

Washington University in St. Louis
Washington University Open Scholarship

All Theses and Dissertations (ETDs)

January 2011

Hidden Markov Models for Heart Rate Variability with Biometric Applications

Michael Walker II

Washington University in St. Louis

Follow this and additional works at: <http://openscholarship.wustl.edu/etd>

Recommended Citation

Walker II, Michael, "Hidden Markov Models for Heart Rate Variability with Biometric Applications" (2011). *All Theses and Dissertations (ETDs)*. 491.

<http://openscholarship.wustl.edu/etd/491>

This Thesis is brought to you for free and open access by Washington University Open Scholarship. It has been accepted for inclusion in All Theses and Dissertations (ETDs) by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS
School of Engineering and Applied Science
Department of Electrical & Systems Engineering

Thesis Examination Committee:

Dr. Joseph A. O'Sullivan

Dr. Paul S. Min

Dr. Robert E. Morley

Dr. John W. Rohrbaugh

HIDDEN MARKOV MODELS FOR HEART RATE VARIABILITY WITH
BIOMETRIC APPLICATIONS

by

Michael R. Walker II

A thesis presented to the Graduate School of Arts and Sciences
of Washington University in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

May 2011
Saint Louis, Missouri

copyright by
Michael R. Walker II
2011

ABSTRACT OF THE THESIS

Hidden Markov Models for Heart Rate Variability with Biometric Applications

by

Michael R. Walker II

Master of Science in Electrical Engineering

Washington University in St. Louis, May 2011

Research Advisor: Dr. Joseph A. O'Sullivan

The utility of hidden Markov models (HMM) for modeling individual heart rate variability (HRV) is presented. Starting with a physiologically based statistical model for HRV from the literature, we justify use of HMMs and present methods for parameterizing the model. The forward-backward algorithm and expectation-maximization algorithm are used to estimate the model and the hidden states for a given observation sequence of inter-beat intervals. Multiple initialization techniques are presented to avoid local maxima. Model order is determined from the data sequence using the Bayesian information criterion. Models are trained on twelve hour recordings. The models are then used to discriminate the identity of an individual using data from a separate set of testing data. For database from 52 individuals, true identity was verified with an equal error rate of roughly 0.36. While initial results do not demonstrate strong performance as a biometric, HMMs are able to capture some individuality in the HRV signal. Consistency in HRV over twelve hour time scales is also demonstrated.

Acknowledgments

I am very grateful for the advice and support of my advisor, Dr. Joseph A. O'Sullivan. Thank you for facilitating a challenging research environment while providing valuable insight. I would like to thank my committee members Dr. John W. Rohrbaugh, Dr. Robert E. Morley, and Dr. Paul S. Min for both their time and insightful comments. Thanks to Dr. Alan Kaplan for providing new perspectives on my research problems.

Michael R. Walker II

Washington University in Saint Louis
May 2011

I dedicate this thesis to my wife, Erin, for her patience and support through the many late nights and weekends I spent at my computer.

Contents

Abstract	ii
Acknowledgments	iii
List of Tables	vi
List of Figures	vii
1 Introduction	1
2 A Physiologically Motivated Model	5
2.1 Model Description	6
2.2 Model Properties	10
2.2.1 Applicability of HMMs	10
2.2.2 Target Mean as a Random Process	11
2.2.3 State Changes and Transition Probabilities	13
3 HMMs and Parameter Estimation	15
3.1 Forward-Backward Algorithm	16
3.2 EM Algorithm	19
3.3 Summary	24
4 Model Selection	26
4.1 Bayesian Information Criterion	27
4.2 Initial Parameter Selection	28
4.2.1 Method 1: Basic	29
4.2.2 Method 2: Increase	29
4.2.3 Method 3: Distribute	30
4.3 Results	32
4.4 Discussion	36
5 Biometric Capability	46
5.1 Recognition Algorithm	46
5.2 Results	47
5.3 Discussion	49
6 Conclusion	51
References	53
Vita	56

List of Tables

2.1	IAGS Model Parameters	9
4.1	Model Performance by Initialization Method	33

List of Figures

2.1	Simulated IBI sequences from unbiased and biased random walks . . .	7
2.2	Distribution of the noise term ψ	11
2.3	Distribution on the number of intervals between updates of the target mean	12
2.4	Distribution on the change in target mean after update	13
4.1	Performance of initial parameter selection methods	38
4.2	Effects of initial parameter selection methods on individual 38	39
4.3	Effects of initial parameter selection methods on individual 33	40
4.4	Histogram of optimal model sizes	41
4.5	Time domain signals of IBI sequences from individual 30	41
4.6	Autocorrelation function and autocorrelation coefficients of IBI sequences from individual 30	42
4.7	Autocorrelation function and autocorrelation coefficients of IBI sequences from individual 19	43
4.8	Average correlation coefficients for individual 30.	43
4.9	Average correlation coefficients for individual 19.	44
4.10	Power spectral density of IBI sequences from individual 30	44
4.11	Power spectral density of IBI sequences from individual 19	45
5.1	Equal error rate vs. testing sequence length	48
5.2	FMR and FNMR vs. threshold	49

Chapter 1

Introduction

Heart rate variability is affected by a variety of physiological activity from respiration and swallowing to mental activity, even anatomical positioning. Much research has focused on quantifying changes in inter-beat intervals (IBI) due to such activities. IBI sequences can provide insight to the physiological state of an individual. However, it is often difficult to identify the stimuli which effected a specific change in the IBI.

Over short time scales, few IBI observations are made. This limits the number of meaningful statistics that can be computed. Over longer time scales, heart rate variability is non-stationary. While many approaches to heart rate variability (HRV) analysis rely on power spectral density estimation, unbiased estimation requires the signal is at least wide-sense-stationary. This has led to the proposal of several detrending methods [1, 2] mitigating the effects of state changes.

Many analytical techniques have been proposed for understanding HRV [3]. However, few of these techniques are able to reproduce or simulate new data sequences. Modeling a process is often insightful in and of itself. Parameter estimates provide new metrics for contrasting training data.

Of the relatively few models presented, drastically different forms are considered. DeBoer *et al.* coupled blood pressure with IBI duration using a set of difference

equations [4]. Kuusela *et al.* used a single difference equation, but added a stochastic component [5]. Ivanov *et al.* [6] modeled IBI sequences in the context of biased random walks.

The model presented by Ivanov *et al.* demonstrated an ability to capture both short- and long-term variability in IBI duration. In this model they present competing biased walkers which represent unique feedback control mechanisms of the cardiac system. Parameter selection for each biased walker was physiologically justified, and the model was demonstrated to generate normal sequences over a span of parameter assignments. However, no method was presented to individualize parameter selection.

The model presented by Ivanov *et al.* [6] suggests the notion of an underlying state that governs the statistics of short-term variability. In this way, hidden Markov models can be considered an extension of the physiologically justified model presented by Ivanov. Algorithms for estimating parameters of HMMs are well known [7], thus providing an algorithmic approach to parameter selection from training data.

HMMs are perhaps best known for their application in speech and language processing. HMMs have been used to model cardiac pulse waveforms [8, 9], but few publications present the results of applying HMMs to IBI sequences directly, or the results of generating IBI sequences from HMMs. Baier *et al.* modeled HRV, but also included blood pressure data in their observation sequences [10]. At the time of this writing, the only publications known to the author which present the results of applying HMMs to IBI sequences alone are by Silipo *et al.*, in which hour-long segments of IBI data were used as training sequences [11, 12].

We focused on long (12-hour and 24-hour) IBI sequences, following the work of Ivanov *et al.* [6]. We obtained long-term recordings of IBI data from the publicly available archive PhysioBank [13]. Specifically, we used the “Normal Sinus Rhythm RR Interval

Database,” which contains roughly 24 hours of intervals from 54 individuals (30 men, aged 28.5 to 76, and 24 women, aged 58 to 73). The IBI data were originally obtained from ECG recordings sampled at 128 Hz.

Common methods for parameterizing HMMs require the likelihood function to be convex over the parameter values to ensure maximum-likelihood (ML) estimates of the parameters. In selecting parameters to optimize the likelihood of a particular data sequence, initial results show the likelihood function contains many local maxima over the range of feasible parameter values. Additionally, evaluation of the likelihood function is computationally expensive for long sequences of IBIs ($>10,000$). For parameterizing models of 12-hour IBI sequences, the choice of optimization algorithms is severely limited.

It is also important to consider the number of states necessary to accurately model a sequence of interest. A well-known tradeoff exists between (over) fitting training data and the uncertainty of new testing data. We used a principled approximation, the Bayesian information criterion (BIC), for optimally selecting model complexity.

To assess the utility of applying HMMs to IBI data, we consider using IBI data as a biometric. More specifically, we consider the capability of matching an observed sequence of IBIs with a known model. In this sense, we measure the ability of HMMs to capture an individual’s uniqueness in IBI sequences, as well as the consistency or similarity across IBI segments from one individual. IBI data has received little attention as a biometric. In some cases, IBI data was explicitly removed from signals before considering biometric capabilities [14]. Many models for HRV do not focus on normal variations across individuals, but instead focus on recognizing abnormal phenomenon across population groups [6, 3]. However, improving models of normal variations could be used to improve reliability of detecting specific conditions.

This thesis is organized as follows. In Chapter 2 we present a physiologically justified model from the literature and identify similarities with HMMs. In Chapter 3, the algorithm for parameterizing the HMM is presented. We discuss a constraint for sizing the model in Chapter 4 and the results of model selection on 12- and 24-hour data. An experiment for testing biometric capabilities of HMMs trained on HRV data is described in Chapter 5.

Chapter 2

A Physiologically Motivated Model

Living organisms are subject to an external environment that is ever changing. Despite external stimuli, it is important to maintain control of internal physiological functions. This ability to maintain control following perturbations is commonly referred to as homeostasis. The cardiac cycle is certainly one physiological function to which the principle of homeostasis applies. For example, while resting after moments of physical exertion heart rate tends toward a normal resting rate.

However, homeostasis does not account for HRV while resting. Homeostasis, alone, would suggest that heart rate would reach a steady state after long periods of resting. This, simply, has not been observed. Balancing HRV with homeostasis lead Ivanov, Amaral, *et al.* to model sequences of IBIs as biased random walks [6, 15]. The authors first demonstrate the ability of a random walk to balance homeostasis with short term HRV. They then expand upon a basic biased walker to accommodate variability over many time scales.

In this chapter we take a detailed look at the model presented by Ivanov, Amaral, Goldberger, and Stanley in [6, 15]. For simplicity, we will refer to this model as the IAGS model. In Sec. 2.1 we will first present the IAGS model and derive a

complementary expression for the model. We will then use this alternative expression to reveal characteristics of the IAGS model in Sec. 2.2.

2.1 Model Description

Using y_n to denote the n^{th} IBI duration, consider the random walk

$$y_n = y_{n-1} + w_n \quad (2.1)$$

where w_n represents the n^{th} observation of a random variable. If we briefly consider a case where w_n are independent and identically distributed (i.i.d.) Gaussian random variables with mean μ and variance σ^2 , it is clear that $y_n - y_0 \sim \mathcal{N}(n\mu, n\sigma^2)$. Even for $\mu = 0$, variance increases with n and homeostasis is not preserved. For this reason, the authors introduce a bias to the walker. The biasing method proposed affects the distribution for w_n such that

$$w_n = \begin{cases} \xi(1 + \eta_n) & \text{for } y_{n-1} < \tau \\ -\xi(1 + \eta_n) & \text{for } y_{n-1} \geq \tau \end{cases}. \quad (2.2)$$

This approach biases the walker toward a target mean τ , where ξ represents a weighting term. Noise is introduced through the terms η_n and are assumed independent, zero-mean, Laplace random variables with standard deviation σ . Since the Laplace distribution is symmetric, we can separate the conditional bias from the noise term. We can equivalently express (2.1) as

$$y_n = y_{n-1} + b_n + \acute{\eta}_n. \quad (2.3)$$

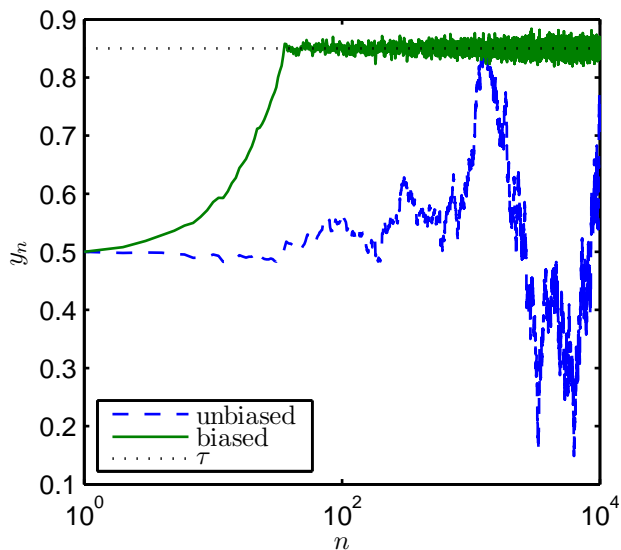


Figure 2.1: Simulated IBI sequences from both unbiased and biased random walks using (2.3). This contrasts the resulting sequences from a system which starts in a perturbed state. The values $y_0 = 0.5$, $\sigma = 5 \times 10^{-3}$, and $\xi = 0.01$ were used for both models. For the biased walker, $\tau = 0.85$. For the unbiased walker, $b_n = 0$. The biased walker drifts toward the target mean τ and the statistics of the observed IBI sequences appear consistent over large n . For the unbiased walker, the variance of y_n increases with n .

Here, $\hat{\eta}_n$ represents a zero-mean Laplace random variable with standard deviation $\xi\sigma$ and

$$b_n = \begin{cases} \xi & \text{for } y_{n-1} < \tau \\ -\xi & \text{for } y_{n-1} \geq \tau \end{cases}. \quad (2.4)$$

Clearly, (2.3) represents a biased walker. The walker becomes unbiased by simply setting $b_n = 0$, $\forall n = \{1, 2, \dots\}$. Fig. 2.1 demonstrates the effects of including the bias term for a simulated sequence of IBIs. The sequence shown in Fig. 2.1 simulates the response of a system which starts in a perturbed state. The initial value, $y_0 = 0.5$, was selected far from the desired mean $\tau = 0.85$. Qualitatively, this demonstrates the ability of biased walker to return to a mean value after a perturbation while simultaneously facilitating slight variability about the target mean. Conversely, the unbiased walker does not display either of these features.

Fig. 2.1 also demonstrates some of the shortcomings of the model as presented thus far. The biased walker model appears to balance homeostasis with short term variability in IBI sequences. However, over long time scales, HRV must adapt to changes in sleep cycle, physical activity and similar changes in state. The authors modify the biased walker model in two ways to accommodate long term variability.

In the first modification, the authors allow the desired mean τ to vary over time. In this way, τ_n represents a discrete-time continuous-valued random process. In the second modification, the authors expand the model to include K independent random processes in the update term such that

$$y_n = y_{n-1} + \sum_{k=1}^K w_{k,n}, \quad (2.5)$$

where $w_{k,n}$ represents the n^{th} observation of the k^{th} random process w_k and

$$w_{k,n} = \begin{cases} \xi_k(1 + \eta_{k,n}) & \text{for } y_{n-1} < \tau_{k,n} \\ -\xi_k(1 + \eta_{k,n}) & \text{for } y_{n-1} \geq \tau_{k,n} \end{cases}. \quad (2.6)$$

It is through the random processes, w_k , by which the authors model physiological control mechanisms influencing IBI duration. $\tau_{k,n}$ represents the n^{th} observation of the k^{th} random process τ_k . $\eta_{k,n}$ are still assumed i.i.d. Laplace random variables with mean 0 and standard deviation σ . The weighting terms ξ_k vary by process k , but are assumed constant over all n .

It is by (2.5) and (2.6), the IAGS model is described. As mentioned previously, we can alternatively express (2.5) as

$$y_n = y_{n-1} + \beta_n + \psi_n. \quad (2.7)$$

Table 2.1: IAGS Model Parameters

	SA	PS	SS
k	1	2	3, ..., 9
ξ_k	0.01	0.03	0.01
σ	0.5	0.5	0.5
τ_k	0.6	U(0.9, 1.5)	U(0.2, 1.0)

In this case, β_n represents the sum of all bias terms at the current time step

$$\beta_n = \sum_{k=1}^K b_{k,n}, \quad (2.8)$$

where

$$b_{k,n} = \begin{cases} \xi_k & \text{for } y_{n-1} < \tau_{k,n} \\ -\xi_k & \text{for } y_{n-1} \geq \tau_{k,n} \end{cases}. \quad (2.9)$$

In (2.7), ψ_n is a random variable representing the sum of K independent, zero-mean, Laplace random variables

$$\psi_n = \sum_{k=1}^K \acute{\eta}_{k,n} \quad (2.10)$$

where the standard deviation of $\acute{\eta}_{k,n}$ is $\xi_k \sigma$.

In [6], the authors present results from a model with parameter values that were physiologically justified. $K = 9$ separate random processes, $w_{k,n}$, were used to model three inputs: the sinoatrial (SA) node, the parasympathetic (PS) and sympathetic (SS) branches of the autonomic nervous system. One process each was used to model the SA and PS inputs. Seven processes were used to model the SS inputs. For the SA input, $\tau_k = 0.6$ was held constant. For the PS and SS inputs, the values τ_K were drawn from uniform distributions. Parameter values are listed in Table 2.1.

The update times for τ_k described in [6, 15] were ambiguous. In both publications, update times were considered normally distributed with mean $T = 1000$ but without

mention of variance. Perhaps a suitable alternative is to assume a geometric distribution with mean $T = 1000$ intervals. The geometric distribution, along with its continuous time analog the exponential distribution, is memoryless and is often used in queuing theory to describe inter-arrival times. Still, other distributions may be more appropriate for modeling these update times.

2.2 Model Properties

The expression (2.7) differs substantially from that proposed by Ivanov in [6], but the models are the same. From this new expression, several interesting features are apparent. In this section we identify several characteristics not previously presented by the authors.

2.2.1 Applicability of HMMs

It is interesting to note that the statistics of the noise term ψ_n do not change over time nor by outcome τ_k . This term follows a nonstandard distribution, but a likelihood function, computed empirically, for ψ_n is shown in Fig. 2.2. Furthermore, it was computed that $|\psi_n| < 0.041$ with probability 0.95.

The value β_n is deterministic given $\Phi_n = \{y_{n-1}, \tau_{1,n}, \dots, \tau_{K,n}\}$. If the random processes τ_k are memoryless (e.g. update times are geometrically distributed), then the likelihood function $f(\Phi_{n+1} | \Phi_n, \Phi_{n-1}, \dots) = f(\Phi_{n+1} | \Phi_n)$. In which case, the sequence Φ_n would represent a discrete-time infinite-state Markov process. Clearly, many of the parameters which Φ_n comprises are unobservable. But, this formulation suggests the applicability of HMMs for modeling HRV.

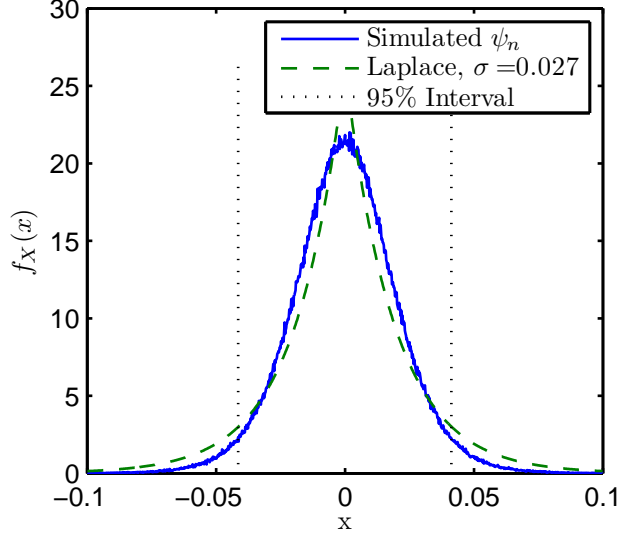


Figure 2.2: Likelihood functions related to the noise term ψ . A likelihood function was computed empirically for ψ . A zero-mean Laplace distribution was fit to the computed likelihood function through selection of standard deviation σ . The parameter was selected which minimized the mean squared error of the resulting likelihood function over the domain $[-0.1, 0.1]$. Finally, 95% confidence interval is indicated for the observed data.

2.2.2 Target Mean as a Random Process

While the term target mean was precise for the single bias walker, for $K > 1$ the term is ambiguous. Consider the threshold $\tau_{0,n} \in \{\tau_{1,n}, \tau_{2,n}, \dots, \tau_{K,n}\}$ where

$$\beta_n < 0, \text{ for } y_n \geq \tau_{0,n} \tag{2.11}$$

$$\beta_n > 0, \text{ for } y_n < \tau_{0,n}. \tag{2.12}$$

By this argument, the $\tau_{0,n}$ approximately represents the target mean of the walker. If ξ_k are constant for all k , $\tau_{0,n}$ is simply the median. The target mean can then be considered a discrete time, continuous valued random process itself. Simulating an outcome for this process using the parameters in Table 2.1, statistics on update times and change in target mean were computed. The results are shown in Fig. 2.3 and Fig. 2.4. These figures suggest that the target mean can be considered as a random

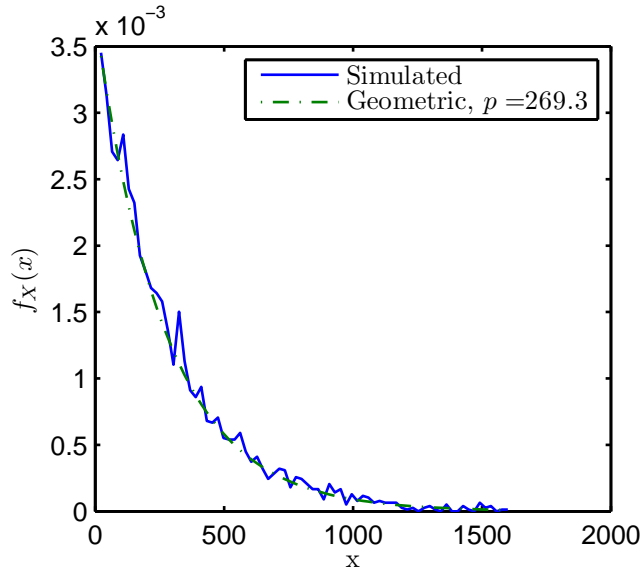


Figure 2.3: Distribution on the number of intervals between updates of the target mean $\tau_{\kappa_M, n}$. The process by which the target mean is determined was simulated using $K = 9$ random processes τ_1, \dots, τ_K with the parameters listed Table 2.1. The update times for each process were geometrically distributed with mean $T = 1000$ intervals. A geometric distribution was fit to the simulated data using the estimated mean $p = 269.3$ intervals.

walk with geometrically distributed update times, where the change in target mean follows a zero-mean Laplace distribution.

Ranking $\tau_{k, n}$ in increasing order using the length K ordered set κ where $\tau_{\kappa_1, n} \leq \tau_{\kappa_2, n} \leq \dots \leq \tau_{\kappa_K, n}$, we can determine the index M such that

$$\tau_{\kappa_M, n} = \tau_{0, n}. \quad (2.13)$$

In this way, it is possible to identify the threshold value $\tau_{\kappa_{M+1}, n}$ and $\tau_{\kappa_{M-1}, n}$ which directly exceeds and is directly less than the target mean τ_0 respectively. Furthermore it is possible to identify the corresponding weighting values ξ_{κ_i} . If we consider the stationary behavior where the K random processes τ_k all remain constant for many time steps, y_n is expected to fluctuate about the mean $\tau_{\kappa_M, n}$ where $\|\beta_n\| = \xi_{\kappa_M}$. However, if many thresholds are clustered about the target mean, or $\tau_{\kappa_{M+1}, n} - \tau_{\kappa_M, n} \approx$

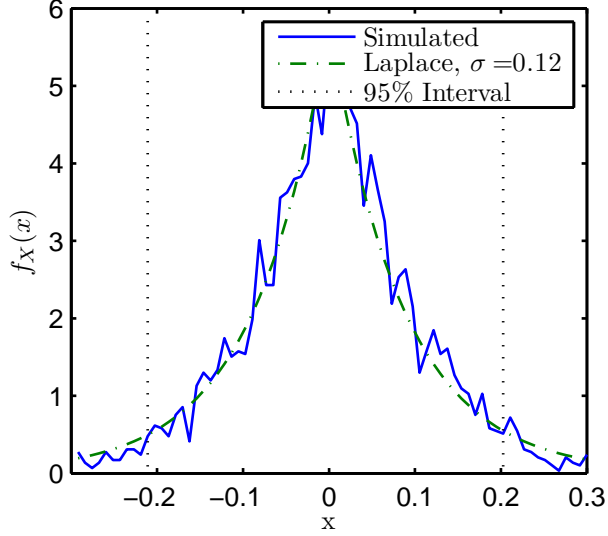


Figure 2.4: Distribution of observed change in target mean after update: $\tau_{\kappa_M, n+1} - \tau_{\kappa_M, n}$. A zero-mean Laplace distribution was fit to the simulated data using the estimated standard deviation $\sigma = 0.12$.

ξ_{κ_M} , the expected magnitude of β_n increases. In either case, the sign of β_n is likely to change with each time step. Qualitatively this suggests that if the spacing of $\tau_{k,n}$ about τ_0 are close, the energy of the high frequency content of the HRV signal will increase. Nonsymmetric distribution of $\tau_{k,n}$ about τ_0 will affect the third moment of the observed y_n . With this model it is possible that two different states with equal target means could generate sequences with very different stationary distributions.

2.2.3 State Changes and Transition Probabilities

We conclude this chapter with a qualitative analysis of state changes of the IGAS model. Consider the case where update times for the process τ_k are geometrically distributed with mean $T = 1000$. Each random process τ_k changes after a time step with likelihood 1×10^{-3} . Given $K = 9$ independent random processes, the likelihood that all τ_k will remain constant after one time step is roughly 0.991. Furthermore, the target mean only changes with the median value $\tau_{\kappa_M, n}$. This suggests that the biased

walker will spend most of its time in one of two modes: 1, transitioning to a new target mean; or 2, in a stationary mode varying about a target mean. Contrasting the maximum value $\max(\|\beta_n\|) = 0.11$ with the standard deviation in change of target mean $\sigma = 0.12$ demonstrated in Fig. 2.4, few time steps are generally required to bias the walker toward the target mean. With an expected $p = 262$ time-steps between update of the target mean, the process is expected to spend most of its time in the second mode. In this mode, β_n will alternate between only a few values to counteract observations of ψ_n . This suggests that the distribution of observed IBI, while the target mean remains constant, could be approximated as stationary with a mean approximately equal to the target mean. After an update to the target mean occurs, the stationary distribution would change. The difference between the mean of the new stationary distribution and the mean of the previous stationary distribution will follow the change in the target mean as demonstrated in Fig. 2.4.

Using this argument, we could define a state by the stationary distribution of observed IBIs while the target mean remains constant. In this way, we could describe an observed sequence of IBIs y using Markov models. The observation y_n would follow the stationary distribution of the state at time n . State transitions would likely occur between states for which the means of the stationary distributions are similar.

Chapter 3

HMMs and Parameter Estimation

A discrete time Markov process is a time-varying random process in which the distribution of future events depend only on the current outcome. For a discrete time random process with realization $x = (x_1, \dots, x_n)$, this principle can be stated as

$$p(x_{n+1} | x_n, x_{n-1}, x_{n-2}, \dots) = p(x_{n+1} | x_n). \quad (3.1)$$

In HMMs, the state of the Markov process is hidden or unknown. The observations are, instead, random variables with distributions conditioned on the state of the “hidden” process.

We choose to model sequences of IBI $y = (y_1, \dots, y_N)$ as a discrete-time, discrete-state HMMs. Each observation is assumed normally distributed according to

$$p(y_n | x_n = m) \sim \mathcal{N}(\mu_m, \sigma_m^2). \quad (3.2)$$

The hidden state sequence $x = (x_1, \dots, x_N)$, $x_n \in \{1, \dots, M\}$, is parameterized by A , representing the transition probabilities

$$A_{i,j} = p(x_{n+1} = j | x_n = i) \quad (3.3)$$

for $i, j \in \{1, 2, \dots, M\}$. Note

$$\sum_{j=1}^M A_{i,j} = 1, \quad \forall i \in \{1, 2, \dots, M\}. \quad (3.4)$$

The conditional mean $\mu = (\mu_1, \dots, \mu_M)$ and variance $\sigma^2 = (\sigma_1^2, \dots, \sigma_M^2)$ of the observations, in addition to A , represent the unknown parameters of the HMM to be estimated. We group these parameters into the ordered set $\theta_M = \{\mu, \sigma^2, A\}$, where the subscript M represents the number of states. In the following, this subscript is omitted when it is not necessary to distinguish model sizes.

In seeking a maximum likelihood (ML) estimate for θ we employ a two-step iterative algorithm. In the first step, we use an estimate for θ , $\hat{\theta}^p$, to compute several marginal likelihood functions. In step two, we use these likelihood functions to compute new estimates $\hat{\theta}^{p+1}$. This structure follows directly from application of the EM algorithm [16]. Step one can be implemented efficiently through application of the forward-backward algorithm. In the following, we first discuss application of the forward-backward algorithm in Sec. 3.1. In Sec. 3.2 we discuss application of the EM algorithm. We conclude with notes on initialization and an algorithm summary in Sec. 3.3.

3.1 Forward-Backward Algorithm

Two likelihood functions are necessary to update estimates of θ : the *a posteriori* probability functions of the hidden states $p(x_n = i | y; \hat{\theta}^p)$ and $p(x_n = j, x_{n-1} = i | y; \hat{\theta}^p)$. Evaluating these functions for all i, j , and n independently would be rather expensive computationally for even modest values N and M . However, these functions can be broken down to the product of several auxiliary functions which depend only on a

subset of the dimensions of x and y . This computation was more recently covered in terms of factor graphs in [17]. However, for brevity, we follow the method outlined in [18].

We define the auxiliary functions

$$\alpha_n(k) = p(y_1, \dots, y_n, x_n = k) \quad (3.5)$$

$$\beta_n(k) = p(y_{n+1}, \dots, y_N | x_n = k). \quad (3.6)$$

We note that $\alpha_i(k)$ can be computed sequentially with increasing i by

$$\alpha_n(k) = \sum_{i=1}^K p(y_1, \dots, y_n, x_n = k, x_{n-1} = i) \quad (3.7)$$

$$= \sum_{i=1}^K p(y_1, \dots, y_{n-1}, x_{n-1} = i) p(y_n, x_n = k | x_{n-1} = i) \quad (3.8)$$

$$= p(y_n | x_n = k) \sum_{i=1}^K \alpha_{n-1}(i) p(x_n = k | x_{n-1} = i). \quad (3.9)$$

Once the state of a Markov process is known, future events are independent of past events. The expression (3.8) is a result of this property.

Similarly, $\beta_n(k)$ can be computed sequentially with decreasing i by

$$\beta_n(k) = \sum_{i=1}^K p(y_{n+1}, \dots, y_N, x_{n+1} = i | x_n = k) \quad (3.10)$$

$$= \sum_{i=1}^K p(y_{n+2}, \dots, y_N | x_{n+1} = i) p(y_{n+1}, x_{n+1} = i | x_n = k) \quad (3.11)$$

$$= \sum_{i=1}^K \beta_{n+1}(k) p(y_{n+1} | x_{n+1} = i) p(x_{n+1} = i | x_n = k). \quad (3.12)$$

Computing the product of the auxiliary functions yields

$$\alpha_n(k)\beta_n(k) = p(y, x_n = k). \quad (3.13)$$

For any n , integrating this product over k provides the likelihood of the observation sequence

$$p(y) = \sum_{i=1}^K \alpha_n(k)\beta_n(k). \quad (3.14)$$

This term can be used to normalize (3.13) such that

$$p(x_n = k | y) = \frac{\alpha_n(k)\beta_n(k)}{\sum_{i=1}^K \alpha_n(k)\beta_n(k)}. \quad (3.15)$$

In seeking an expression for the joint likelihood of sequential states, we note

$$\alpha_n(i) p(x_{n+1} = j | x_n = i)\beta_n(j) p(y_{n+1} | x_{n+1} = j) = p(y, x_n = i, x_{n+1} = j). \quad (3.16)$$

Normalizing this equation by (3.14), we obtain

$$p(x_n = i, x_{n+1} = j | y) = \frac{\alpha_n(i) p(x_{n+1} = j | x_n = i)\beta_n(j) p(y_{n+1} | x_{n+1} = j)}{\sum_{i=1}^K \alpha_n(k)\beta_n(k)}. \quad (3.17)$$

We note that $\alpha_n(k)$ and $\beta_n(k)$ depend on (3.2) and (3.3) which require values for μ , σ^2 and A . It is assumed that these values are taken from the “current” estimate for θ , or $\hat{\theta}^p$. In the next section, these likelihood functions are used to compute $\hat{\theta}^{p+1}$.

3.2 EM Algorithm

It would be computationally infeasible to determine the maximum likelihood estimate (MLE) of θ_M directly from y by brute-force analysis of

$$\hat{\theta}_{\text{ML}} = \arg \max_{\theta_M} \ln f(y; \theta_M). \quad (3.18)$$

However, if the hidden state sequence x were known, finding the MLE of θ would be straight forward. Starting with initial parameter estimates, $\hat{\theta}^0$, we can estimate the probability of any state sequence x . We can then use these probability estimates to improve parameter estimates $\hat{\theta}^1$. The EM algorithm [16] is an iterative algorithm which exploits this concept to converge upon parameter estimates which provide at least a local maximum of the log-likelihood function of incomplete data. In the following, we derive the equations necessary to implement the EM algorithm to parameterize the HMM through selection of θ .

For the problem at hand, we consider the “incomplete data” to include $\{y\}$ and the “complete data” to include $\{x, y\}$. We represent the log-likelihood function of the complete data with $\ln p(x, y; \theta)$, where the likelihood function is parameterized by the values θ . In the expectation step (E-step), we determine the expected value of the log-likelihood function of the complete data

$$U(\theta, \hat{\theta}^p) = E \left[\ln p(x, y; \theta) \mid y; \hat{\theta}^p \right]. \quad (3.19)$$

Due to the iterative nature of the algorithm, we use θ^k to represent the estimated parameters during the k^{th} iteration. This distinguishes the free parameters θ which are used to select estimates for the next iteration, θ^{k+1} , through the maximization

step (M-step)

$$\hat{\theta}^{p+1} = \arg \max_{\theta} U \left(\theta, \hat{\theta}^p \right). \quad (3.20)$$

In this paper, we assume (3.19) to be convex over θ for all values of interest. This significantly reduces the complexity of the M-Step by allowing many dimensions of θ to be optimized independently. The MLE for each μ_i can be obtained by setting the partial derivative of (3.19), with respect to μ_i , equal to zero and solving for μ_i . The values σ_i^2 can be found similarly, although they depend on μ_i . The MLE for A can be found through the method of Lagrange multipliers under the constraint (3.4). If (3.19) is not convex, this algorithm will still converge on estimates providing at least a local maximum. In the following, we first express (3.19) in terms of A , μ , and σ^2 . Then, we derive estimates for the parameters.

By the definition of conditional probability, we can expand (3.19) using

$$U \left(\theta, \hat{\theta}^p \right) = U_A \left(\theta, \hat{\theta}^p \right) + U_B \left(\theta, \hat{\theta}^p \right), \quad (3.21)$$

where

$$U_A \left(\theta, \hat{\theta}^p \right) = E \left[\ln p \left(y \mid x; \theta \right) \mid \mathbf{y}; \hat{\theta}^p \right] \quad (3.22)$$

$$U_B \left(\theta, \hat{\theta}^p \right) = E \left[\ln p \left(x; \theta \right) \mid \mathbf{y}; \hat{\theta}^p \right]. \quad (3.23)$$

Noting that each observation is independent given the “hidden” state,

$$\begin{aligned} U_A \left(\theta, \hat{\theta}^p \right) &= \sum_{n=1}^N E \left[\ln p \left(y_n \mid x_n; \theta \right) \mid \mathbf{y}; \hat{\theta}^p \right] \\ &= \sum_{n=1}^N \sum_{i=1}^M \left[\ln p \left(y_n \mid x_n = i; \theta \right) p \left(x_n = i \mid \mathbf{y}; \hat{\theta}^p \right) \right]. \end{aligned} \quad (3.24)$$

Using the normal distribution for the observations given the state, we have

$$U_A(\theta, \hat{\theta}^p) = \sum_{i=1}^M \sum_{n=1}^N \left[\left(-\frac{1}{2} \ln(2\pi\sigma_i^2) - \frac{(y_n - \mu_i)^2}{2\sigma_i^2} \right) p(x_n = i | y; \hat{\theta}^p) \right]. \quad (3.25)$$

In this equation, μ_i and σ_i^2 can be determined from θ . However, the expression does not depend on A .

Before expanding (3.23), we use the properties of a Markov process to show

$$\ln p(x; \theta) = \ln p(x_1; \theta) + \sum_{n=2}^N \ln p(x_n | x_{n-1}; \theta), \quad (3.26)$$

which leads to

$$U_B(\theta, \hat{\theta}^p) = E \left[\ln p(x_1; \theta) | y; \hat{\theta}^p \right] + \sum_{n=2}^N E \left[\ln p(x_n | x_{n-1}; \theta) | y; \hat{\theta}^p \right] \quad (3.27)$$

and

$$\begin{aligned} U_B(\theta, \hat{\theta}^p) &= \sum_{i=1}^M \ln p(x_1 = i; \theta) p(x_1 = i | y; \hat{\theta}^p) \\ &\quad + \sum_{n=2}^N \sum_{i=1}^M \sum_{j=1}^M \left[\ln p(x_n = j | x_{n-1} = i; \theta) \right. \end{aligned} \quad (3.28)$$

$$\left. p(x_n = j, x_{n-1} = i | y; \hat{\theta}^p) \right].$$

Combining (3.21)(3.25)(3.28), we arrive at

$$\begin{aligned}
U(\theta, \hat{\theta}^p) &= \sum_{i=1}^M \sum_{n=1}^N \left[\left(-\frac{1}{2} \ln(2\pi\sigma_i^2) - \frac{(y_n - \mu_i)^2}{2\sigma_i^2} \right) p(x_n = i | y; \hat{\theta}^p) \right] \\
&\quad + \sum_{n=2}^N \sum_{i=1}^M \sum_{j=1}^M \ln(A_{i,j}) p(x_n = j, x_{n-1} = i | y; \hat{\theta}^p) \\
&\quad + \sum_{i=1}^M \ln p(x_1 = i; \theta) p(x_1 = i | y; \hat{\theta}^p)
\end{aligned} \tag{3.29}$$

At this point we have an expression for (3.19) which depends on A , μ , σ^2 , the initial conditions $p(x_1 = i)$, and the functions $p(x_n = i | y; \hat{\theta}^p)$ and $p(x_n = j, x_{n-1} = i | y; \hat{\theta}^p)$. It should be noted, however, that these function depend on only past estimates $\hat{\theta}^p$, and can be considered constant for the M-Step (3.20). We ignore the initial distribution on the state sequence $p(x_1 = i)$ by assuming its contribution to the change in $U(\theta, \hat{\theta}^p)$ with respect to any element of θ is insignificant.

In seeking the MLE for μ , we differentiate (3.29) with respect to $\mu_i \forall i \in \{1, 2, \dots, M\}$ and set the result equal to zero. Solving for μ_i we arrive at

$$\hat{\mu}_i^{p+1} = \frac{\sum_{n=1}^N y_n p(x_n = i | y; \hat{\theta}^p)}{\sum_{n=1}^N p(x_n = i | y; \hat{\theta}^p)}. \tag{3.30}$$

Similarly, we differentiate (3.29) with respect to $\sigma_i^2 \forall i \in \{1, 2, \dots, M\}$ and set the result equal to zero. Solving for σ_i^2 we arrive at

$$\hat{\sigma}_i^{2p+1} = \frac{\sum_{n=1}^N (y_n - \hat{\mu}_i^{p+1})^2 p(x_n = i | y; \hat{\theta}^p)}{\sum_{n=1}^N p(x_n = i | y; \hat{\theta}^p)}. \tag{3.31}$$

In estimating A we are interested in maximizing (3.29) subject to the constraints (3.4). For solving this constrained optimization problem, we use the method of Lagrange multipliers. We define the objective function ¹

$$f(A) = \sum_{n=2}^N \sum_{i=1}^M \sum_{j=1}^M \ln(A_{i,j}) p(x_n = j, x_{n-1} = i | y; \hat{\theta}^p). \quad (3.32)$$

We define M constraint functions

$$h_i(A) = 1 - \sum_{j=1}^M A_{i,j} \quad (3.33)$$

for $i = 1, 2, \dots, M$. The resulting Lagrangian function is

$$L(A, \boldsymbol{\lambda}) = f(A) - \sum_{i=1}^M \lambda_i h_i(A), \quad (3.34)$$

where λ_i represents the i^{th} element of the M -vector $\boldsymbol{\lambda}$ of Lagrange multipliers. Taking partial derivatives of the Lagrangian function, with respect to each element of A and $\boldsymbol{\lambda}$, and setting the results equal to zero provides $M^2 + M$ equations. In solving this system of equations, all elements of A and $\boldsymbol{\lambda}$ will be determined. For each element of A ,

$$\frac{\partial L(A, \boldsymbol{\lambda})}{\partial A_{i,j}} = \sum_{n=2}^N \frac{p(x_n = j, x_{n-1} = i | y; \hat{\theta}^p)}{A_{i,j}} - \lambda_i. \quad (3.35)$$

Setting this equation equal to zero, solving for $A_{i,j}$, and finding λ_i which satisfy the constraints, we arrive at

$$\hat{A}_{i,j}^{p+1} = \frac{\sum_{n=2}^N p(x_n = j, x_{n-1} = i | y; \hat{\theta}^p)}{\sum_{j=1}^M \sum_{n=2}^N p(x_n = j, x_{n-1} = i | y; \hat{\theta}^p)}. \quad (3.36)$$

¹The objective function used for the method of Lagrange multipliers includes only those terms from (3.29) which depend on A . In subsequent steps, we are only interested in the partial derivative of $f(A)$ with respect to the elements of A . Therefore, this simplifies the expression without changing the results.

Equations (3.30), (3.31), and (3.36) provide rather intuitive expressions for their parameter estimates.

3.3 Summary

We summarize the algorithm described in the previous sections in three steps: initialization, expectation, and maximization.

Initialization

The algorithm described in the previous sections require an initial estimate $\hat{\theta}_M^0$. In applications where (3.19) is not convex over θ , results may vary significantly with selection of the initial estimate $\hat{\theta}_M^0$. Methods for selecting these initial estimates are contrasted in Chapter 4.

E-Step

(3.9) and (3.12) require $\alpha_1(k)$ and $\beta_M(k)$ to be initialized at each iteration of the algorithm. We set

$$\alpha_1(k) = p(y_1 | x_1 = m) \tag{3.37}$$

$$\beta_M(k) = 1 \tag{3.38}$$

for all value $k \in \{1, \dots, M\}$. After computing (3.9) and (3.12), the functions (3.14), (3.15), and (3.17) are evaluated for all $n = \{1, \dots, N\}$ and $m = \{1, \dots, M\}$.

M-Step

New estimates $\hat{\theta}_M^{p+1}$ are computed by (3.30), (3.31), and (3.36). Return to *E-Step*.

Conclusion

The *E-Step* and *M-Step* are repeated until a specific stopping criterion is reached.

The process could continue until the change in the score function (3.14) or change in parameter estimates after each iteration becomes negligible.

Chapter 4

Model Selection

How many states are necessary: a seemingly innocuous question of how to select M . In chapter 3, algorithms were presented to improve estimates of model parameters. However, selection of M was not addressed. In many publications, model size selection is paid little attention [6]. Sometimes it is selected to optimize performance related to a particular problem[10]. In this chapter we focus on incorporating model size as a parameter itself, determined by the data.

Using (3.14) we can compute $p(y; \theta)$ as a score function quantifying the overall fit of model parameters θ to y . Employing slightly different notation for the parameter estimates, we define the ML estimate

$$\hat{\theta}_{\text{ML}}(y, M) = \underset{\theta_M}{\operatorname{argmax}} \ln p(y | \theta_M). \quad (4.1)$$

Using these estimates, we define

$$g(y, M) = \ln p\left(y | \hat{\theta}_{\text{ML}}(y, M)\right). \quad (4.2)$$

This function is monotonically non-decreasing with M . Stated another way, it is always possible to add a state without decreasing the score function such that

$$\ln p(y | \theta_{M+1}) \geq \ln p(y | \theta_M). \quad (4.3)$$

This reasoning would suggest that the largest model possible can best model an observation sequence. However, our interest is in modeling an underlying random process as opposed to a specific outcome. If new data are introduced, re-parameterizing the model on the extended observation sequence may significantly affect the parameter estimates. In general, the expected change in $\hat{\theta}$, due to the inclusion of new data, increases with M in what is often referred to as “overfitting” the data. This suggests an optimization problem which takes into account both the score function (3.14) and model size (or model complexity). To establish a cost function, we require a penalty term.

4.1 Bayesian Information Criterion

One classic approximation applicable to selection of model size is

$$\ln p(y, M) \approx \ln p\left(y | \hat{\theta}_{\text{MAP}}(y, M)\right) - \frac{d}{2} \ln(N). \quad (4.4)$$

Requiring optimization of this expression over M is referred to as the Bayesian Information Criterion (BIC) [19]. Here N represents the number of observations in y , and d is the model description length. In our case $d = M^2 + M$, where the model parameter A can be represented with $M(M - 1)$ elements under constraint (3.4), and the parameters μ and σ^2 together require $2M$ elements. In this way, θ_M can be represented as a vector of length d .

Note that (4.4) assumes knowledge of maximum *a posteriori* estimation of the parameters $\hat{\theta}_{\text{MAP}}(y, M)$. It also requires consistency with increasing N . When considering long IBI sequences (50,000), parameter consistency with increasing N is a reasonable assumption. However, the proposed algorithm for parameter selection does not guarantee ML estimates of the parameters, and no prior distribution on $\hat{\theta}$ has been proposed. In the next section we contrast score functions while increasing model size. For some initial parameter selection methods, the BIC appears applicable.

4.2 Initial Parameter Selection

In chapter 3, an algorithm was presented to improve estimates of model parameters. The algorithm seeks to select parameters $\hat{\theta}_M^0$ which optimize the score function (3.14). However, this algorithm may provide suboptimal estimates when the score function is not convex. In our problem convexity does not hold, making it difficult to quantify the benefit of increasing model order M .

One difficulty in optimizing nonconvex functions is distinguishing global from local maximums. However, in our case (4.3) can be used to identify parameter selections which do not represent a global maximum. While a global maximum cannot be confirmed, if a set of parameters θ_{M-1} exist such that $\ln p(y | \theta_M) \leq \ln p(y | \theta_{M-1})$, θ_M does not represent a ML estimate.

Calculating the score function over long sequences of observations, even for models with few states, becomes computationally expensive. For this reason, we consider nonlinear search algorithms to be inaccessible. Instead, we focus on initial parameter selections submitted to the EM algorithm. In the following we discuss three methods for initial parameter selection, contrasting performance.

4.2.1 Method 1: Basic

As a first pass at initial parameter selection, μ was selected such that its elements were evenly spaced between the maximum and minimum values of y . The values $\hat{\sigma}^2$ were selected such that

$$\sigma_i^2 = \left(\frac{\mu_2 - \mu_1}{8} \right)^2 \quad (4.5)$$

for all values $i \in \{1, \dots, M\}$. Similarly, $A_{i,j} = 1/M$ for all $i, j \in \{1, \dots, M\}$. Contrasting initial results from this method demonstrated the susceptibility of the EM algorithm to converge on local maximum. Scoring each set of parameters with increasing M , through (4.2), revealed several local maximums for all sequences of training data y used.

4.2.2 Method 2: Increase

In an effort to reliably increase model score with model size, a method was devised in which states would be added to the models incrementally. This method starts with parameterizing a single state model. Increasing model size then included selection of parameters μ_{M+1} and σ_{M+1}^2 and expanding the transition matrix A . Qualitatively, this approach seeks to identify μ_{M+1} which is most likely to increase (3.14). Adding nontrivial values to A can decrease (3.14). However, if the initial value A suggests that it is unlikely to enter the new state, the EM algorithm will largely ignore the new state parameters. Therefore, the values σ_{M+1}^2 and updates to A must be selected to minimize a reduction in (3.14), without driving the likelihood of entering the new state to zero.

Consider the stationary distribution of one IBI, y_0 , for the current model

$$p(y_0; \theta_M) = \sum_{m=1}^M p(x_0 = m) p(y_0 | x_0 = m), \quad (4.6)$$

where $p(x_0 = m)$ represents the stationary distribution of the states. This stationary distribution on the states can be determined from the eigenvector of A with the corresponding eigenvalue 1. Seeking to minimize the difference between the stationary distribution and the distribution of the data, $f_Y(y)$, we select

$$\mu_{M+1} = \max_{y_0} f_Y(y_0) - p(y_0; \theta_M). \quad (4.7)$$

The value $\sigma_{M+1}^2 = (3/128)^2$ was selected as the original ECG recording was sampled at 128 Hz. The transition matrix A was updated such that

$$p(x_{n+1} = m | x_n = i) \approx 1 \times 10^{-4} \quad \text{for } i = 1, \dots, M + 1 \quad (4.8)$$

$$p(x_{n+1} = i | x_n = m) \approx \frac{1}{M + 1} \quad \text{for } i = 1, \dots, M. \quad (4.9)$$

In this case, the approximations are due to the necessary normalization of A under constraint (3.4). For some IBI sequences, this approach significantly improved the score function (4.2). However, this method would enter conditions in which it would assign sequential states similar means. While designed to match the stationary distribution of the model to the distribution of the observations, it proved counterproductive as demonstrated in Fig. 4.3.

4.2.3 Method 3: Distribute

In method 2, many states would be assigned to a small range of IBI. While this increased the overall score function in some cases, the resulting stationary distribution

did not match the distribution of the data. In method 1, the spacing of the initial means changes significantly with increasing model size. For many observed sequences, the distribution is peaked. It is critical that the trained model place states with target means near the peaks of the distribution. By uniformly distributing the state means, Method 1 did not provide consistency with increasing model size.

As a third method for assigning initial parameter values, we distribute the means according to the distribution of the data. Using $F_Y(y_0)$ to denote the cumulative distribution of the data, we select values for μ such that

$$F_Y(\mu_i) = \frac{i}{M+1} \quad (4.10)$$

for $i = 1, \dots, M$. Variance parameters were selected such that

$$\sigma_i^2 = (\max(\mu_i - \mu_{i-1}, \mu_{i+1} - \mu_i))^2. \quad (4.11)$$

The transition matrix was initialized to a constant $A = \frac{1}{M}$.

This method consistently resulted in models yielding higher scores compared with methods 1 and 2. Additionally, this method significantly reduced the variance between the stationary distribution of the resulting models represented the distribution of the observed data when contrasted with the resulting models using Method 2. In the few cases method 2 yielded models which exceeded the score of the model trained using method 3, the stationary distribution of the method 2 model varied significantly from the distribution of the data. These results are demonstrated in Fig. 4.1 contrasting results for two separate individuals. Resulting parameters for model size 24 are shown for the two individuals in Fig. 4.2 and Fig. 4.3

4.3 Results

Using training data obtained from the “Normal Sinus Rhythm RR Interval Database” with PhysioBank [13], we sought to find the optimal model size satisfying (4.4). For each individual, we first removed intervals less than 0.2s which were assumed double-counted beats. We also segmented the data at missed beats, where the interval duration exceeded the moving average by a factor of 1.6. For each segment (3.15), (3.17) and (3.14) were computed independently. However, for (3.30), (3.31), and (3.36), the summations were performed over all sets. After removing ectopic beats, the average number of IBIs used for training was 53,667 with a standard deviation of 5,445 intervals. Training each model, for each size, 400 iterations of the EM algorithm were computed to establish $\hat{\theta}_M$ for each sequence.

We estimated model parameters for each individual using three methods proposed in this chapter. If the identified model parameters represented the ML estimates, (4.3) would hold true for each neighboring model size. In application, many models did not follow this inequality, demonstrating ML estimates were not identified. Focusing on model sizes 19-25, we contrasted the parameters selected using each of the three initialization methods over the entire sample population of 54 sequences. For each initialization method, it was possible to identify up to 324 degenerate models which did not satisfy (4.3). The percentage of degenerate models identified in this range are listed as a percentage by initialization method in Table 4.1. For each model size and each individual, we also determined the initialization method which identified parameters yielding the highest score using (3.14). Out of the 378 models trained, the percentage of highest-scoring models determined by each initialization method are also listed in Table 4.1. Note that this column sums to 1 as the initialization methods are only contrasted with each other. It is possible that no method identified the ML estimate.

Table 4.1: Model Performance by Initialization Method

	Degenerate Models	Models with Highest Score
Method 1: Basic	25.0%	0.0%
Method 2: Increase	35.5%	13.8%
Method 3: Distribute	3.1%	86.2%

In general, the $g(y, M)$ increased monotonically with M , using method 3, suggesting the BIC is meaningful. Using the BIC to determine optimal model size, a histogram demonstrating the relative frequency of model size is shown in Fig. 4.4. Overall, the sample mean for all model sizes is 25.64, but Fig. 4.4 demonstrates an appreciable difference in model complexity from one individual to another.

We provided a method for quantitatively measuring the fitness of the model to the observed sequence Fig. 4.2(c) and Fig. 4.3(c) by contrasting the distribution of the data with the stationary distribution of the model. As these models are capable of generating IBI sequences, we can also contrast simulated data with the original training data. We trained 26-state models on full 24-hour recordings to capture one cycle of the circadian rhythm. Using these models, we simulated one IBI sequence for every trained model. In Fig. 4.5, we show IBI sequences from the training data and the resulting simulated sequence over different timescales. For comparison, we also included simulated data generated by the IAGS model. However, the IAGS model was parameterized using the values in Table 2.1. Parameters were not specifically selected to match the training data.

Fig. 4.5 qualitatively demonstrates the ability of the model to capture variability observed in the training data. The simulated data appears to both capture HRV and preserve Homeostasis, as does the IAGS model. Unfortunately, many extremely low IBI durations are encountered in the training data. These IBIs are physiologically unlikely and suggests problems with the original data set. While the simulated data suggests that the models capture these pathologies in the training data, it is likely

that these features are characteristics of the data capture and processing methods rather than an underlying physiological process.

As a quantitative comparison of the data, we estimate the autocorrelation function and autocorrelation coefficients of IBI sequences over time. Estimates are shown for individuals 30 and 19 in Fig.4.6 and Fig.4.7, respectively. Plots are shown for training data, simulated data (parameters estimated from the training data), and data generated using the IAGS model.

In estimating autocorrelation functions, IBI signals were divided into segments of 400 sample windows with a 50% overlap. The middle 200 samples were then time shifted for $\tau \in [-100, \dots, 100]$ samples. Each point of the autocorrelation function was determined by averaging the results over the 200 samples. Each column represents one estimated autocorrelation function from one window. The x-axis demonstrates change in the autocorrelation function over time. Note, the color scale is consistent across (a), (c), and (e) in Fig. 4.6.

The estimated autocorrelation functions of the training data for individuals 30 and 19, shown in Fig. 4.6 (a) and Fig. 4.7 (a), respectively, both demonstrates a vertical striping pattern. More specifically, a wide red stripe toward the center of the autocorrelation plots demonstrate a long period of highly correlated IBIs in the middle of the training data. This likely corresponds to the sleep cycle of the individual. This dramatic, yet sustained, change in correlation is not exhibited by the simulated data.

Autocorrelation coefficients were estimated using the same windowing method used for estimating the autocorrelation functions. The mean and variance were assumed consistent across the entire 400 sample window, although they were allowed to change between windows. Average correlation coefficients across all time windows are shown in Fig. 4.8 and Fig. 4.9, for individuals 30 and 19 respectively.

Contrasting the autocorrelation coefficients in both Fig. 4.6 and Fig. 4.7, all plots exhibit similar variability over time. Again, the estimates in the 30,000 - 60,000 interval range for the training data exhibit a slightly different pattern than those observed in the simulated data. In this case, a patterns of negative coefficients emerges. Contrasting the average autocorrelation coefficients in Fig. 4.8 and Fig. 4.9, the simulated data using HMM shows an increased autocorrelation for $5 < |\tau| < 65$ compared with the original data sequence. Conversely, the IAGS model exhibits a decreased correlation over these time shifts τ .

In medical applications, the power spectral density (PSD) of IBI sequences is often estimated. Variability over different frequency bands has been linked to specific physiological functions and influences [20]. We contrast the estimated PSD of the training data with simulated sequences in Fig. 4.10 and Fig. 4.11 for individuals 30 and 19, respectively.

Computing the DFT of a sequence of IBI will estimate signal energy at frequencies of biological time. However, many influences of HRV such as 0.25 Hz respiratory influences and the so-called 10s rhythm are not expressed in biological time. One method from converting the DFT of an IBI sequence to Hz is interpret the sampling rate as the average IBI duration [21]. In this way, the normalized frequencies at which the DFT is computed can be divided by the average IBI duration. When estimating the PSD of sequences of IBI over long time periods, the average IBI duration is likely to change. To estimate the PSD in Hz, we divided each signal into non-overlapping segments of 2048 samples. Following Barlett's method, we computed the periodogram over each window. We also computed the average IBI duration over each window. This allowed scaling of the frequency range of each periodogram in Hz. An estimated PSD in Hz was obtained by averaging these values using a simple binning method with 1024 frequencies from 0 to 0.5 Hz.

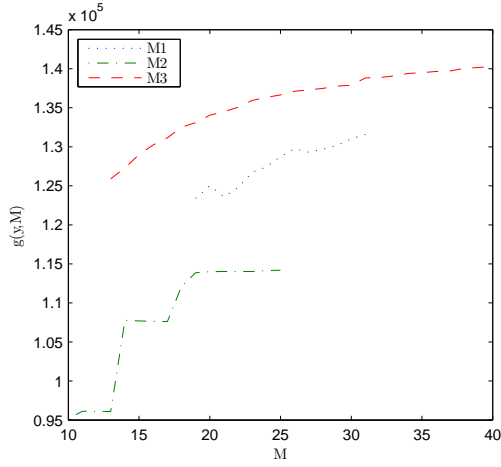
The estimated PSD of the simulated data for individual 30 appears to capture the PSD of the training data as demonstrated by Fig. 4.10. However, for individual 19, the PSD estimates are very different for the training data and simulated data. The training data exhibits strong respiratory influences at approximately 0.25 Hz. Also, there is a significant change in PSD at approximately 0.1 Hz. While both of these features are relatively common across subjects, these models simply do not capture these features.

4.4 Discussion

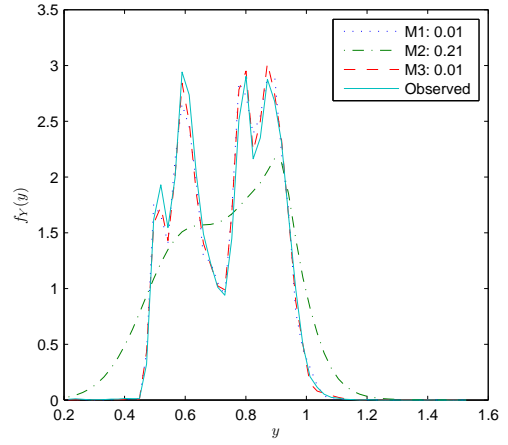
In this chapter we presented initialization methods and a model sizing criterion which can be used to select a model for a particular sequence of IBIs. Model fitness was demonstrated in Fig. 4.1(b) and Fig. 4.1(d) through similarities in the distribution of the training data and the stationary distribution of the respective models. Qualitatively, these models are capable of generating IBI sequences that exhibit HRV while preserving homeostasis as demonstrated by Fig. 4.5. These models also demonstrate the $1/f$ power spectrum which is characteristic of IBI sequences. These properties, which were used to validate the IAGS model initially, are satisfied.

We have also exposed several features of the training data which were not captured in the model. In particular, the 10s rhythm and 0.25 Hz effects due to respiration are not present in the simulated data as shown in Fig. 4.11. Both the HMMs and IAGS models have no reference to absolute time; all state transitions occur in biological time. In Fig. 4.10, the models demonstrate some capacity to fit the contour of the PSD of the data sequence. However, when a strong 0.25 Hz component is present in the training data, overall performance of the model suffers as demonstrated in Fig. 4.11. Performance may be improved by modeling 0.25 Hz influences separately.

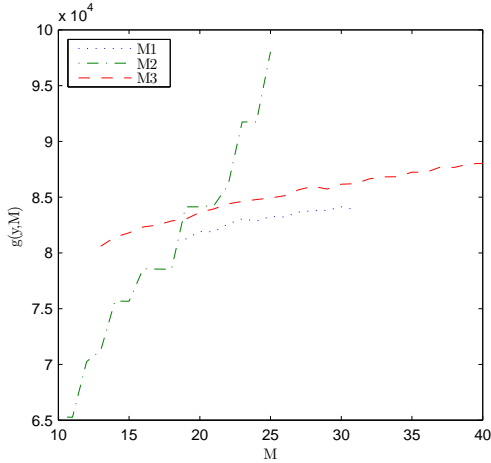
In estimating model parameters, Fig. 4.8 and Fig. 4.9 demonstrate the inability to accurately model correlations over $5 < |\tau| < 65$. Note that neither the simulated data from the HMMs nor the IAGS model appear to represent the correlation coefficients faithfully. Simulated data generally demonstrated increased variability at frequencies below 0.03 Hz. This is evident in Fig.4.10 and Fig. 4.11. Plots of the autocorrelation function in Fig.4.6 (c) and Fig. 4.7 (c) do not demonstrate long-range correlations exhibited by the training data. However, periodic state changes due to, for example, the circadian rhythm may require a deterministic component.



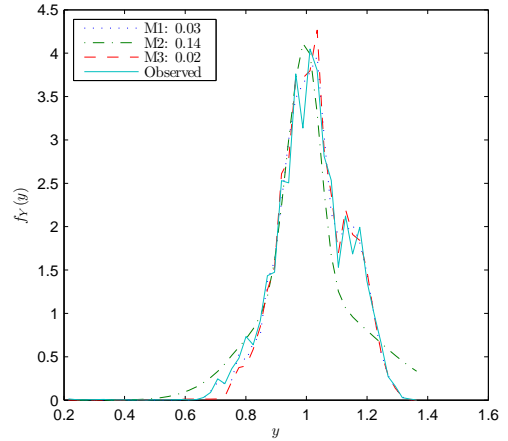
(a) Individual 38, Score vs. Model Size



(b) Individual 38, Stationary Distributions

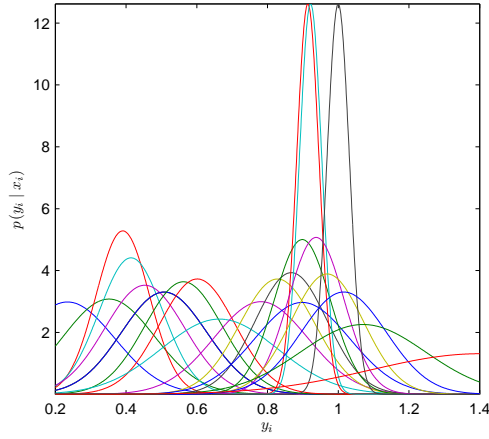


(c) Individual 33, Score vs. Model Size

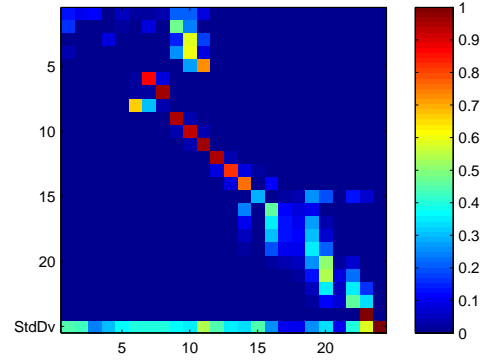


(d) Individual 33, Stationary Distributions

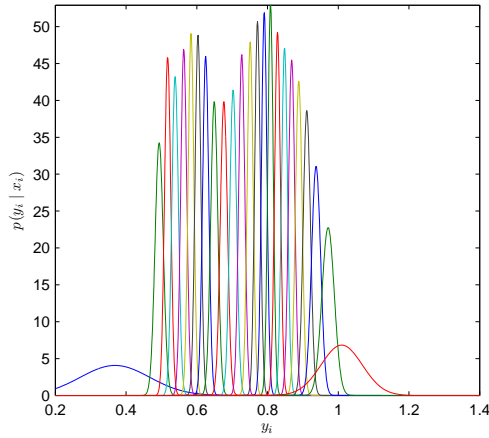
Figure 4.1: Performance of initial parameter selection methods. (a) and (c) show the resulting score function 4.2 for individuals 38 and 33 respectively. Scores are computed for many model sizes using the three proposed initialization techniques. (b) and (d) show the stationary distributions for the 24 state models estimated using each of the three initialization methods for individuals 38 and 33 respectively. Stationary distributions are contrasted with the distribution of the data. For each model, the mean squared error was estimated contrasting the stationary distribution from the parameters and the distribution of the data. These estimates are shown in the legends of (b) and (d). The mean squared error was consistently larger for initialization method 2. For most individuals, and most model sizes, initialization method 3 provided the highest score $g(y, M)$. However, (c) demonstrates that for some model sizes, for individual 38, the score for the models resulting from method 2 exceed those of the others.



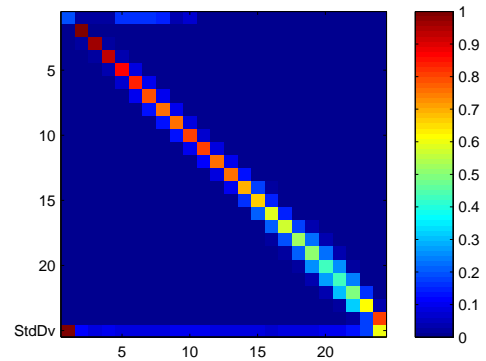
(a) Method 2, *Posterior* distributions given state



(b) Method 2, Transition matrix

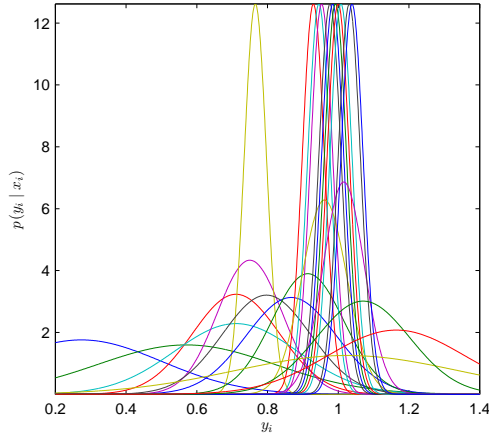


(c) Method 3, *Posterior* distributions given state

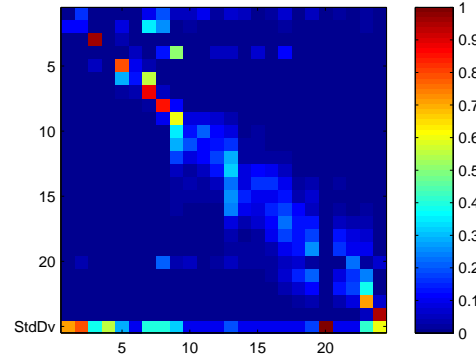


(d) Method 3, Transition matrix

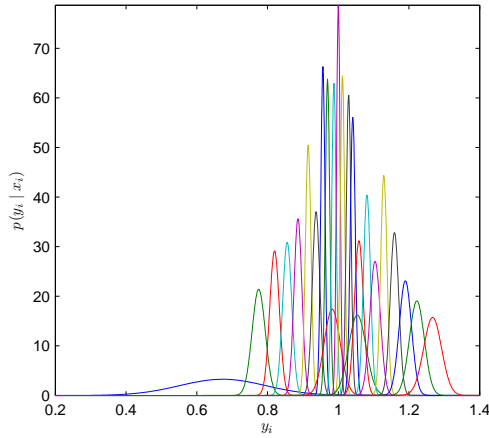
Figure 4.2: Effects of initial parameter selection methods on a 24-state model for individual 38. *Posterior* distributions for IBI duration given state information are shown for method 2 and method 3 in (a) and (c) respectively. Transition matrices are shown for method 2 and method 3 in (b) and (d) respectively.



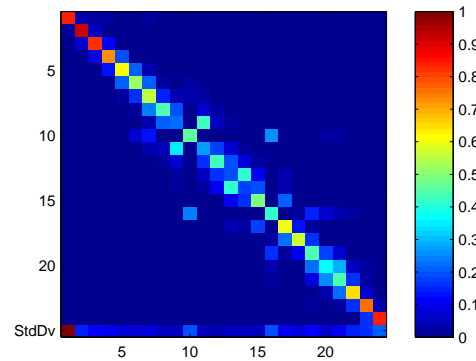
(a) Method 2, *Posterior* distributions given state



(b) Method 2, Transition matrix



(c) Method 3, *Posterior* distributions given state



(d) Method 3, Transition matrix

Figure 4.3: Effects of initial parameter selection methods on a 24-state model for individual 33. *Posterior* distributions for IBI duration given state information are shown for method 2 and method 3 in (a) and (c) respectively. Transition matrices are shown for method 2 and method 3 in (b) and (d) respectively.

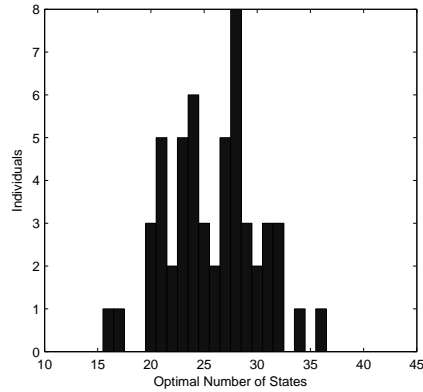


Figure 4.4: Histogram depicting the number of sequences by their optimal model size determined using the BIC (4.4).

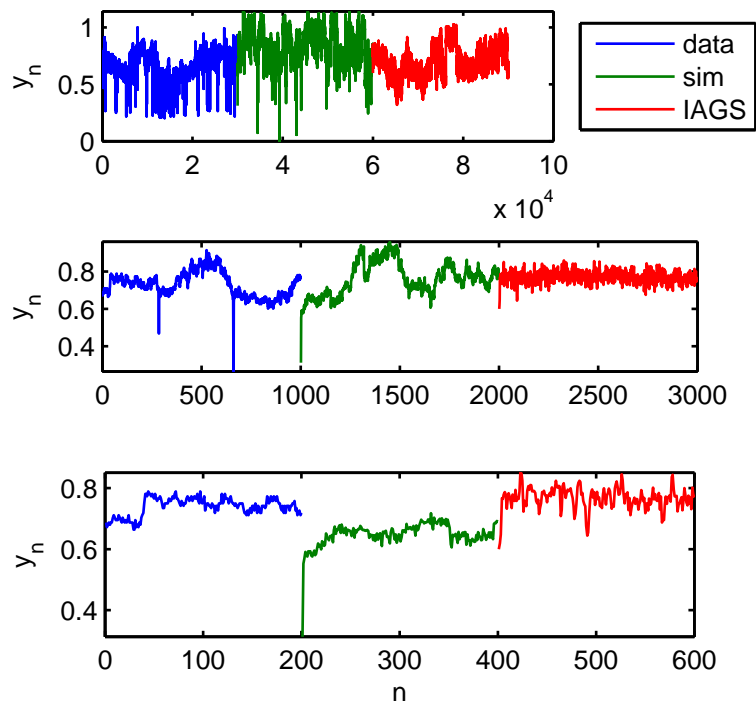


Figure 4.5: IBI sequences over different time scales based on training data from individual 30. Three sequences are plotted consecutively: the training data, simulated data generated using model parameters estimated from the training data, and simulated data from the IAGS model using parameters found in Table 2.1.

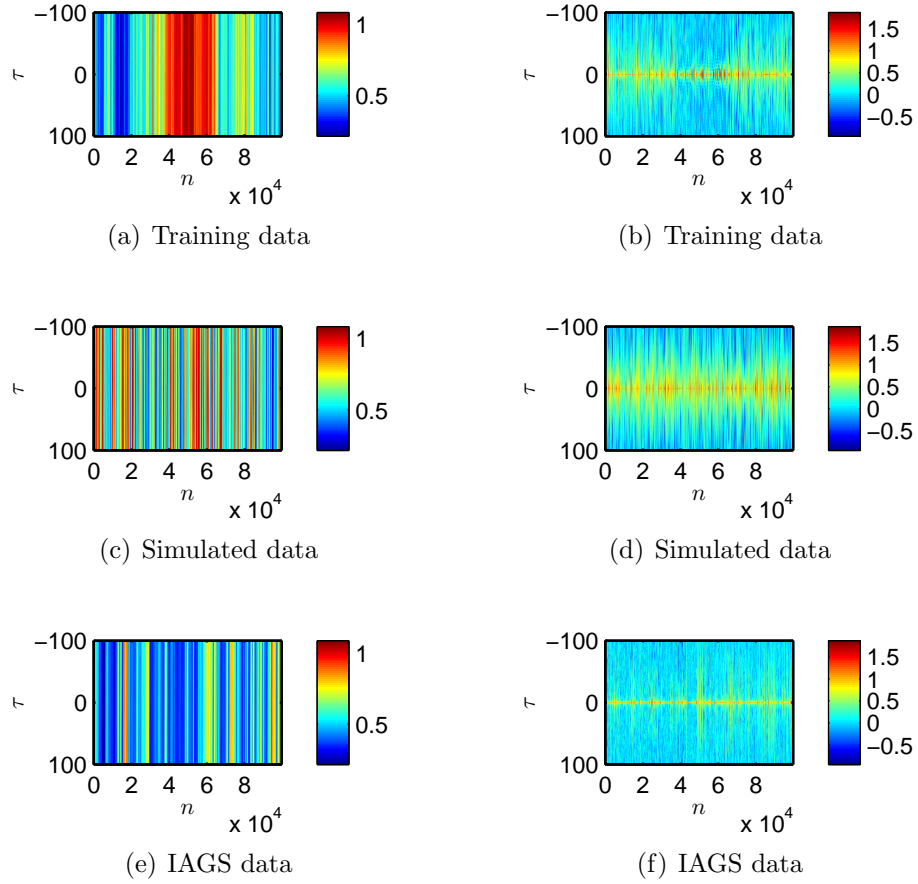


Figure 4.6: Estimated autocorrelation functions and autocorrelation coefficients for sequences of IBIs. Estimated autocorrelation functions for training data, simulated data and data generated using the IAGS model are shown in (a), (c), and (e) respectively. In each plots, each column represents an autocorrelation function, where τ represents the time shift. The x-axis demonstrates the change in the autocorrelation function over time. Similarly, (b), (d), and (f) show the change in estimated autocorrelation coefficients over time.

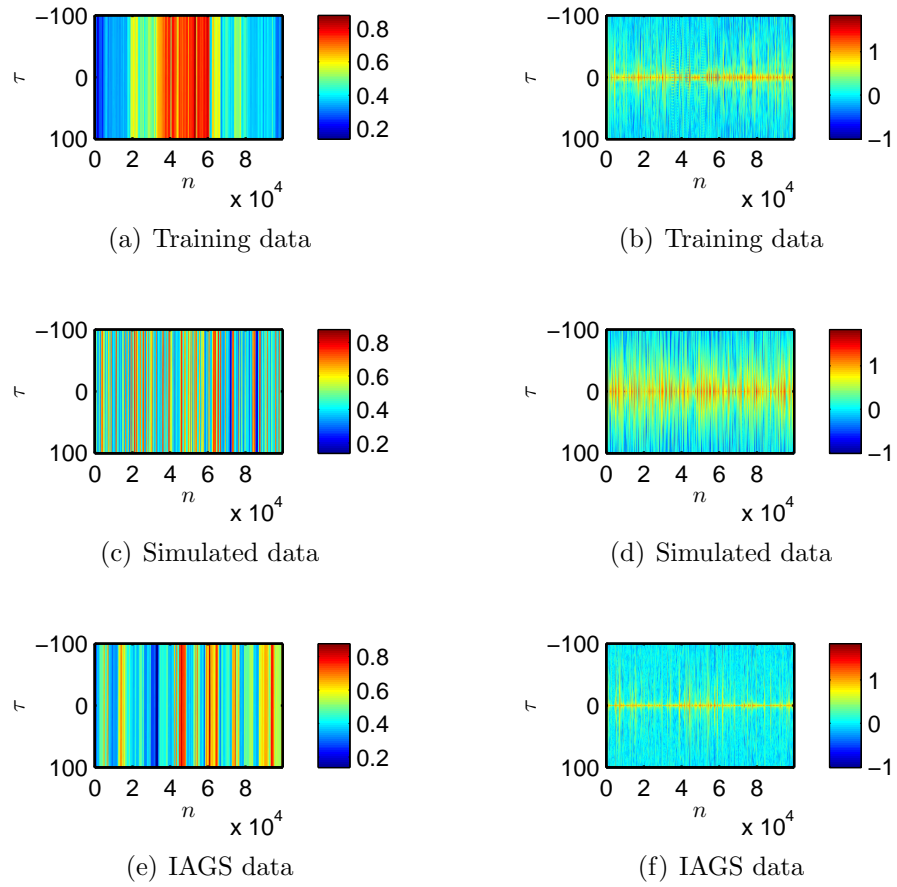


Figure 4.7: Autocorrelation function and autocorrelation coefficients of IBI sequences from individual 19

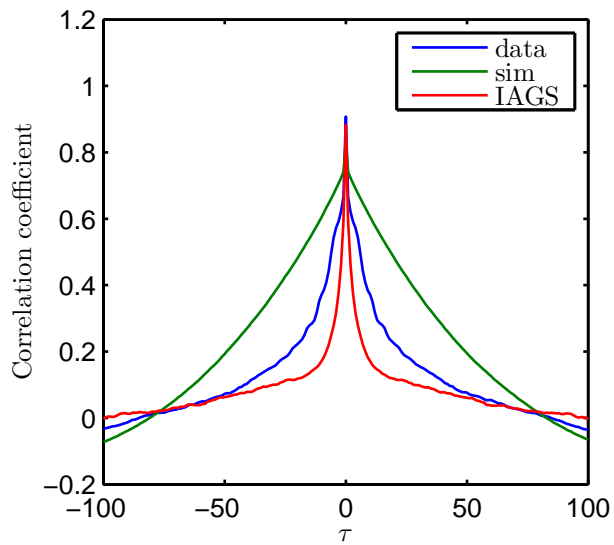


Figure 4.8: Average correlation coefficients for individual 30.

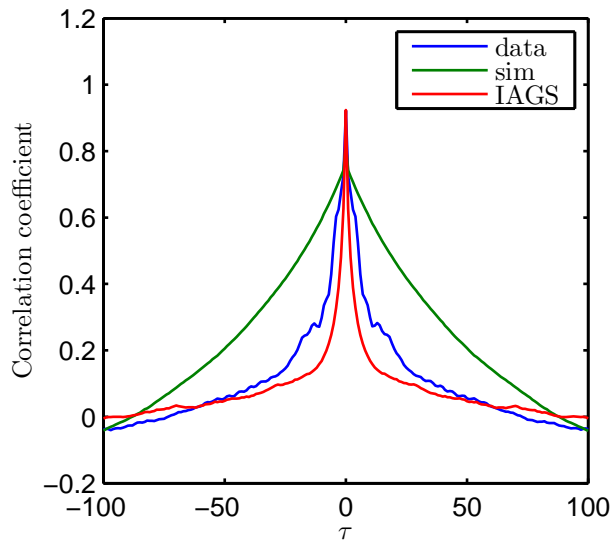


Figure 4.9: Average correlation coefficients for individual 19.

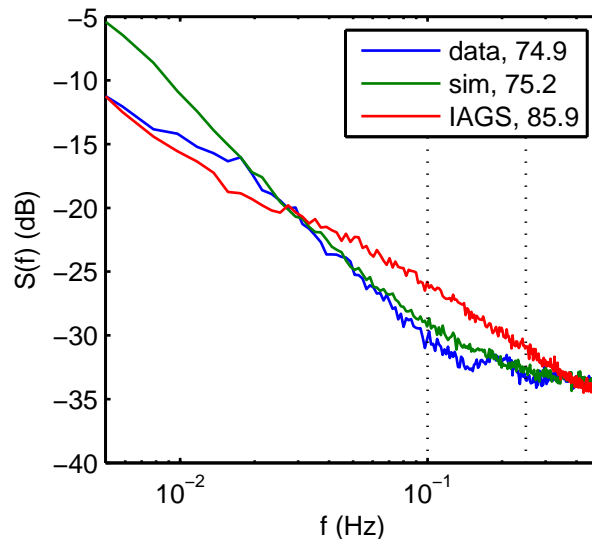


Figure 4.10: Estimated PSD of IBI sequences. Estimates were obtained by computing length 2048 periodograms. Estimates were obtained in Hz by scaling the frequency range by the average IBI duration for each periodogram. Dotted lines indicate 0.1 Hz and 0.25 Hz. Average heart rates are shown in the legend for each sequence.

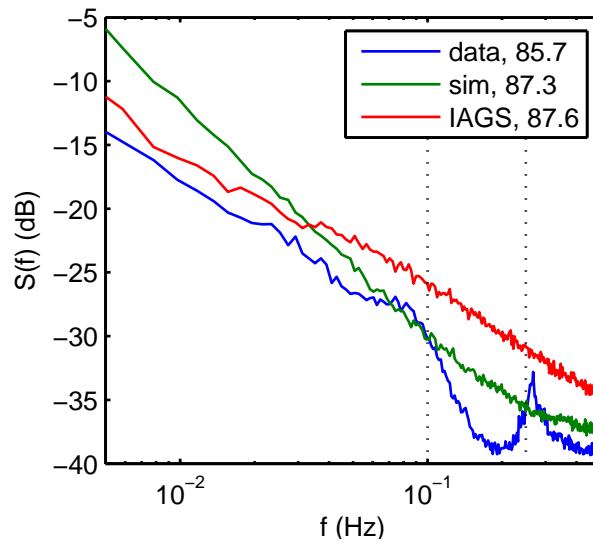


Figure 4.11: Power spectral density of IBI sequences from individual 19

Chapter 5

Biometric Capability

As a use-case for modeling HRV, we consider confirming the identity of an individual based on a brief observation sequence. We focus strictly on the confirmation or rejection of a presented identity. In this case, there are two types of errors: failing to authenticate the correct individual and incorrectly authenticating an impostor. The likelihood of each of these outcomes is commonly abbreviated false non-match rate (FNMR) and false match rate (FMR) respectively. We first present two criterion for matching a model to a sequence in Sec. 5.1. Results of applying these criteria are presented in Sec. 5.2.

5.1 Recognition Algorithm

We use two separate methods for matching an identity to a testing sequence. In the first case, individual scores are contrasted against a single, generic, model H_0 . In our case, the H_0 model was trained using approximately 50,000 IBIs each, from 52 individuals. We denote the score of a testing sequence y using (3.14) and the trained model H_0 as

$$s_{H_0} = \ln p(y; \theta_{H_0}). \quad (5.1)$$

We similarly score the testing sequence using the asserted identify, l , such that

$$s_l = \ln p(y; \theta_l). \quad (5.2)$$

We then choose to match the l^{th} model, θ_l to the observation sequence y if the inequality

$$s_l - s_{H_0} > T \quad (5.3)$$

is true for a predetermined threshold T . This tunable threshold can be used to balance FMR with FNMR.

For the second test, we assume a library of L models representing unique individuals $\theta = \{\theta_1, \dots, \theta_L\}$, where θ_l represents the model for the l^{th} individual. Over the entire library, we define the sample mean and variance as

$$\mu_s = \frac{1}{L} \sum_{l=1}^L s_l \quad (5.4)$$

$$\sigma_s^2 = \frac{1}{L-1} \sum_{l=1}^L (s_l - \mu_s)^2. \quad (5.5)$$

We then choose to match the l^{th} model, θ_l to the observation sequence y if the inequality

$$\frac{s_l - \mu_s}{\sigma_s} > \gamma \quad (5.6)$$

is satisfied, which is commonly referred to as a χ^2 test.

5.2 Results

Using data obtained from the “Normal Sinus Rhythm RR Interval Database” with PhysioBank [13], we parameterized a 26-state model for each individual. For each

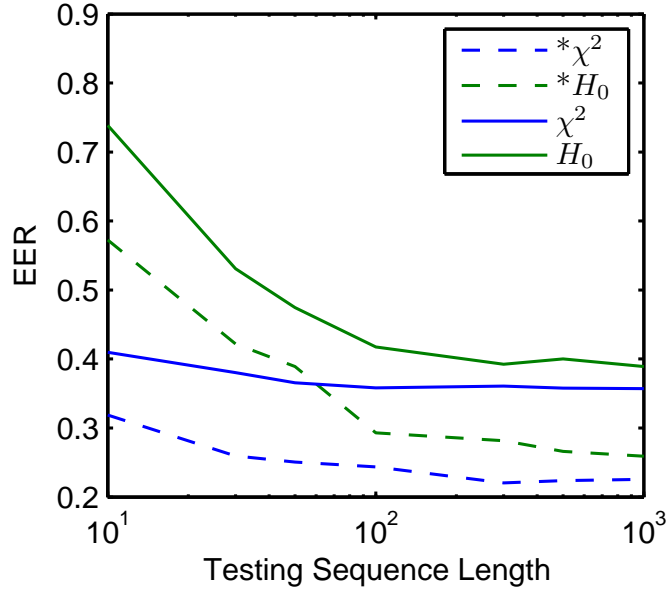


Figure 5.1: Measured equal error rates as a function of test sequence duration for both decision criteria. The χ^2 test clearly out-performs the H_0 test for testing segments of all sizes. In the legend, * indicate results where testing data was drawn from the training data, or intra-session testing. Legend entries without a * indicate that models were not trained on the testing data, or cross-session testing.

individual, approximately 24 hours of IBI data were recorded. The first half (approximately 12 hours) was used for training. Testing sequences were then drawn from the 24 hour data. A set of libraries of testing data were established, each containing sequences of different length. For each library, 30 equally-spaced, equal-duration, testing segments of IBI data were identified for each individual. The lengths of training sequences identified ranged from 10 - 1000 IBIs.

Each testing sequence was scored by each model and an identity confirmation decision was made according to (5.3) and (5.6). For each decision criteria, the threshold yielding an equal error rate was computed. The resulting equal error rates are shown in Fig. 5.1 and plotted for various lengths of testing data.

Focusing on the χ^2 decision metric, Fig. 5.2 shows the error rate performances for FMR and FNMR as a function of the threshold γ . Testing data was reliably matched

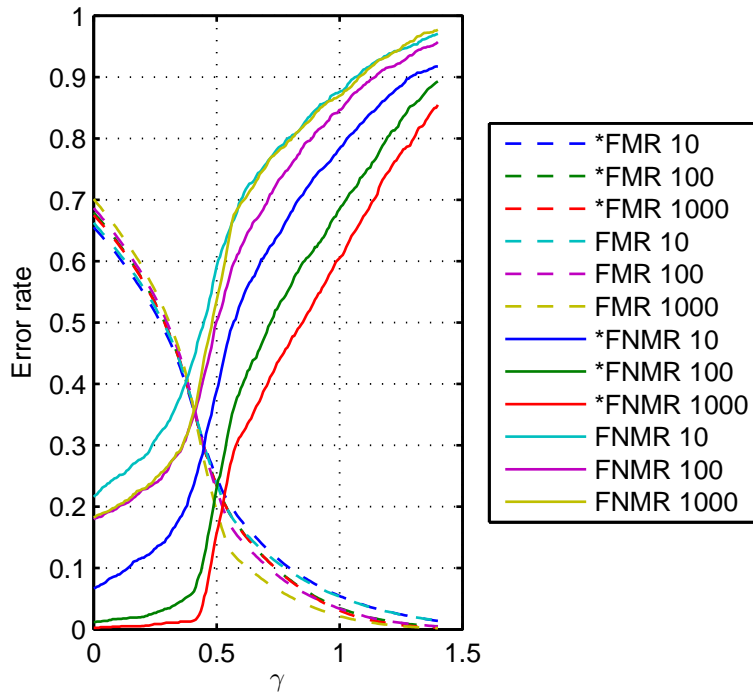


Figure 5.2: Measured error rates as a function of threshold γ . In testing, models could be matched to segments of 300 IBIs with an equal error rate below 0.37. However, for FMR below 0.01, FNMR exceeds 0.93. In the legend, * indicate results where testing data was drawn from the training data. Legend entries without a * indicate that models were not trained on the testing data.

to its corresponding model with an equal error rate of roughly 0.37. Enforcing a FMR below 0.01, the FNMR increases to roughly 0.93.

5.3 Discussion

The results in this chapter do not present a compelling argument for the use of HRV as a biometric. While some individuality was captured, the reliability is far below the minimum requirements for even a screening application. The data used in this application lacked many desirable features. It is possible that performance would improve with new data.

In the previous chapter, abnormally brief IBI sequences were identified in the training data (see Fig. 4.5). While these beats were likely an artifact of the data acquisition or processing methods, they were captured by the models. It is unclear whether these features added to the individuality of the models. However, their presence likely affected performance.

The data from PhysioBank did not contain any time reference. For this reason, we trained models on long sequences (12 hours) of IBI data. Using the Markovian assumption we considered each individual would enter, and return to, specific states throughout a day. Over long time scales, we could train over a larger sampling of the possible states to which an individual is likely to return. In this way, the models would become more robust to new testing data. However, HRV is highly susceptible to external influence. Physical activity, positioning, and stress can all greatly affect both mean IBI duration and variability. Without information as to these external influences, we could not account for them while training models. It is likely that our models captured environmental influences as much as individual physiological differences between training sequences.

In application of an identity confirmation system, there is likely consistency of environmental influence. Not only would the acquisition of 12-hour recordings of IBI data be impractical, it's utility is questionable for such application. A future test could be designed where each subject is conditioned by similar environmental influence during the acquisition of both testing and training data.

Chapter 6

Conclusion

HRV is susceptible to many influences. With the relatively low frequency of observations, IBI sequences represent a challenging signal to model. In fact, few beat-to-beat models are presented in the literature.

In this thesis, we built upon physiologically justified models to demonstrate the applicability of HMMs for modeling HRV. Using this form, we were able to represent specific observation sequences, selecting parameters on an individualistic basis. These models were then able to generate data statistically similar to the original data.

Additionally, we explored methods for initial parameter selection and model complexity. One method reliably performed better than the others across all individuals and model sizes. However, it did not identify ML estimates of the parameters without exception. One area for additional research is optimization techniques for ensuring ML estimates of the parameters.

The models presented are applicable for both analyzing and generating IBI sequences. We contrasted simulated IBI sequences with their original training sequences to measure model performance. In doing so, we identified several features of the training sequences which were not reproduced in the simulated data. Most notably, the 0.25 Hz respiratory influences and 10s rhythm were not present in the simulated data.

We evaluated the utility of HMM trained on HRV for use as a biometric. Admittedly, this is a rather lofty goal given the nature of HRV. However, these models could easily be applied in other test scenarios. These models could be applied to other detection and inference problems where HRV signals are available.

The data used in this study lacked many desirable features. The sampling rate for the original ECG was only 128 Hz. Higher resolution of IBI data might improve individuality in the models. Additionally, no time reference was provided with the data. It would be interesting to contrast results training and testing on similar phases of the sleep cycle. Also, it would be interesting to note consistency in HRV over longer periods between training and testing.

References

- [1] M. Tarvainen, P. Ranta-aho, and P. Karjalainen, “An advanced detrending method with application to HRV analysis,” *Biomedical Engineering, IEEE Transactions on*, vol. 49, no. 2, pp. 172–175, 2002.
- [2] D. A. Litvack, T. F. Oberlander, L. H. Carney, and J. P. Saul, “Time and frequency domain methods for heart rate variability analysis: A methodological comparison,” *Psychophysiology*, vol. 32, no. 5, pp. 492–504, 1995.
- [3] T. F. of the European Society of Cardiology the North American Society of Pacing Electrophysiology, “Heart rate variability : Standards of measurement, physiological interpretation, and clinical use,” *Circulation*, vol. 93, no. 5, pp. 1043–1065, Mar. 1996.
- [4] R. W. DeBoer, J. M. Karemaker, and J. Strackee, “Hemodynamic fluctuations and baroreflex sensitivity in humans: a beat-to-beat model,” *American Journal of Physiology - Heart and Circulatory Physiology*, vol. 253, no. 3, pp. H680 –H689, 1987.
- [5] T. Kuusela, T. Shepherd, and J. Hietarinta, “Stochastic model for heart-rate fluctuations,” *Phys. Rev. E*, vol. 67, no. 6, p. 061904, Jun 2003.
- [6] P. C. Ivanov, L. A. N. Amaral, A. L. Goldberger, and H. E. Stanley, “Stochastic feedback and the regulation of biological rhythms,” *Europhysics Letters (EPL)*, vol. 43, no. 4, pp. 363–368, 1998.
- [7] L. Rabiner, “A tutorial on hidden markov models and selected applications in speech recognition,” *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257 –286, feb 1989.
- [8] D. Coast, R. Stern, G. Cano, and S. Briller, “An approach to cardiac arrhythmia analysis using hidden markov models,” *Biomedical Engineering, IEEE Transactions on*, vol. 37, no. 9, pp. 826–836, 1990.
- [9] A. D. Kaplan, J. A. O’Sullivan, E. J. Sirevaag, S. D. Kristjansson, P.-H. Lai, and J. W. Rohrbaugh, “Hidden state dynamics in laser doppler vibrometry measurements of the carotid pulse under resting conditions,” in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, 2010, pp. 5273–5276.

- [10] V. Baier, M. Baumert, P. Caminal, M. Vallverdu, R. Faber, and A. Voss, “Hidden markov models based on symbolic dynamics for statistical modeling of cardiovascular control in hypertensive pregnancy disorders,” *Biomedical Engineering, IEEE Transactions on*, vol. 53, no. 1, pp. 140–143, 2006.
- [11] R. Silipo, G. Deco, R. Vergassola, and C. Gremigni, “A characterization of HRV’s nonlinear hidden dynamics by means of markov models,” *Biomedical Engineering, IEEE Transactions on*, vol. 46, no. 8, pp. 978–986, 1999.
- [12] R. Silipo, G. Deco, R. Vergassola, and H. Bartsch, “Dynamics extraction in multivariate biomedical time series,” *Biological Cybernetics*, vol. 79, no. 1, pp. 15–27, 1998.
- [13] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, “PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals,” *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000 (June 13).
- [14] M. Chen, J. A. O’Sullivan, N. Singla, E. J. Sirevaag, S. D. Kristjansson, P.-H. Lai, A. D. Kaplan, and J. W. Rohrbaugh, “Laser doppler vibrometry measures of physiological function: Evaluation of biometric capabilities,” *Information Forensics and Security, IEEE Transactions on*, vol. 5, no. 3, pp. 449–460, 2010.
- [15] L. A. N. Amaral, A. L. Goldberger, P. C. Ivanov, and H. Stanley, “Modeling heart rate variability by stochastic feedback,” *Computer Physics Communications*, vol. 121–122, pp. 126–128, 1999.
- [16] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, Jan. 1977.
- [17] F. Kschischang, B. Frey, and H. Loeliger, “Factor graphs and the sum-product algorithm,” *Information Theory, IEEE Transactions on*, vol. 47, no. 2, pp. 498–519, 2001.
- [18] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, “Optimal decoding of linear codes for minimizing symbol error rate (Corresp.),” *Information Theory, IEEE Transactions on*, vol. 20, no. 2, pp. 284–287, 1974.
- [19] A. D. Lanterman, “Schwarz, wallace, and rissanen: Intertwining themes in theories of model selection,” *International Statistical Review*, vol. 69, no. 2, pp. 185–212, 2001.
- [20] D. Cysarz, D. von Bonin, P. Brachmann, S. Buetler, F. Edelhäuser, K. Laederach-Hofmann, and P. Heusser, “Day-to-night time differences in the relationship between cardiorespiratory coordination and heart rate variability,” *Physiological Measurement*, vol. 29, no. 11, pp. 1281–1291, Nov. 2008.

- [21] R. W. DeBoer, J. M. Karemaker, and J. Strackee, “Comparing spectra of a series of point events particularly for heart rate variability data,” *IEEE Transactions on Bio-Medical Engineering*, vol. 31, no. 4, pp. 384–387, Apr. 1984, PMID: 6745974.

Vita

Michael R. Walker II

- Date of Birth** May 9, 1982
- Place of Birth** Muskegon, Michigan
- Degrees** B.Sc., Electrical Engineering May 2004
Cedarville University, Cedarville, Ohio
- M.Sc., Electrical Engineering May 2011
Washington University in St. Louis, St. Louis, Missouri
- Affiliations** Member, IEEE
- Current Position** Communication System Engineer, Advanced Systems R&D
Aclara Power-Line Systems, Inc., St. Louis, MO.
- Publications** David W. Rieken and **Michael R. Walker II** "*Ultra Low Frequency Power-Line Communications Using a Resonator Circuit,*" in *IEEE Transactions on Smart Grid*, vol.2, no.1, pp.41-50, March 2011
- David W. Rieken and **Michael R. Walker II** "*Distance Effects in Low-Frequency Power Line Communications,*" in *Power Line Comm. and Its Applications, 2010 IEEE Int. Symposium on*, 2010, pp. 22-27.
- May 2011