2 basic operations for each of the T time indices in a sequence. Given the backward messages, the cost of forward sampling scales as $\mathcal{O}(TK)$, since at each of the T time indices a finite distribution with support of size K must be normalized and sampled. The quadratic dependence on K is important when selecting a weak limit approximation level, since an approximation with twice as many states may require as much as four times the computation time per iteration.

It is common to fit such models to many observation sequences. With N observation sequences of average length T, the computational cost of each sampler iteration typically requires $\mathcal{O}(NTK^2)$ basic operations. Since the observation sequences are conditionally independent given the model parameters, it is straightforward to parallelize such computations. However, in these MCMC algorithms it is still necessary to "touch" the full dataset at each iteration, a requirement that can make such algorithms difficult to scale to very large datasets. It is a subject of ongoing research to scale MCMC (Angelino *et al.* 2014; Bardenet *et al.* 2014; Korattikara *et al.* 2014; Maclaurin & Adams 2014; Nishihara *et al.* 2014) and other Bayesian inference algorithms (Hoffman *et al.* 2013; Johnson & Willsky 2014) to very large datasets.

11.3 Materials and methods

This section describes the utilized datasets, as well as the parameter settings of the Bayesian nonparametric techniques for discovery of shared dynamics among patients.

11.3.1 Data sets

11.3.1.1 Tilt-table experiment

Time series of HR and MAP were acquired from ten healthy subjects undergoing a tilt-table experiment. The mean age was 28.7 ± 1.2 years. The details of the protocol are described in previous publications (Heldt *et al.* 2003; Heldt 2004). Briefly, subjects were placed in a supine position. Tilting was performed from horizontal position to vertical position and back to supine. Since we were interested in the dynamics of interaction between HR and MAP in the frequency range pertinent to sympathetic and parasympathetic regulation, time series of HR and MAP were high-pass filtered to remove the steady-state baseline and any oscillation in the time series slower than 100 beats/cycle. This filtering was done using a seventh order Butterworth digital filter with cutoff frequency of 0.01 cycles/beat. All time series were further normalized to have a standard deviation of one.

11.3.1.2 MIMIC II data set

The MIMIC II waveform database (version 2) (Saeed *et al.* 2011) included approximately 4000 sets of high resolution physiological waveforms with associated minute-by-minute vital sign trends. This study included only the adult patients with clinical information, and with at least 8 hours of continuous minute-by-minute invasive BP trends during the first 24 hours of their ICU stays. Patients with more than 15% of missing or invalid (i.e., outside physiologically plausible bounds of 20 to 200 mmHg for mean pressures) BP samples were excluded. In order to compare with SAPS-I score, we restricted our analysis to patients with SAPS-I scores during the first 24 hours of their ICU stays, yielding a total of 453 patients. The median SAPS-I score for this cohort is 16 (interquartile range [13, 18]). 16% of patients in this cohort died before hospital discharge. The data set contained approximately 9700 hours of minute-by-minute mean arterial blood pressure measurements (20.2 hours per patient on average). Gaussian noise was used to fill in the missing or invalid values. The median age of this cohort was 69 with an inter-quartile range of (57, 79). About 59% of the patients were male. Approximately 15% (67 out of 453) of patients in this cohort died in the hospital; 28-day mortality of this cohort was approximately 19% (85 out of 453).

11.3.2 Bayesian nonparametric model settings

For the tilt data sets, we modeled the beat-by-beat HR/BP time series as a switching AR(5) process to model most of the parasympathetic responses and at least some of the sympathetic effects, without introducing an unduly complex model. Minute-by-minute BP time series from MIMIC II were modeled as a switching AR(3) process to capture a real oscillation and a possible trend per mode.

11.3.2.1 MNIW prior settings

An inverse-Wishart prior $\texttt{InvWishart}(S_0, n_0)$ was placed on Σ_k, and a matrix-normal prior $\texttt{MN}(A_k; M_0, \Sigma_k, K_0)$ on A_k, given Σ_k. The MNIW prior was given $M_0 = 0$, the

matrix normal hyperparameter $K_0 = 10 \times \mathbf{I}_d$, $\nu_0 = d + 2$, where d was the dimension of the observations. The mean covariance matrix was set from data. For the tilt data set, the scale matrix S_0 was set to 0.5 times the empirical covariance of the observations. For the MIMIC II blood pressure data, the scale matrix S_0 was set to the empirical covariance of the observations to allow for more variability in the observed behaviors. For the MIMIC II heart rate data, we set the scale matrix S_0 to 0.75 times the empirical covariance of the observations.

11.3.2.2 Sticky HDP-AR-HMM settings

We sampled over hyperparameters to infer the number of states and degree of self-transition bias from the data. The sampling updates for the hyperparameters of the sticky HDP-AR-HMM were described in (Fox 2009). To simplify the inference procedure, we introduced an additional hyperparameter ρ as in (Fox 2009):

$$\alpha = (1 - \rho)(\alpha + \kappa), \tag{11.38}$$

$$\kappa = \rho(\alpha + \kappa). \tag{11.39}$$

Instead of sampling over α and κ directly, we sampled over $\alpha + \kappa$ and ρ using a gamma and beta distribution respectively.

For the tilt data set, we used a $\texttt{Gamma}(1, 1)$ hyperprior on γ, which specifies the concentration parameter for the base distribution β. We used a $\texttt{Beta}(100, 1)$ hyperprior on ρ, and $\texttt{Gamma}(1, 1)$ hyperprior on $\alpha + \kappa$. We report the classification performance using the model output after 10 000 MCMC iterations.

11.3.2.3 HDP-AR-HSMM settings

We model the state duration as a geometric distribution parameterized by a success (transition) probability of p_z. We place a $\texttt{Beta}(\alpha_0, \beta_0)$ prior on the transition probability p_z. Thus, the ratio β_0/α_0 controls the average duration in state Z. For the tilt data set, the inverse mean parameter of the geometric duration distribution was given a beta prior ($\alpha_0 = 1$, $\beta_0 = 300$). We report the classification performance using the model output after 10 000 MCMC iterations. For the MIMIC II data set, the inverse mean parameter of the geometric duration distribution was given a beta prior ($\alpha_0 = 1$, $\beta_0 = 60$). We report the performance using the model output after 3000 MCMC iterations. The HDP priors γ and α were set to 1. The MNIW prior setting was specified as before.

11.3.2.4 BP-AR-HMM settings

For the tilt data set, the hyperparameter α, which specifies the distribution over the total number of modes, was given a $\texttt{Gamma}(1, 1)$ prior. We used a $\texttt{Gamma}(1, 1)$ hyperprior on γ, which specifies the concentration parameter for the symmetric Dirichlet prior on each time series' Markov switching dynamics. The κ hyperparameter determines the preference for self-transition and it was given a $\texttt{Gamma}(100, 1)$ prior. We report the classification performance using the model output after 10 000 MCMC iterations.

11.3.3 Evaluation methods and statistical analysis

We define *mode proportion* $MP_k^{(i)}$ as the proportion of time the i-th patient spends within the k-th mode. Given the latent mode assignment z_t, we have

$$MP_k^{(i)} = \frac{1}{T^{(i)}} \sum_{t=1}^{T^{(i)}} \delta(z_t^{(i)}, k), \tag{11.40}$$

where $\delta(z_t^{(i)}, k)$ indicates the Kronecker delta and is one if $(z_t^{(i)} = k)$, and zero otherwise. For classification and prediction purposes, we characterize each time series with its corresponding mode proportion (a $1 \times K$ feature-vector), and use a logistic regression classifier to make predictions about the outcome variables of interest.

11.3.3.1 Time series classification and patient risk stratification

For the tilt-table experiment, we used the mode proportions within each segment (e.g., supine vs. non-supine) as inputs to a logistic regression classifier, and report the classification performance in discriminating between two different postural positions (supine vs. non-supine) in the tilt data set.

To assess the predictive power of the dynamic modes, we performed a ten-fold cross-validation study. Ten models were learned on the training set of each of the folds, followed by mapping the corresponding mode proportions to outcomes (e.g., hospital mortality) using logistic regression. Next, mode assignments of time series in the test set of each fold was inferred based on the modes learned from the corresponding training set, and the regression weights from the training fold was used to predict outcomes. We compare the mortality prediction performance of the dynamic modes with the existing acuity metrics.

11.3.3.2 MIMIC association analysis

We used univariate and multivariate logistic regressions to examine the associations between dynamic mode proportions and hospital mortality. We built a separate multivariate logistic regression model for each of the discovered dynamic modes, with the mode proportion as the primary predictive variable, and APACHE-IV as a covariate. For each mode, we reported its p value, odds ratio (OR, with 95% confidence interval), and adjusted OR (after including the covariate). The Hosmer–Lemeshow p values (HL p values) were reported to assess the model fit. The odds ratios were per 10% increase in the mode proportion. Two-sided p values less than 0.05 were considered statistically significant. The analysis was performed to quantify the mortality risk associated with each dynamic mode; modes with significant ($p < 0.05$) associations with mortality were established as either *low-risk* (OR < 1), or *high-risk* (OR > 1) dynamics depending on their odds ratios. Dynamic modes without statistically significant associations with mortality were *neutral* modes. Test of statistical significance was based on p-values after correcting for the false discovery rate (FDR) using the technique described in (Benjamini & Hochberg 1995).

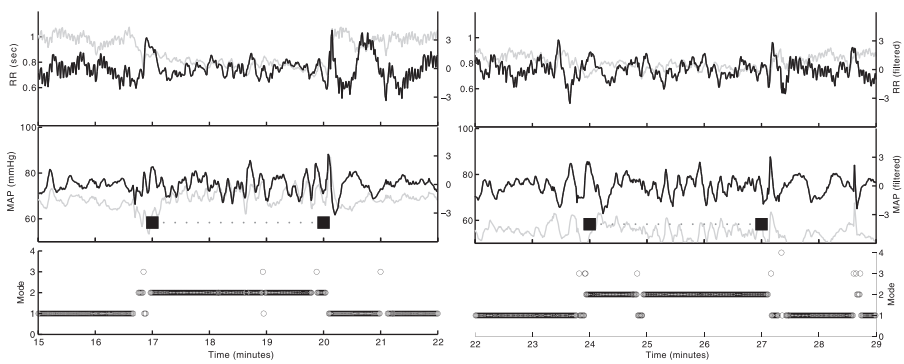
11.4 Results

11.4.1 Tilt-table experiment

Figure 11.3 shows the segmentation results for two subjects using HDP-AR-HSMM. Note that the two subjects share the same inferred non-supine dynamics (mode 1) and supine dynamics (mode 2); the algorithm consistently assigns the mode 1 to the non-supine position for both subjects. Application of logistic regression with ten-fold cross-validation yielded a median AUC of 1.00 with an interquartile range of (0.93, 1.00). The sticky HDP-AR-HMM and BP-AR-HMM yielded similar median and IQR performance of 1.00 (0.88, 1.00) and 1.00 (0.90, 1.00) respectively.

11.4.2 MIMIC II: performance in estimating mortality risks of patients

The median number of dynamic modes discovered (over ten folds) using the HDP-AR-HMM approach were 16 and 18 respectively for the HR and BP time series (modes with less than 1% of the overall samples were not considered), with the top ten dynamic modes capturing over 80% of the overall measurements in the entire data set. Table 11.1 evaluates the prognostic power of HR and BP dynamic features (HR_{dyn} and BP_{dyn}) learned from the HDP-AR-HSMM approach. We used mode proportions of the top ten dynamic modes as features for mortality prediction. SAPS-I and APACHE-IV were used as the baselines. Median AUCs (from ten-fold cross-validation) and the interquartile range are shown. Note that the BP dynamics out-performed the HR dynamic features in mortality prediction. Subsequent analysis focuses on the predictive power of the BP



(a) Tilt subject 1 (b) Tilt subject 6

Figure 11.3 Tilt-table study: segmentation from HDP-AR-HSMM. Two examples out of the ten recordings of HR and mean arterial blood pressure (MAP) from the tilt-table experiment are shown. Parts (a) and (b) show a 7-minute recording of RR intervals and MAP while the subjects transition to/from supine to a non-supine position after a fast tilt procedure. The two solid square markers denote the onset of the fast tilt procedures with the dotted horizontal lines indicating the duration while the subjects remain in a non-supine position. Actual HR/BP values are in light gray and filtered values are in black. Note that the two subjects shared the same inferred non-supine dynamics (mode 2). The supine position for both subjects are captured by mode 1.

Table 11.1 Mortality prediction performance of dynamic mode proportions inferred from vital signs from the first 24 hours in the ICU.

	Hospital mortality (AUC)	28-day mortality (AUC)
SAPS-I	0.65 (0.59, 0.71)	0.64 (0.56,0.70)
HR_{dyn}	0.60 (0.55, 0.70)	0.66 (0.54, 0.72)
BP_{dyn}	0.65 (0.59, 0.74)	0.62 (0.59, 0.72)
HR_{dyn}+SAPS-I	0.68 (0.63, 0.74)	0.66 (0.62, 0.76)
BP_{dyn}+SAPS-I	0.77 (0.60, 0.79)	0.73 (0.63, 0.79)
APACHE-IV	0.82 (0.77, 0.85)	0.83 (0.74, 0.86)
HR_{dyn}+APACHE-IV	0.82 (0.80, 0.89)	0.83 (0.77, 0.85)
BP_{dyn}+APACHE-IV	0.84 (0.74, 0.88)	0.82 (0.80, 0.85)

dynamics in comparison to the baseline. For each baseline, we show the performance from the baseline alone, and the combined approach (combining BP dynamics and the baseline).

Application of 10-fold cross-validation demonstrated that dynamic features from blood pressure alone achieved a median AUC 0.65 comparable to the performance from the SAPS-I (which required 13 different lab tests). In comparison, using standard deviation of mean arterial blood pressure resulted in a median AUC (IQR) of 0.55 (0.43, 0.63).

Combining dynamic blood pressure features with SAPS-I resulted in an improved prediction power both in hospital mortality prediction and 28-day mortality prediction. These results indicate that the dynamic features from vital signs contain complementary information to the SAPS-I scores. State-of-the-art risk score APACHE-IV achieved better prediction performance than the BP dynamic features alone. Adding BP dynamics to APACHE-IV improved the median hospital mortality prediction performance slightly, but the performance gain was not statistically significant.

The performance reported in Table 11.1 (using the HDP-AR-HSMM approach with geometric state duration distribution) is consistent with our prior results using BP-AR-HMM (Lehman *et al.* 2012) and SVAR (Lehman *et al.* 2014a). The baseline SAPS-I and APACHE-IV performance for the same patient cohort was previously reported (Lehman *et al.* 2014a).

11.4.3 MIMIC II association analysis

Table 11.2 presents logistical regression analysis to test the associations between the proportion of time patients spent in each of the top ten most common BP dynamics and hospital mortality. Figure 11.4 shows examples of low-risk and high-risk dynamical modes learned using the HDP-AR-HSMM technique (see Table 11.2 for the odds ratio (OR) associated with each mode).

Dynamic modes were numbered based on their prevalence across the entire cohort (i.e. mode 1 is the most common dynamic mode). Our results indicate that five of the

Table 11.2 Associations of blood pressure dynamic modes and hospital mortality. (OR = odds ratio; HL_p = Hosmer–Lemeshow p value.)

Mode	p-value	OR (95%CI)	Adjusted p	Adjusted OR (95%CI)	HL p
6	0.0000	1.81 (1.43, 2.28)	0.0004	1.59 (1.23, 2.06)	0.31
7	0.0005	2.01 (1.36, 2.97)	0.0198	1.71 (1.09, 2.67)	0.50
8	0.0058	1.30 (1.08, 1.56)	0.6827	1.05 (0.84, 1.30)	0.38
3	0.0034	0.63 (0.46, 0.86)	0.0271	0.71 (0.52, 0.96)	0.97
4	0.0039	0.53 (0.35, 0.82	0.1828	0.74 (0.47, 1.15)	0.09
1	0.0084	0.71 (0.55, 0.92)	0.0815	0.78 (0.60, 1.03)	0.50
10	0.0962	0.41 (0.14, 1.17)	0.0226	0.24 (0.07, 0.82)	0.13
9	0.1098	1.47 (0.92, 2.35)	0.8635	1.05 (0.58, 1.91)	0.39
5	0.1711	0.82 (0.62, 1.09)	0.7152	0.95 (0.71, 1.27)	0.07
2	0.7575	1.03 (0.86, 1.22)	0.4038	1.09 (0.89, 1.32)	0.69

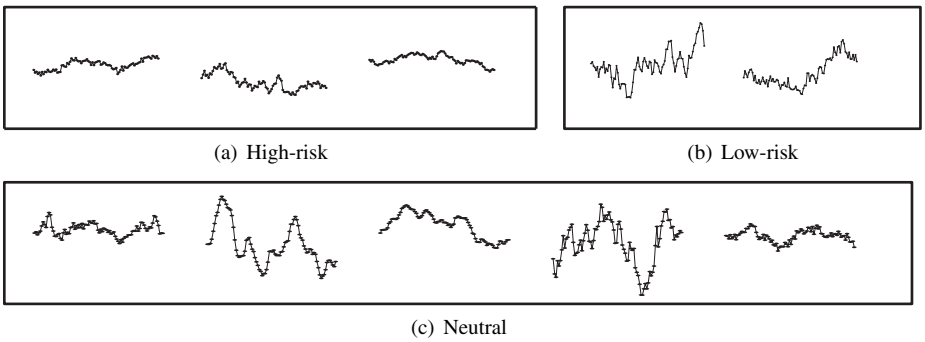


Figure 11.4 Discovered dynamic modes of mean arterial blood pressure of 453 patients during the first 24 hours in the ICU. Figure shows the top ten most common dynamic modes, simulated using the AR coefficients from each dynamic mode. (a) High-risk dynamic modes (from left to right): 6, 7 and 8. (b) Low-risk dynamic modes: 3 and 4. (c) Neutral dynamic modes: 1, 10, 9, 5 and 2. All modes were simulated and plotted with the same time duration (90 minutes) and amplitude scale.

modes (modes 6, 7, 8, 3 and 4) had significant associations with hospital mortality after FDR correction.

Three dynamic modes (modes 6, 7 and 8) were "high-risk" modes ($p < 0.0001$, $p < 0.001$, and $p < 0.01$) in which increased proportions of time in these modes were associated with higher hospital mortality with odds ratios 1.81 (1.43, 2.28), 2.01 (1.36, 2.97) and 1.30 (1.08, 1.56), respectively. Dynamic modes 3 and 4 were "low-risk" modes in which increasing proportions of time in these modes were significantly ($p < 0.01$) associated with a decreased risk of hospital mortality, with odds ratios less than one.

For the multivariate analysis (Table 11.2), each row is a separate multivariate model, in which the mode proportion for a given target mode is the primary predictive variable, and APACHE-IV is added as a control variable in the multivariate model. Results

from multivariate logistic regression indicate that two of the modes (modes 6 and 7) remain significant predictors of patients' outcome even after adjustment for APACHE-IV scores, indicating that the proportion of time patients spent in these two dynamic modes during the first 24 hours in the ICU are independent risk predictors of hospital mortality.

11.4.4 Example blood pressure dynamics of survivors vs. non-survivors

As examples, BP time series from four patients are presented in Figure 11.5. Figure 11.5(a) shows two of the patients (within the same test set) with a high proportion of time in high-risk dynamics (modes 6, 7, and 8); both patients died in the hospital. Figure 11.5(b) shows two patients with the top two highest proportions of time in the low-risk dynamics during the second half of their first day in the ICU; both patients survived the hospital stay. All four patients were from the same test set, with mode assignment inferred based on dynamic modes learned from the corresponding training set.

11.4.4.1 Evolution of cardiovascular dynamics of survivor vs. non-survivor

We also provide illustrative examples of the evolving blood pressure dynamics of the survivors and the non-survivors' during the first 24 hours in the ICU. Figure 11.6 shows blood pressure time series for two patients with different trajectories in the evolution of their blood pressure dynamics during the first day in the ICU. Both patients were from the medical ICU. Figure 11.6 (a) shows a patient with increasing high-risk mode proportions during the 24 hours in the ICU; patient died two days after admission to the ICU. Figure 11.6 (b) shows a patient with decreasing trend in the high-risk mode proportions during the first day in the ICU. Note that as time progresses, the patient in (a) spends more time in the "high-risk" dynamic modes. In contrast, the patient in (b) shows a decreasing trend in high-risk mode proportion and transitions to lower-risk and neutral dynamic modes over the course of the first 24 hours in the ICU. The high-risk mode proportion (shown on the right-hand side y-axis) was computed as sum of the mixture weights for high-risk modes 6, 7 and 8 in a six-hour sliding window, updated on an hourly basis.

These illustrative examples support results from our prior investigation which demonstrated that, at the population level, patients who did not survive the hospital stay exhibited different evolution in their vital sign dynamics than those who survived (Lehman *et al.* 2013). In particular, as time progressed, the non-survivors tend to have increasing trends in the proportion of time in the high-risk modes.

11.5 Discussion and conclusion

We presented a Bayesian nonparametric switching Markov processes framework to systematically learn and identify dynamic behaviors from multivariate vital sign time

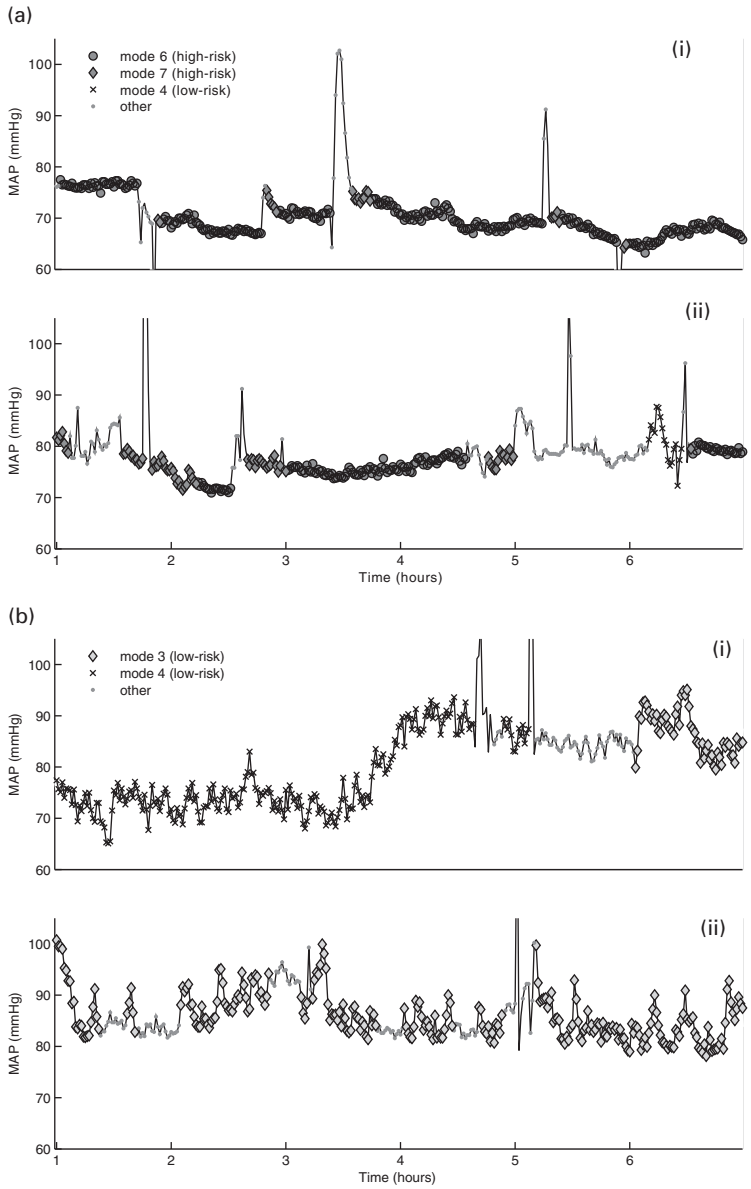


Figure 11.5 Example mean arterial blood pressure (MAP) of four patients sampled during the first 24 hours in the ICU. Blood pressure measurements plotted in original units (before de-trending). All four patients were from the same test set, with dynamic modes and their associated mortality risks learned from the corresponding training set. (a) Patients with high proportions of high-risk modes (6, 7 and 8) during their first day in the ICU. Both patients died in the hospital. Patients were from the medical ICU (i) and the coronary care ICU (ii) respectively. (b) Patients with the highest proportions of low-risk modes (3 and 4) during their first day ICU stays. Patients were from the cardiac surgery recovery unit (i) and the surgical ICU (ii). Both patients survived the hospital stay.

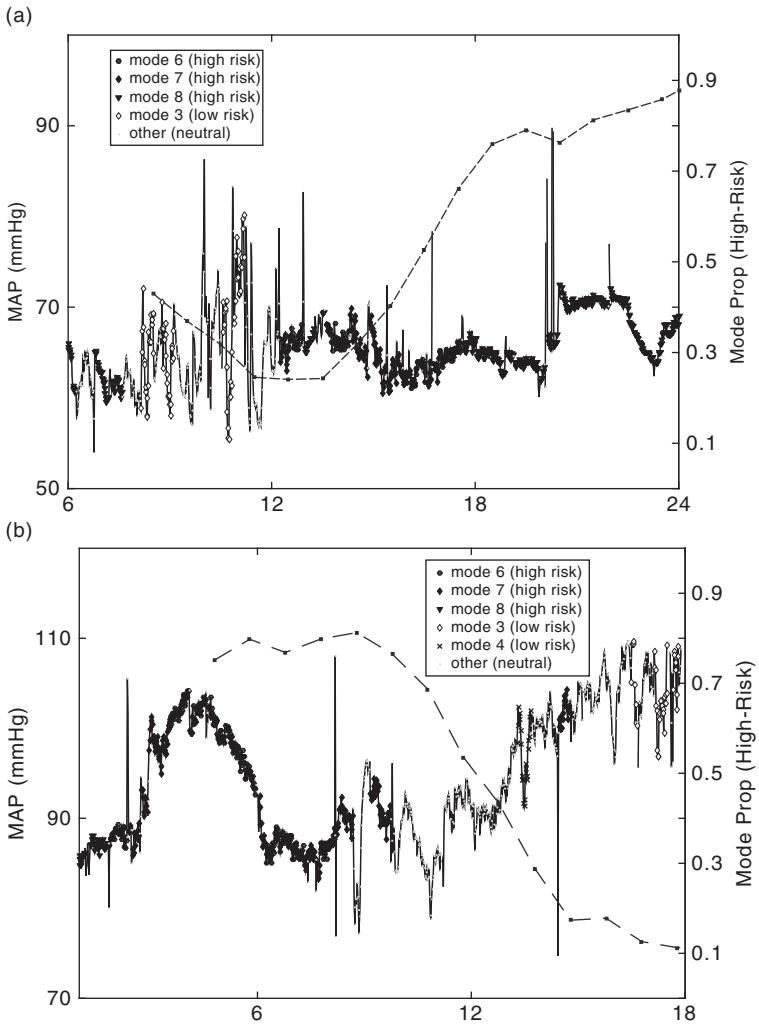


Figure 11.6 High-risk mode proportions and mean arterial blood pressure (MAP) of two patients during their first 24 hours in the ICU. Blood pressure measurements plotted in original units (before de-trending). The high-risk mode proportion was computed as sums of the mixture weights for high-risk modes 6, 7 and 8 in a six-hour sliding window updated on a hourly basis, and were plotted as dashed gray lines with scale indicated by *y*-axes on the right-hand side of each graph. (a) Mean arterial blood pressure (MAP) of a patient with increasing trend in high-risk dynamics over the first 24 hours in the ICU. Right-hand axis shows mode proportions in high-risk modes. Patient (from the medical ICU) died two days after admission to the ICU. Note the increasing trend in high-risk mode proportion as the patient transitions from low-risk and neutral dynamics to high-risk dynamics. (b) Mean arterial blood pressure (MAP) of a patient with decreasing trend in high-risk dynamics over the first 24 hours in the ICU. Right-hand axis shows mode proportion in high-risk modes. Note the decreasing trend in high-risk mode proportion as the patient transitions from high-risk dynamics to low-risk and neutral dynamics. Patient was from the medical ICU and survived the hospital stay.

series within a patient cohort. We explored several variants of Bayesian nonparametric approaches to model changes in dynamics of physiological time series as switching between a set of linear dynamical systems. We showed that the proposed framework is able to automatically capture changes in the dynamics of HR and BP due to external perturbations (i.e., positional changes in the tilt-table experiment). In evaluating the prognostic value of the dynamic modes, we focused on the predictive power of the discovered dynamic modes and their associations with hospital mortality.

Commonly used acuity scores for patient prognosis, such as APACHE and SAPS (Le Gall *et al.* 1984, 1993; Knaus *et al.* 1991; Zimmerman *et al.* 2006), are based on snap-shot values of these vital signs, typically the worst values during a 24 hour period. A growing body of literature points to the clinical utility of vital signs time series dynamics to inform prognosis (Saria *et al.* 2010; Moorman *et al.* 2011; Lehman *et al.* 2012, 2014a; Wiens *et al.* 2012; Mayaud *et al.* 2013), and to provide early predictors of potentially life-threatening conditions in the ICU (Blount *et al.* 2010; Lehman *et al.* 2013).

In this work, we showed that the proportion of time each patient spent within the different dynamic modes is a significant predictor of hospital mortality risks. It is interesting to note that the BP time series dynamics alone achieved similar performance to that of the SAPS-I score which uses age, Glasgow coma score, and the most extreme values of 13 lab variables. Furthermore, our results indicate that the blood pressure dynamics may contain complimentary information to existing acuity metrics, which assess the health of multiple organ systems based on a variety of physiological and lab variables. Specifically, combining the dynamics of BP time series and SAPS-I provided a more accurate assessment of patient survival or mortality in the hospital than using SAPS-I alone.

Association analysis using the minute-by-minute MIMIC-II BP time series revealed that the high-risk modes often correspond to less variable dynamical patterns. Interestingly, such low-frequency variability, observed at the minute-to-minute scale, is associated with an enhanced chance of survival, corresponding well to existing HR/BP variability literature using beat-by-beat vital sign time series (Riordan *et al.* 2009; Moorman *et al.* 2011; Parati *et al.* 2013). The working hypothesis of our ongoing research is that the observed dynamical patterns are due to patients' underlying physiology, patient-specific response to clinical interventions, and measurement artifacts.

Since the proposed framework is built on the dynamical systems framework (which includes the class of VAR models), the discovered modes can be used to reveal the oscillations that are present within the individual time series, and therefore can be used to extract useful indices of HR and BP variability (assuming beat-to-beat time series). Moreover, in the multivariate case one may use the learned dynamics to derive the directional transfer functions of the system (e.g., baroreflex control of HR and BP) (Nemati *et al.* 2011).

This approach provides an improvement over time series similarity measures based on trend-detection (Avent & Charlton 1990), wavelet-based symbolic representations (Saeed & Mark 2006), or Gaussian mixture modeling (Lehman *et al.* 2008) due to its compact representation and sharing of the model parameters within and across

time series. Prior work using a factorial SLDS for patient monitoring focused on detection of events associated with artifactual measurements and pathological states (Quinn *et al.* 2009). Our work, in contrast, jointly models multiple time series across a large patient cohort to identify phenotypic dynamical patterns for patient outcome prediction.

We used the HDP-AR-HSMM with geometric duration distributions to test the prognostic values of the common dynamic modes. This approach provides a more direct and interpretable Bayesian prior in controlling the self-transition bias than the sticky model, without incurring additional computational cost to the Gibbs sampling algorithm. In contrast to the sticky HDP-AR-HMM approach, in which the global self-transition bias is shared among all states (Johnson & Willsky 2013a), the HSMM framework has the additional advantage of being able to model state-specific duration distributions. Although we focus on the geometric duration distribution in this chapter and did not fully exploit the learned state-specific duration information, the models presented in (Johnson & Willsky 2013a; Johnson 2014) provide a more expressive and powerful framework to model non-Markovian state durations in general. Future work involves learning highly interpretable dynamic behaviors from vital sign time series to capture non-geometric state durations.

The beta process framework, explicitly models the dynamic behaviors to determine both the shared and unique dynamic behaviors from a collection of time series (Fox *et al.* 2014; Fox 2009). Through this feature-based representation, BP-AR-HMM permits time series specific transition behaviors. The HDP prior, on the other hand, assumes that all time series share the same dynamic modes and transition between them in the same manner (Fox 2009). As the focus of this current investigation is on the prognostic value of the common (instead of rare) dynamic behaviors, the proposed variant of the HDP-AR-HSMM approach (with geometric duration distribution) provides a computationally efficient approach to learn common prognostic dynamic behaviors from a large patient cohort. There are other aspects of the learned dynamical models that need to be investigated further, for example, dynamic mode distributions and discovery of rare or unique dynamic behaviors. Future work aims to conduct further performance comparison of these techniques, and fully characterize the effects of the HDP and beta process priors and hyperparameter settings on the discovered dynamic behaviors in the context of physiological patient monitoring.

In summary, we presented a framework to discover prognostic dynamical behaviors from vital sign time series in a critical care setting. In particular, our results demonstrate that the discovered dynamics provide additional predictive values to conventional snapshot-based acuity metrics, and thus the proposed approach holds promise to providing additional insights to state of health of patients. Future and ongoing work involve combining the switching dynamics framework with clinical data, including lab tests, medication records, progress notes, and clinical interventions (administration of fluids, pressors, and titration of medications) to further investigate the clinical and physiological interpretation of the discovered dynamic modes (Lehman *et al.* 2014b), and to devise a comprehensive risk score capable of continuous patient monitoring and treatment decision support.

Acknowledgments

The authors thank Dr. Thomas Heldt (MIT Institute for Medical Engineering & Science) for kindly providing the tilt-table data analyzed in this study, and Dr. Louis Mayaud for deriving the APACHE-IV scores used in this manuscript. This work was supported by the National Institutes of Health (NIH) grant R01-EB001659 and R01GM104987 from the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the James S. McDonnell Foundation Postdoctoral grant, and the DARPA Young Faculty Award N66001-12-1-4219 grant. The content of this chapter is solely the responsibility of the authors and does not necessarily represent the official views of the NIBIB or the NIH.

References

Angelino, E., Kohler, E., Waterland, A., Seltzer, M. & Adams, R. P. (2014). Accelerating MCMC via parallel predictive prefetching. In *Proceedings of 30th Conference on Uncertainty in Artificial Intelligence*.

Avent, R. K. & Charlton, J. D. (1990). A critical review of trend-detection methodologies for biomedical monitoring systems. *Critical Reviews in Biomedical Engineering* **17**(6), 621–59.

Bardenet, R., Doucet, A. & Holmes, C. (2014). Towards scaling up Markov chain Monte Carlo: an adaptive subsampling approach. In *Proceedings of the 31st International Conference on Machine Learning*, pp. 405–413.

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* **57**(1), 289–300.

Bernardo, J. M. & Smith, A. F. (2009). *Bayesian Theory*, New York: Wiley.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*, New York: Springer.

Blount, M., Ebling, M. R., Eklund, J. M., James, A. G., McGregor, C., Percival, N., Smith, K. P. & Sow, D. (2010). Real-time analysis for intensive care: development and deployment of the artemis analytic system. *IEEE Engineering in Medicine and Biology Magazine* **29**(2), 110–118.

Costa, M., Goldberger, A. L. & Peng, C. K. (2002). Multiscale entropy analysis of complex physiologic time series. *Physical Review Letters* **89**(6), 068102.

Fox, E. (2009). Bayesian nonparametric learning of complex dynamical phenomena. PhD thesis, MIT.

Fox, E. B., Sudderth, E. B., Jordan, M. I. & Willsky, A. S. (2008). An HDP-HMM for systems with state persistence. In *Proceedings of the International Conference on Machine Learning*.

Fox, E. B., Sudderth, E., Jordan, M. & Willsky, A. (2010). Bayesian nonparametric learning of Markov switching processes. *IEEE Signal Processing Magazine* **27**, 43–54.

Fox, E., Hughes, E., Sudderth, E. & Jordan, M. (2014). Joint modeling of multiple related time series via the beta process with application to motion capture segmentation. *Annals of Applied Statistics* **8**, 1281–1313.

Fox, E., Sudderth, E., Jordan, M. & Willsky, A. (2009). Sharing features among dynamical systems with beta processes. *Advances in Neural Information Processing Systems 22* pp. 549–557.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2013). *Bayesian Data Analysis*, 3rd edn, CRC press.

Ghahramani, Z., Griffiths, T. L. & Sollich, P. (2006). Bayesian nonparametric latent feature models. In *Proceedings of ISBA 8th World Meeting on Bayesian Statistics*.

Heldt, T. (2004). Computational Models of Cardiovascular Response to Orthostatic Stress. PhD thesis, MIT.

Heldt, T., Oefinger, M. B., Hoshiyama, M. & Mark, R. G. (2003). Circulatory response to passive and active changes in posture. *Proceedings of the Computers in Cardiology* **30**, 263–266.

Hoffman, M. D., Blei, D. M., Wang, C. & Paisley, J. (2013). Stochastic variational inference. *Journal of Machine Learning Research* **14**(1), 1303–1347.

Ishwaran, H. & Zarepour, M. (2002). Exact and approximate sum representations for the Dirichlet process. *Canadian Journal of Statistics* **30**(2), 269–283.

Ivanov, P. C., Amaral, L. A., Goldberger, A. L., Havlin, S., Rosenblum, M. G., Struzik, Z. R. & Stanley, H. E. (1999). Multifractality in human heartbeat dynamics. *Nature* **399**, 461–465.

Johnson, M. & Willsky, A. (2013a). Bayesian nonparametric hidden semi-markov models. *Journal of Machine Learning Research* **14**, 673–701.

Johnson, M. & Willsky, A. (2014). Stochastic variational inference for Bayesian time series models. In *Proceedings of the 31st International Conference on Machine Learning*, pp. 1854–1862.

Johnson, M. J. (2014). Bayesian time series models and scalable inference. PhD thesis, MIT.

Johnson, M. J. & Willsky, A. S. (2013b). Bayesian nonparametric hidden semi-markov models. *Journal of Machine Learning Research* **14**(1), 673–701.

Knaus, W. A., Wagner, D. P., Draper, E. A., Zimmerman, J. E., Bergner, M., Bastos, P. G., Sirio, C. A., Murphy, D. J., Lotring, T., Damiano, A. & Harrell, F. (1991). The APACHE III prognostic system. *Chest* **100**(6), 1619–1636.

Korattikara, A., Chen, Y. & Welling, M. (2014). Austerity in MCMC land: cutting the Metropolis-Hastings budget. In *Proceedings of The 31st International Conference on Machine Learning*, pp. 181–189.

Le Gall, J. R., Lemeshow, S. & Saulnier, F. (1993). A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. *Journal of the American Medical Association* **270**, 2957–2963.

Le Gall, J. R., Loirat, P., Alperovitch, A., Glaser, P., Granthil, C., Mathieu, D., Mercier, P., Thomas, R. & Villers, D. (1984). A simplified acute physiology score for ICU patients. *Critical Care Medicine* **12**(11), 975–977.

Lehman, L. H., Adams, R. P., Mayaud, L., Moody, G. B., Malhotra, A., Mark, R. G. & Nemati, S. (2014a). A physiological time series dynamics-based approach to patient monitoring and outcome prediction. *IEEE Journal of Biomedical and Health Informatics* **18**, in press.

Lehman, L. H., Nemati, S., Adams, R. P. & Mark, R. G. (2012). Discovering shared dynamics in physiological signals: Application to patient monitoring in ICU. In *Proceedings of the IEEE Engineering in Medicine and Biology Society*, pp. 5939–5942.

Lehman, L. H., Nemati, S., Adams, R. P., Moody, G., Malhotra, A. & Mark, R. G. (2013). Tracking progression of patient state of health in critical care using inferred shared dynamics in physiological time series. In *Proceedings of IEEE Engineering in Medicine and Biology Society*, pp. 7072–7075.

Lehman, L. H., Nemati, S., Moody, G. B., Heldt, T. & Mark, R. (2014b). Uncovering clinical significance of vital sign dynamics in critical care. In *Proceedings of the Computing in Cardiology*.

Lehman, L. H., Saeed, M., Moody, G. & Mark, R. (2008). Similarity-based searching in multi-parameter time series databases. In *Proceedings of the Computers in Cardiology*, pp. 653–656.

Maclaurin, D. & Adams, R. P. (2014). Firefly Monte Carlo: exact MCMC with subsets of data. In *Proceedings of 30th Conference on Uncertainty in Artificial Intelligence*.

Mancia, G. (2012). Short-and long-term blood pressure variability present and future. *Hypertension* **60**(2), 512–517.

Mayaud, L., Lai, P. S., Clifford, G. D., Tarassenko, L., Celi, L. A. & Annane, D. (2013). Dynamic data during hypotensive episode improves mortality predictions among patients with sepsis and hypotension. *Critical Care Medicine* **41**(4), 954–962.

Moorman, J. R., Delos, J. B., Flower, A. A., Cao, H., Kovatchev, B. P., Richman, J. S. & Lake, D. E. (2011). Cardiovascular oscillations at the bedside: early diagnosis of neonatal sepsis using heart rate characteristics monitoring. *Physiological Measurements* **32**(11), 1821–1832.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*, Cambidge, MA: MIT Press.

Nemati, S. (2012). Identifying evolving multivariate dynamics in individual and cohort time-series, with application to physiological control systems. PhD thesis, MIT.

Nemati, S., Edwards, B. A., Sands, S. A., Berger, P. J., Wellman, A., Verghese, G. C., Malhotra, A. & Butler, J. P. (2011). Model-based characterization of ventilatory stability using spontaneous breathing. *Journal of Applied Physiology* **111**(1), 55–67.

Nemati, S., Lehman, L. H., Adams, R. P. & Malhotra, A. (2012). Discovering shared cardiovascular dynamics within a patient cohort. In *Proceedings of IEEE Engineering in Medicine and Biology Society*, pp. 6526–6529.

Nishihara, R., Murray, I. & Adams, R. P. (2014). Parallel MCMC with generalized elliptical slice sampling. *Journal of Machine Learning Research* **15**, 2087–2112.

Parati, G., Ochoa, J. E., Lombardi, C. & Bilo, G. (2013). Assessment and management of blood-pressure variability. *Nature Reviews Cardiology* **10**, 143–155.

Peng, C. K., Havlin, S., Stanley, H. E. & Goldberger, A. L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos* **5**, 82–87.

Quinn, J. A., Williams, C. K. & McIntosh, N. (2009). Factorial switching linear dynamical systems applied to physiological condition monitoring. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **31**(9), 1537–1551.

Riordan, W. P., Norris, P. R., Jenkins, J. M. & Morris Jr, J. A. (2009). Early loss of heart rate complexity predicts mortality regardless of mechanism, anatomic location, or severity of injury in 2178 trauma patients. *Journal of Surgical Research* **156**(2), 283–289.

Robert, C. P. & Casella, G. (2004). *Monte Carlo Statistical Methods*, New York: Springer.

Saeed, M. & Mark, R. (2006). A novel method for the efficient retrieval of similar multiparameter physiologic time series using wavelet-based symbolic representations. In *Proceedings of the AMIA Annual Symposium*, pp. 679–683.

Saeed, M., Villarroel, M., Reisner, A. T., Clifford, G., Lehman, L. H., Moody, G., Heldt, T., Kyaw, T. H., Moody, B. & Mark, R. G. (2011). Multiparameter intelligent monitoring in intensive care (MIMIC II): a public-access intensive care unit database. *Critical Care Medicine* **39**(5), 952–960.

Saria, S., Rajani, A. K., Gould, J., Koller, D. & Penn, A. (2010). Integration of early physiological responses predicts later illness severity in preterm infants. *Science Translational Medicine* **2**, 48–65.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4**, 639–650.

Teh, Y., Jordan, M., Beal, M. & Blei, D. (2006). Hierarchical Dirichlet processes. *Journal of American Statistical Association* **101**, 1566–1581.

Van Gael, J., Saatci, Y., Teh, Y. W. & Ghahramani, Z. (2008). Beam sampling for the infinite hidden Markov model. In *Proceedings of the 25th International Conference on Machine Learning*, pp. 1088–1095.

Wainwright, M. J. & Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning* **1**(1-2), 1–305.

Wiens, J., Horvitz, E. & Guttag, J. V. (2012). Patient risk stratification for hospital-associated C. *diff* as a time-series classification task. In *Advances in Neural Information Processing Systems 25*, pp. 476–484.

Zimmerman, J. E., Kramer, A. A., McNair, D. S. & Malila, F. M. (2006). Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Critical Care Medicine* **34**(5), 1297–1310.