SYSTEM IDENTIFICATION OF DYNAMICAL MODELS FOR SIGNALS RELATED TO THE HUMAN USE OF ETHANOL

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Ipek Ozil

August 2014

SYSTEM IDENTIFICATION OF DYNAMICAL MODELS FOR SIGNALS RELATED TO THE HUMAN USE OF ETHANOL

Ipek Ozil, Ph.D.

Cornell University 2014

The influence of genetics on the risk for alcoholism is a major theme in alcoholism research. Genetic research depends on phenotyping. However, accurate phenotyping of human use of alcohol is difficult. What are essentially video games with alcohol as a reward are being used to examine human use of alcohol in controlled circumstances. A generative model (containing parameters with unknown values) of a simple game involving a progressive work paradigm is described along with the associated point-process signal processing that allows system identification of the model. The system is demonstrated on human subject data. The same human subject playing the game under different circumstances, e.g., with and without a psychoactive drug, is assigned different parameter values. Potential meanings of the different parameter values are described.

Physiologically based pharmacokinetic models have been used to describe the distribution and elimination of ethanol after intravenous administration. Mathematically, these models are nonlinear ordinary differential equations. These equations are solved and optimized, by using their gradient, to formulate and refine parameter identification and control strategies. The Hessian information is then used to design an optimal input to the system.

BIOGRAPHICAL SKETCH

Ipek Ozil was born in Istanbul, Turkey on December 8, 1986. She graduated from her high school, Robert College, with high honors in 2004. She was awarded a Presidential and an International Fellowship at Worcester Polytechnic Institute, where she earned her Bachelor of Science degree in Electrical and Computer Engineering in 2004. She was a member of the Engineering Honor Society, Tau Beta Pi, and a member of the Electrical and Computer Engineering Honor Society, Eta Kappa Nu.

She received her Master of Science degree from Cornell University in 2012. At Cornell University, she worked as a graduate research assistant to Dr. Peter Doerschuk and worked on system identification and modeling. She served as a Teaching Assistant in 2008 and as a Teaching Assistant Fellow 2009-2012.

She worked as an Associate at JP Morgan in the Fixed Income Strategy division in 2012 and in 2013 she joined the Chief Investment Office in JP Morgan as a Senior Portfolio Manager. She resides in Brooklyn, New York.

I dedicate this thesis to my grams, Turkan Sokullu, who has always been the rock in my life. I am sorry for taking five years away from our time together.

This is for you.

ACKNOWLEDGEMENTS

I am most grateful to my advisor, Dr. Peter Doerschuk, who supported me through the years and who has constantly given me his time, his ideas and who was generous to let me explore my own passions. I have learned so much from him that I will carry with me for the rest of my life.

I am incredibly grateful to my committee members, Dr. Lang Tong and Dr. Kevin Tang, who have been more than supportive and who have offered me their expertise.

Nothing would have been possible without the support of my parents, Eralp and Yaprak Ozil, who have always pushed me to work a little harder and who made me believe that I am capable of achieving anything I set my mind to. I know that it hasn't been easy but we made it through.

I also would like to thank our very close friends Munire and Mahir Arikol, who have endured the stresses and the joys with me. Thank you for being there.

My friends, you know who you are. Thank you for being with me through the good and the bad times. Thank you for always listening to me and supporting me. Belin, thank you for all the food that you cooked when I was missing home - you made me survive.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Ethanol is a naturally produced drug used by humans for thousands of years because of its psychoactive properties. Beneficial when used in moderation [30], excessive use of ethanol can be devastating. Of people who use ethanol, nearly eight percent will become addicted during the course of their life, and about a third of them will die of complications attributable to the addiction [19, 23]. Alcohol has its psychoactive effects by acting on the brain directly. Measuring characteristics of these effects, such as brain alcohol concentration, directly is extremely difficult. Therefore, mathematical models that relate the input of alcohol to the brain alcohol concentration of a person play an important role in alcoholism research.

A major theme in alcoholism research is the influence of genetics on risk of addiction, e.g., [30]. Accurate phenotyping is critical for genetic studies. However, phenotyping human use of alcohol in the community is difficult. One approach is to replace community use by use in controlled laboratory settings, e.g., [27]. There are two different components to phenotyping subjects in laboratory settings: designing an experiment and modelling the brain alcohol concentration to give standardized doses.

O'Connor and colleagues have developed several experiments, that are essentially video games, where the reward is alcohol based on a progressive work paradigm in order to measure how much effort a human subject is willing to invest in order to get alcohol. In a progressive work game, the first dose of alcohol (given by intravenous infusion) requires relatively little work but the amount of work required to get successive doses progressively increases. Using the ideas of [24], the dose can be normalized to achieve brain alcohol concentration changes that are the same for all subjects independent of age, sex, weight, *etc*.

The most simple such game is considered. In particular, the work is just the pressing of a button where the progressive increase of work is a progressive increase in the number of button presses required in order to receive the next dose of alcohol. The game is made more complicated by the fact that any button presses beyond three presses per minute are not counted and the subject is provided with feedback at every button press that indicates whether that particular button press was counted.

The manner in which the subject plays the game of the previous paragraph is exactly described by the sequence of times at which the subject presses the button. Therefore, a point process model, e.g., [29], whose arrivals are the button-pressing times provides a generative model of the subject's performance.

In order to correctly model the brain alcohol concentration, physiologically-based pharmacokinetic models (PBPK) can be used. In previous work [24] the authors describe ordinary differential equation models with two and three states which have different number of parameters. These models have been widely used with parameter values that are linear transformations of morphometrics (e.g., gender, age, height, weight) [20]. A key aspect of this work is that the models do not represent populations of subjects but rather individual subjects. This greatly constrains the complexity of the model and the type and amount of data that is available to identify the parameters in the model. However, it is essential for the clinical and research applications of these models.

The pharmacokinetic data available on a subject comes from the following type of experiment: A solution of ethanol in normal saline is injected by intravenous infusion. The volume flow of the infusion is known. Periodically, at intervals of roughly two minutes, arterial ethanol concentration is measured by performing a breath-analyser measurement. A correctly performed measurement requires the subject to produce end-expiratory air. Especially when the subject is drunk, not every measurement is correctly

performed. When a measurement is correctly performed, it is well documented that the measurement is equivalent to a arterial ethanol concentration measurement, which is one of the variables in the PBPK model [24].

The work done in this thesis focuses on system identification on the progressive games and the PBPK models. The thesis is organized in the following manner. The first section deals with the progressive game experiments. A point process model in which arrivals of the process are the button-pressing times and which has parameters with unknown values is described for the first game Section 3.1. System identification for the model is equivalent to estimation of the parameter values and a maximum likelihood approach is described Section 3.2. An algorithm for computing the maximum likelihood estimator is described Section 3.3. The algorithm of Section 3.3 is not exact and so numerical simulations are described which demonstrate the performance of the algorithm Section 3.4. The algorithm is demonstrated on two paired sets of human subject data Section 3.5. The statistical tests done by using the Hessian information are described in that section. A different approach to the EM algorithm by using Monte Carlo Markov Chain simulations discussed in Section 3.6. The results of MCMC simulations are shown in Section 3.7. The second section of the thesis is about the PBPK models and their analysis. The PBPK model and the measurement model associated with the experiments are described in detail Section 4.1. The maximum likelihood estimator and the algorithm are shown Section 4.3. Performance of the estimator is calculated via Hessian matrices Section 4.6. Experimental results are shown in Section 4.7. The Hessian information is then used to design an optimal input in Section 4.8.

CHAPTER 2

LITERATURE REVIEW

2.1 Statistics

Statistical analysis is needed to correctly interpret any observational data and is therefore important in the study of alcoholism. For this reason a short review of relevant aspects of statistics will be made here. Statistics can be studied from many points of view and these points of view are included in this chapter: (i) the study of populations, (ii) as the study of variation and (iii) as the study of methods of the reduction of data.

The concept of statistics as the study of variation is the natural outcome of viewing the subject as the study of populations. A population of individuals in which each individual is in all respects identical is completely described by a description of any one individual, together with the number in the group. The populations which are the object of statistical study always display variation in one or more respects. The study of variation leads immediately to the concept of a frequency distribution. Frequency distributions are of various kinds; the number of classes in which the population is distributed may be finite or infinite; in the case of quantitative variates, the intervals which separate the classes may be finite or infinitesimal. The idea of a frequency distribution is applicable either to populations which are finite in number, or to infinite populations, but it is more usefully and more simply applied to the latter. A finite population can only be divided in certain limited ratios, and cannot in any case exhibit continuous variation. Moreover, in most cases only an infinite population can exhibit accurately, and in their true proportion, the whole of the possibilities arising from the causes actually at work, and which we wish to study.

The third aspect of statistics deals mainly with the reduction of the bulk of the data. Results of an experimental study should be expressed in a relatively small number of numerical values which can be understood by the community of researchers. An exact description of the data usually requires far more data than are required to answer the questions posed by the experiments. In consequence, much of the information supplied by any body of data is irrrelevant. It is the object of the statistical processes employed in the reduction of data to exclude this irrelevant information, and to isolate the whole of the relevant information contained in the data. [12]

For the purpose of estimating any parameter, such as the center of a normal distribution from a sample of n independent measurements, it is usually possible to invent many statistics, such as the arithmetic mean, or the median, etc., each of which has for larger n an error variance that falls off inversely. But for large samples of a fixed size the error variance of these different statistics will generally be different. Consequently, special importance belongs to a smaller group of statistics, the error distributions of which tend to the normal distribution with the least possible variance as the sample is increased. These statistics form a sub group of consistent statistics with a special value and these are known as efficient statistics. According to [12], an efficient statistic can in all cases be found by the Method of Maximum Likelihood; that is, by choosing statistics so that the estimated population should be that for which the likelihood is greatest.

Fisher showed that in certain special cases, namely the location and scale models, all of the information in the sample is recoverable by using an appropriately conditioned sampling distribution for the maximum likelihood estimator, called ancillary statistics. Fisher in [11] showed that when these log likelihood functions are differentiable successive portions of the information loss may be recovered by using as ancillary statistics, in addition to the maximum likelihood estimate, the second and higher differential coeffi-

cients at the maximum. The function of the ancillary statistic is analogous to providing a true weight for the value of the estimate.

Efron and Hinkley in [10] assume that inference is accomplished by attaching a standard error, which is the standard deviation of the difference between the unknown true value and the estimate provided by the statistic, to the maximum likelihood estimate, and based on Fishers remarks they use a conditional variance approximation based on the observed second derivative of the log likelihood function evaluated at the maximum likelihood estimate. Many have argued in favour of the variance estimator , where $I(x)$ is the observed information, i.e., minus the second derivative of the log likelihood function at given data x.

2.2 State-Space Model and Kalman Filter

State-space models with a variety of state spaces are widely used in engineering. Such models are defined by an evolution equation for the state and an observation equation describing what can be measured. State space models are sometimes referred to as Hidden Markov Models.

A simple form of state space model is the linear finite dimension model shown in Equations 2.1 - 2.2 where $x(t)$ is the state, A, B, C and D are matrices with appropriate dimensions, $y(t)$ is the observed variable and $w(t)$ is the input variable which drives the evolution of the state [15]:

$$x(t+1) = Ax(t) + Bw(t) \tag{2.1}$$

$$y(t) = Cx(t) + Dv(t). \tag{2.2}$$

When the input $w(t)$ to the state evolution equation and the error $z(t)$ in the measurement equation are both stochastic processes then an important question is to estimate the conditional probability density function (pdf) on the state $x(t)$ given the past measurements $y(\tau)$ $0 < \tau < t$. This is called the filtering problem. If $v(t)$ and $w(t)$ are jointly Gaussian then the Kalman filter solves the filtering problem by providing the conditional mean and the conditional covariance of the pdf, which is a complete characteristic of the pdf since the pdf is Gaussian. Kalman filter provides a recursive solution to the problem, in the sense that each estimate is computed from a previous estimate and a new input to the system [14]. For computer computation a discrete time formulation is often used. The equations analogous to Eqs (2.1) and (2.2) are Eqs. (2.3) - (2.4).

$$x_{k+1} = F_{k+1,k}x_k + w_k \tag{2.3}$$

$$y_k = H_k x_k + v_k, \tag{2.4}$$

where w_k is the process noise, an additive white Gaussian noise with zero mean and v_k is the observation noise, also an additive white Gaussian. These are assumed to be uncorrelated, with known individual covariance matrices. $F_{k+1,k}$ is the transition matrix, which takes state x_k into state x_{k+1}. H_k is the observation matrix [14].

Based upon the state space model, Kalman filter equations for state space estimation and covariance estimation are given in Equations 2.6 - 2.8:

$$\hat{x}_k^- = F_{k,k-1}x_{k-1} \tag{2.5}$$

$$P_k^- = F_{k,k-1}P_{k-1}F_{k,k-1}^T + Q_{k-1} \tag{2.6}$$

$$\hat{x_k} = \hat{x}_k^- + G_k(y_k - H_k\hat{x}_k^-) \tag{2.7}$$

$$P_k = (I - G_k H_k)P_k^-, \tag{2.8}$$

where P_k is the error covariance matrix and G_k is the Kalman gain matrix given by $P_k^- H_k^T[H_k P_k^- H_k^T + R_k]^{-1}$ and Q_k and R_k are the covariance matrices of the Gaussian state and measurement noises [14].

2.3 Expectation-Maximization Algorithm

An important approach to parameter estimation in data analysis is maximum likelihood(ML) estimation. Θ is the set of feasible parameter values, $\theta \in \Theta$ is the specific parameter vector. The conditional pdf on the data y given the parameter vector is $p(y|\theta)$. The likelihood is defined by $L(\theta;y) = p(y|\theta)$. The maximum likelihood estimator is defined as the θ value that makes the likelihood achieve its maximum value[18]:

$$\hat{\theta}_{ML} = argmax_{\theta \in \Theta} L(\theta;y) \tag{2.9}$$

Usually, one looks for the maximum likelihood estimator in the log-likelihood instead of the likelihood itself because the logarithm operation converts a product in the likelihood to a summation, making it more convenient to differentiate:

$$\hat{\theta}_{ML} = argmax_{\theta} \log(L(\theta;y)) \tag{2.10}$$

The computation of an ML estimate is the solution of a maximization problem. In only a few cases is it possible to give an explict solution in terms of the values of the measurement. When an explicit solution is not available, a numerical algorithm is necessary. An important class of such algorithms is expectation maximization algorithms. Expectation-maximization algorithms are based on the idea that an ML problem can be made easier to solve if additional data is measured. While the additional data is actually unavailable, the idea of having such data leads to an iterative algorithm in which each iteration has two steps, an expectation step involving the expectation over the additional data and a maximization step involving maximization over the parameter value. The iteration continues until a convergence criteria is met. There are two sets of data: observed data, denoted y, and missing data, denoted x, that is observed through y [8].

The steps of the EM algorithm are as follows, where p denotes the iterations of the

algorithm and θ denotes the parameter vector.

- E-step: Estimate the complete data sufficient statistic, denoted $t(x)$ by finding:

$$t^{(p)} = E(t(x)|y, \theta^{(p)}) \tag{2.11}$$

- M-step: Determine $\theta^{(p+1)}$ as the solution of the equations:

$$E(t(x)|\theta) = t^{(p)} \tag{2.12}$$

In the most general case of the EM algorithm, one can define a function, denoted Q, as:

$$Q(\theta^{'}|\theta) = E(\log f(x|\theta^{'}|y, \theta). \tag{2.13}$$

Then the EM algorithm can be thought of as computing $Q(\theta|\theta^{(p)})$ in the E-step and then choosing a value of $\theta^{(p+1)}$ that maximizes $Q(\theta|\theta^{(p)})$.

2.4 Physiological Pharmacokinetic Models

One of the important aspects of alcoholism research is the relationship between ethanol input and consequent time trajectory of ethanol in varius tissues and brain's response to such input. Mathematical models have been developed to explain such phenomenon. A subject's response to alcohol input depends on many factors such as gender, height, weight and previous history with alcohol. There are two different kinds of models: physiological and phenomenological. Phenomenological models describe the time series of ethanol concentration in terms of generic compartments where the number of compartments is the order of the linear dynamical system [6]. Physiological models describe the time series in terms of anatomical structures, i.e liver, and physiological principles[24].

The two types of models have different applications, where phenomenological models might be used to model large populations and certain physiological models are used to model single subjects. In accordance with our work with the Indiana School of Medicine, where the tests are performed on single subjects, this thesis will focus on the physiological models. Most physiological models contain more than one anatomical structure.

Effects of alcohol on subjects vary widely. For instance, individuals vary 3- to 4-fold in metabolic rates and systematic concentrations and 2- to 3- fold in subjective and physiologic responses to the drug [25]. This variability makes it difficult to study the effects of alcohol, especially for people who are at risk for alcoholism.

Based upon the paper [17], colleagues the Indiana University School of Medicine lead by Dr. S.J. O'Connor, have developed a three-compartment PBPK model, where the distribution kinetics are based on the cardiac outflow distribution to the liver, tissue and vasculature compartments. The goal in developing such a model was to control the ethanol trajectory in the brain of each individual. This was later used to model the breath alcohol content (BrAC). There is also a two-compartment PBPK model developed by O'Connor and colleagues, which excludes liver from the equations, due to the fact that the liver volume is hard to measure. It was believed that the PBPK2 model was easier to use in comparison to the PBPK3 model.

In the 3-state PBPK model, the vasculature compartment circulates ethanol throughout the whole body whereas the tissue (periphery) compartment acts as a storage unit. The liver compartment is where the ethanol is metabolized. The model is based on how alcohol enters and leaves the liver compartment.

Ethanol can enter the system in two ways: by ingestion and by venous infusion.

Venous infusion is simpler and is the preferred method in a laboratory setting because it eliminates the other issues that might arise with ingestion: taste, smell and absorption rate. Such a method has been developed by the Indiana School of Medicine and is called the alcohol clamp [25]. Since the parameters of the PBPK model are adjusted for each individual, the alcohol clamp method makes it possible to achieve the same alcohol concentration trajectory in each individual.

The three-state and two-state PBPK models can be thought of as electric circuits where alchol input and output follow Kirchoff laws. A circuit diagram for the three-state PBPK model is shown in Figure 2.1. Each compartment in the block diagram of Figure 2.1 is labeled with its name, its volume ($V_{\mathscr{X}}$ variable) and its state variable [$\mu_{\mathscr{X}}(t)$ variable] which is the mass of ethanol in the compartment. The subscripts $\mathscr{V}$, $\mathscr{T}$, $\mathscr{L}$ stand for Vascular, peripheral Tissue and for the Liver parenchyma, respectively. Each edge is labelled with a name and with the ethanol mass [$M_{\mathrm{X}}(t)$ variable] and volume flow [$R_{\mathrm{X}}(t)$ variable] that occur along that edge. The edges are directed, which indicates the positive flux direction[24].

The description of the three-state PBPK model as ordinary differential equations allow for further investigations in system identification, filtering and control system design, which is what this thesis is based upon.

2.5 Genetic Algorithm

Genetic algorithms were introduced by Holland [16] as a method for solving optimization problems based on evolutionary principles and natural selection. They are stochastic global search algorithms. The principle idea is that at every iteration of the algorithm, a new generation is produced and the fittest individual would survive, providing a solution

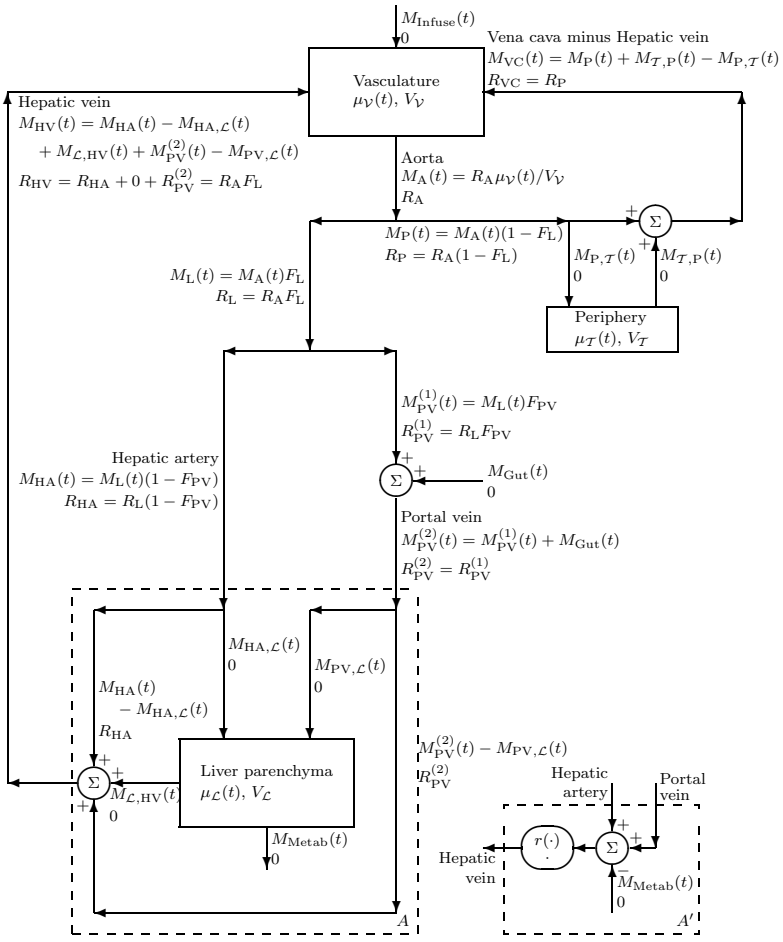


Figure 2.1: Circuit diagram for the 3-state PBPK model.

to the optimization problem.

The genetic algorithm works on a population of individuals where each individual is a potential solution to the optimization problem. Individuals are encoded as strings, which form the chromosomes, and the decision variables are the phenotypes of the genetic material. A binary alphabet is used in such encoding. The objective function, i.e. the cost, in an optimization problem is called the fitness function in genetic algorithms and it assesses the performance of an individual in the population. As in an optimiza-

tion problem, the most ”fit” individual will have the lowest cost. The fitness function chooses which individuals to mate together according to a probability. There are three measures of performance as represented by Baker: bias, spread and efficiency [2]. Bias is defined by, like in the statistical sense, the absolute difference between an individual's actual performance and the expected performance. Spread is the range in the number of trials that an individual may receive. Efficiency is how fast the algorithm converges. A roulette wheel selection algorithm is used in the selection process, which is a stochastic sampling method with replacement.

Once the individuals are chosen, their offspring are produced via recombination (crossover) methods. As the name implies, recombination produces offsprings with genetic material from both parents. Several recombination techniques may be employed in a genetic algorithm, such as multipoint crossover and uniform crossover. Mutation is the last step before the final offsprings are produced. Mutation is applied with low probability to the genes and the goal is to guarantee that the probability of searching any given string is not zero.

Once the offspring are created, they are again tested against the fitness function to see if more generations are necessary or if the fittest individual is a good solution to the problem [13]. Genetic algorithms differ from traditional optimization methods in the sense that they search a population of points at the same time and that they do not require derivative information. Another important difference is that the genetic algorithms return a series of solutions and the user picks the best solution, rather than returning a single solution. The genetic algorithm terminates when a predetermined number of generations is reached.

The pseudocode for a simple genetic algorithm is given in Algorithm 1.

Algorithm 1: **The simple genetic algorithm**

set number of individuals.

set maximum number of generations.

set number of variables.

set precision of variables.

set generation gap.

initialise population.

evaluate objective function for the initial population.

while generations **do**

  assign fitness values to the population.

  select individuals for breeding.

  recombine individuals and apply mutation.

  evaluate individual, call objective function.

  reinsert offspring into the population. [13]

end while

CHAPTER 3

PROGRESSIVE WORK EXPERIMENTS

3.1 A Point Process Model

Let t be the real-valued independent variable which will represent time. A point process is a stochastic process $N(t)$ which takes values in $\{0,1,2,\dots\}$ where $N(t)$ evaluated at a time t_0 is the number of arrivals that have occurred between the starting time (conventionally taken to be time 0) and time t_0. Therefore $N(t)$ is monotonically increasing in t. The interarrival times of the point process are the time intervals between arrivals, i.e., the time intervals between the step increases in $N(t)$. The simplest point process is the Poisson process which is completely described by a single positive real number, denoted by λ, which is the rate. In this process, the interarrival times are independent and identically distributed with an exponential probability density function (pdf) with parameter λ and this property (along with the choice that $N(0) = 0$) is one of the many equivalent definitions for the Poisson process.

The point process used to model the button presses described in Chapter 1 is more general than the Poisson process, in particular, is a doubly-stochastic Poisson process [29, Chapter 7] in which λ is not a deterministic constant but instead is itself a stochastic process. This allows the model to describe the idea that the human subject's pattern of button pushes can change over the course of the video game experiment. In particular, $\lambda(t)$ is related to a first-order Gauss Markov process, which is denoted by $x(t)$, by the following equations:

$$\lambda(t) = \exp(\mu + x(t)) \tag{3.1}$$

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \alpha x(t) + w(t) \tag{3.2}$$

where $w(\cdot)$ is a zero-mean Gaussian white noise process with power spectral density $N_0/2$, the initial condition on $x(0)$ is chosen so that $x(\cdot)$ is a wide-sense stationary stochastic process, and the three parameters in the model, all real valued, are μ, α, and $N_0/2$. The meanings of the parameters are described in the following paragraph.

Because $x(\cdot)$ is wide-sense stationary and Gaussian, $x(\cdot)$ is completely described by its mean function (denoted by $\bar{x}(t)$) and auto correlation function (denoted by $R_x(\tau)$) which have values

$$\bar{x}(t) = E[x(t)] = 0 \tag{3.3}$$

and

$$R_x(\tau) = E[x(t)x(t-\tau)] = \sigma^2 \exp(-|\tau|/\alpha) \tag{3.4}$$

where $E[\cdot]$ is expectation. (One implication of these results it that the statistics of $x(\cdot)$ and $-x(\cdot)$ are identical). Based on Eq. 3.4, the characteristic time over which changes in $x(\cdot)$ occur is $1/\alpha$ and the power in $x(\cdot)$ (i.e., $R_x(\tau = 0)$) is controlled jointly by α and $N_0/2$. Based on Eq. 3.3, the parameter μ controls the typical value of the time-varying rate $\lambda(t)$. Finally, the exponential function in Eq. 3.1 transforms a stochastic process that takes values in $(-\infty, +\infty)$ into a stochastic process that takes values in $(0, \infty)$ which is necessary if $\lambda(\cdot)$ is to be interpreted as a rate. Choosing $\lambda(t)$ as an exponential also simplifies the computation of some expectations that arise in the maximization part in Section 3.3.

As will be described in Section 3.3, we compute certain expectations based on the idea that the rate $\lambda(\cdot)$ is constant over intervals of duration Δ. Therefore, in order to make this choice exact rather than an approximation, we replace Eqs. 3.1–3.2 by the discrete time (sampling interval Δ) system

$$\lambda_n = \exp(\mu + x_n) \tag{3.5}$$

$$x_{n+1} = \rho x_n + w_n \tag{3.6}$$

where w_n is a zero-mean Gaussian white noise process with variance σ^2. Therefore the three parameters in the model, all real valued, are μ, ρ, and σ^2 and

$$\bar{x}_n = E[x_n] = 0 \tag{3.7}$$

and

$$R_x(l) = E[x_n x_{n-l}] = \frac{\sigma^2}{1-\rho^2}\rho^{-|l|}. \tag{3.8}$$

3.2 System Identification by a Maximum Likelihood Estimator

Given the sequence of arrival times (i.e., button pressing times) on the interval $[0,T]$, the goal is to determine the values of the parameters μ, ρ, and $N_0/2$ in Eqs 3.5 and 3.6. This estimation problem is solved by a maximum likelihood (ML) estimator, i.e., the estimated values of μ, α, and $N_0/2$, denoted by $\hat{\mu}$, $\hat{\alpha}$, and $\widehat{N_0/2}$ are defined by

$$\hat{\mu}, \hat{\alpha}, \widehat{N_0/2} = \arg\max_{\mu,\alpha,N_0/2} p(\{N(t) : 0 \le t \le T\} | \mu, \alpha, N_0/2) \tag{3.9}$$

where $p(\cdot|\cdot)$ is the conditional pdf on the arrival times given the parameter values. To solve this problem requires dealing with the real-valued time of each arrival. In order to formulate a simpler discrete-time problem [28], the arrivals are lumped into time bins of width Δ and the data is taken to be the number of arrivals in the bins, which is denoted by $\mathrm{d}N_k$ for the kth bin, i.e., $\mathrm{d}N_k = N(\Delta(k+1)) - N(\Delta k)$,where $k \in 0, \ldots \lfloor (T/\Delta) \rfloor - 1$. The equivalent equation for the moments is Eqs 3.7 and 3.8. Then Eq. 3.9 is replaced by

$$\hat{\mu}, \hat{\rho}, \widehat{\sigma^2} = \arg\max_{\mu,\rho,\sigma^2} p(\{\mathrm{d}N_k : k \in \{0,\ldots,K\}\} | \mu, \rho, \sigma^2) \tag{3.10}$$

where $T = K\Delta$.

Let $\mathrm{d}N$ without a subscript be the entire trajectory of $\mathrm{d}N_k$ for $k \in \{0,\ldots,K\}$ and

likewise for x and x_k. Let $\theta = (\mu, \rho, \sigma^2)$. Then

$$p(\mathrm{d}N|\theta, x) = \prod_{k=1}^{K} Poisson(\mathrm{d}N_k; \exp(\mu + \beta x_{k-1})\Delta) \tag{3.11}$$

and $p(\mathrm{d}N|\theta)$ can be computed from $p(\mathrm{d}N|\theta, x)$ by multiplying by the pdf $p(x|\theta)$ and integrating with respect to x.

As described in Section 2.2, a particularly attractive feature of an ML estimator is that standard theory provides an estimate of the covariance of the difference between the parameter estimates and the true parameter values [10]. This estimate requires computation of the Hessian of the log likelihood at the parameter values that maximize the log likelihood. While this computation is not currently implemented in the system described in this paper, it will be a key component of statistical tests for whether differences in parameter values are significant.

3.3 Computation of the Maximum Likelihood Estimator

An expectation-maximization (EM) algorithm [8] is used to compute the ML estimate described in Eq. 3.10. The nuisance parameters in the algorithm are x. Let $\theta^{(l)}$ be the parameter values θ at the lth iteration of the EM algorithm. The expectation step is to compute

$$Q(\theta|\theta^{(l)}) = \int \ln[L(\theta|\mathrm{d}N, x)] p(x|\mathrm{d}N, \theta^{(l)}) \mathrm{d}x \tag{3.12}$$

where $L(\theta|\mathrm{d}N, x) = p(\mathrm{d}N, x|\theta^{(l)})$ and the maximization step is to determine the θ which maximizes $Q(\theta|\theta^{(l)})$. The new estimate of the parameters, denoted by $\theta^{(l+1)}$, is this maximizing value of θ.

An iteration of the EM algorithm for this problem has two parts [28]. The first part is to compute first and second order conditional moments of x given the data $\mathrm{d}N$. Q can

be written in terms of these moments and the second part is to determine the value of θ that maximizes Q which can also be written in terms of these moments. The necessary moments are

$$\begin{aligned} x_{k|k} &\doteq E[x_k|\{dN_l : l \in \{0,\ldots,k\}\}] && (3.13)\\ \sigma^2_{k|k} &\doteq E[(x_k - x_{k|k})^2|\{dN_l : l \in \{0,\ldots,k\}\}] && (3.14)\\ W_k &\doteq E[x_k^2|\{dN_l : l \in \{0,\ldots,K\}\}] && (3.15)\\ W_{k,k-1} &\doteq E[x_k x_{k-1}|\{dN_l : l \in \{0,\ldots,K\}\}]. && (3.16) \end{aligned}$$

In terms of these moments, the exact $\theta = (\mu, \rho, \sigma^2)$ that maximizes Q is

$$\begin{aligned} \rho^{(l+1)} &= \frac{\sum_{k=1}^K W_{k,k-1}}{\sum_{k=1}^K W_{k-1}} && (3.17)\\ \left(\sigma^2\right)^{(l+1)} &= \frac{1}{K}\Bigg\{ W_k - 2\rho^{(l+1)} W_{k,k-1} + \left[\rho^{(l+1)}\right]^2 W_{k-1} \\ &\qquad + W_0\left[1 - \left[\rho^{(l+1)}\right]^2\right]\Bigg\} && (3.18)\\ \mu^{(l+1)} &= \ln \sum_{k=1}^K dN(k\Delta) \\ &\qquad - \ln \sum_{k=1}^K \exp\left(\beta x_{k|K} + \frac{1}{2}\beta^2 \sigma^2_{k|K}\right)\Delta. && (3.19) \end{aligned}$$

Equation 3.19 benefits from the choice of an exponential in Eq 3.5 because that choice leads to computing the expectation of the exponential of a Gaussian random variable (essentially x_k), which can be done exactly via log-normal methods. Then one step of the EM algorithm requires the following computations [28] in the forward (increasing k) direction,

$$\begin{aligned} x_{k|k} &= \rho x_{k-1|k-1} + \sigma^2_{k|k-1}[y(k\Delta) - \exp(\mu + x_{k_k})\Delta && (3.20)\\ \sigma^2_{k|k} &= -\left[\frac{1}{\sigma^2_{k|k-1}} - \exp(\mu + x_{k|k})\Delta\right]^{-1} && (3.21)\\ x_{k|k-1} &= \rho x_{k-1|k-1} \\ \sigma^2_{k|k-1} &= \rho^2 \sigma^2_{k-1|k-1} + \sigma^2, \end{aligned}$$

with initial condition $x_{0|0} = x_0$, $\sigma^2_{k|k} = \sigma^2(1-\rho^2)^{-1}$ and the following computations [28] in the reverse (decreasing k) direction,

$$A_k = \rho\sigma^2_{k|k}(\sigma^2_{k+1|k})^{-1} \tag{3.22}$$

$$x_{k|K} = x_{k|k} + A_k(x_{k+1|K} - xk+1|k) \tag{3.23}$$

$$\sigma^2_{k|K} = \sigma^2_{k|k} + A^2_k(\sigma^2_{k+1|k} - \sigma^2 k+1|l), \tag{3.24}$$

where $k = K-1,\ldots,1$ and the initial conditions are $x_{K|K}$ and $\sigma^2_{K|K}$ from the forward phase. Finally, using state-space covariance ideas [7],

$$W_{k,k+1} = A_k\sigma^2_{k+1|K} + x_{k|K}x_{k+1|K} \tag{3.25}$$

$$W_k = \sigma^2_{k|K} + x^2_{k|K}. \tag{3.26}$$

Unlike Equations 3.17- 3.19, which are exact, the computing of moments cannot be done exactly with a feasible amount of calculation. Therefore, approximations are necessary. Furthermore, while Eqs 3.13- 3.14 are conditioned on dN_k for $l \in \{0,\ldots,k\}$ and are therefore filtering problems, Eqs 3.15- 3.16 are conditioned on dN_l for $l \in \{0,\ldots,K\}$ and therefore computing W_k and $W_{k,k-1}$ are smoothing problems. Standard nonlinear filtering ideas are used to approximately solve this problem in a forward filtering and backward smoothing filter algorithm [28]. The forward phase equation for $x_{k|k}$ is determined by maximum likelihood computed by a Newton method. The forward phase equation for $\sigma^2_{k|k}$ is calculated by setting the variance of the Gaussian random variable to the inverse of an approximation of its Fisher information matrix [10]. The equations for $x_{k|k-1}$ and $\sigma^2_{k|k-1}$ are standard Kalman filter equations [1]. The backward phase uses the backwards part of a forward-backward Kalman smoother [22]. Finally, state-space covariance ideas [7], allow the evaluation of W_k and $W_{k,k-1}$. The initial condition used in the moment calculations for the first iteration of the EM are $\mu = [N(T) - N(0)]/K$, $\rho = 0$, $\sigma^2 = \mu/4$ and $x_0 = 0$. The moment calculations at iteration l are the results of iteration $l-1$.

Let $H(\theta)$ be the Hessian of the log likelihood at parameter value θ, a 3×3 matrix. Then the negative inverse of the Hessian evaluated at the maximum likelihood parameter value, $-H^{-1}(\theta^*)$, is the empirical Fisher information matrix [10] where θ^* is the maximum likelihood parameter value. All of these partial derivatives can be calculated in terms of the moments described earlier in this section. The matrix is always transpose symmetric and, at a maximum of the likelihood, is positive semidefinite. The Hessian matrix can be used to describe the error in the estimate. In particular, the error is approximately Gaussian with mean zero and covariance $-H^{-1}(\theta^*)$ where θ^* is the maximum likelihood parameter value. Due to the structure in the definition of the log-likelihood, the second partial derivative of the log-likelihood with respect to μ and ρ and with respect to μ and σ^2 are 0. However, the Hessian is still full rank, and due to the matrix inverse, the approximate covariance matrix has nonzero entries.

3.4 Simulations

The algorithm of the previous section has two components - computing the moments and computing the ML estimator of μ, ρ, and σ^2. In simulation, it is possible to separately evaluate the performance of both components.

The first simulation covers the filter that computes $x_{k|k}$. The simulated data is from Eqs 3.5 and Eqs 3.6with $\mu = 6$, $\rho = 0.8$, $\sigma^2 = 1$, $K = 100$, $\Delta = 0.3$ seconds and x_0 set to the corresponding steady state value. The initial conditions for the filter are the initial conditions for the overall algorithm described following Eq. 3.19. In summary of the results, Figure 3.1 shows the $x_{k|k}$ from Eq. 3.20 for one trajectory demonstrating that many qualitative features of x_k are successfully preserved in $x_{k|k}$ in spite of the exponential nonlinearity.

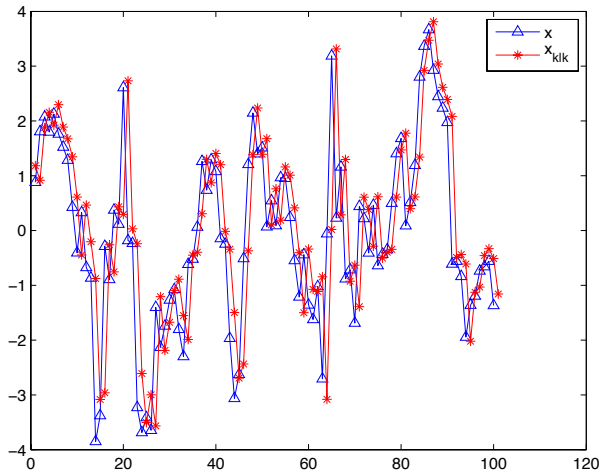


Figure 3.1: x before filtering vs. $x_{k|k}$, after filtering

The primary goal of the estimation process is the identification of the system parameters, μ,ρ and σ^2. We have done Monte Carlo calculations (J=20 trajectories, each K=1000 samples long) of the bias and the variance of the estimate for $\mu \in [0.2,2]$, $\rho \in [0.002,0.92]$,$\sigma^2 \in [0.01,0.5]$ and $\Delta \in 1,2$, which covers the values we have observed in the experimental data. The EM algorithm is always started for the initial conditions observed following Eq 3.24.

Overall, the results in Table 3.1 show parameter dependent biases and small variances around the biased value. We summarize the results as a function of μ because large values of μ versus small values of μ really influence whether the rate of arrivals is large or small. Among the results concerning μ, we focus on simulated data where the true values of all the parameters are near the values seen in the experimental data, specifically, $\mu \in [1,1.25,1.5,1.75,2.0]$, $\rho = 0.1$,$\sigma^2 = 0.1$ and $\Delta = 2$ seconds. The mean of the estimate of μ as a function of the true parameter values by Monte Carlo is $\bar{\mu} = \frac{1}{J}\sum_{j=1}^{J} \hat{\mu^{(j)}}$ where j indexes Monte Carlo trials. The bias is $b_\mu = \bar{\mu} - \mu^{true}$, and the standard deviation is $S_\mu = \left[\frac{1}{J-1}\sum_{j=1}^{J}(\mu^{(j)} - \bar{\mu})^2\right]^{1/2}$.

Table 3.1: Summary of simulation results concerning μ

μ^{true}	1	1.25	1.5	1.75	2.0
$\bar{\mu}$	0.9387	1.2623	1.4888	1.6907	2.0194
b_μ	-0.0391	-0.0371	-0.0261	-0.0350	-0.0209
S_μ	0.0220	0.0182	0.0157	0.0160	0.0173

As shown in Table 3.4, the bias is less than 4% and the standard deviation is less than 2% of the true values and the dependence on the true value is weak.

3.5 Experimental Results

All of the data investigated in this thesis had been collected by Dr. Sean J. O'Connor under protocols that were approved by the Indiana University Institutional Review Board. All subjects provided written informed consent. The experimental results for two subjects are shown in Figure 3.2. The two curves have differences at both large and small time scales. For instance, at the time scale of 10^3 samples the deviations from 0 are entirely different in pattern, even in direction. On a small time scale, Figure 3.2(a) has more variation than Figure 3.2(b) shown here as the appearance of a thicker curve for Figure 3.2(a) than Figure 3.2(b).

The histogram of interarrival times is shown in Figure 3.3. The data is only measured to within 1 second so the histogram is discretized. In spite of the discretization, the histogram does not appear to originate from an exponential interarrival time pdf (since the curve is not a straight line), motivating the doubly stochastic model of Eqs. 3.5–3.6.

The data proceeds in epochs, where each epoch is the time interval in which the subject presses the buttons. The estimated dynamical system parameters μ, ρ, σ^2 and

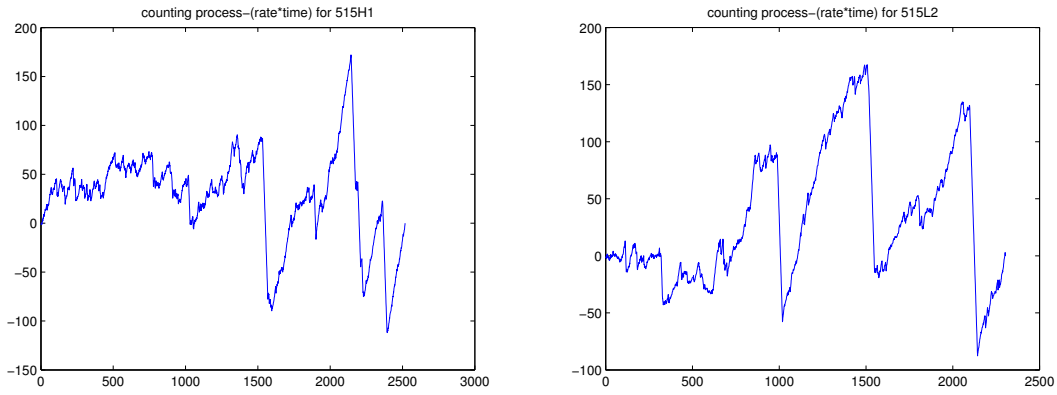


Figure 3.2: Experimental results on two subjects. The curve is the counting process for button presses as a function of the time minus the average rate multiplied by time so that the difference remains near zero. The average is exactly the μ used in initialization of the estimator

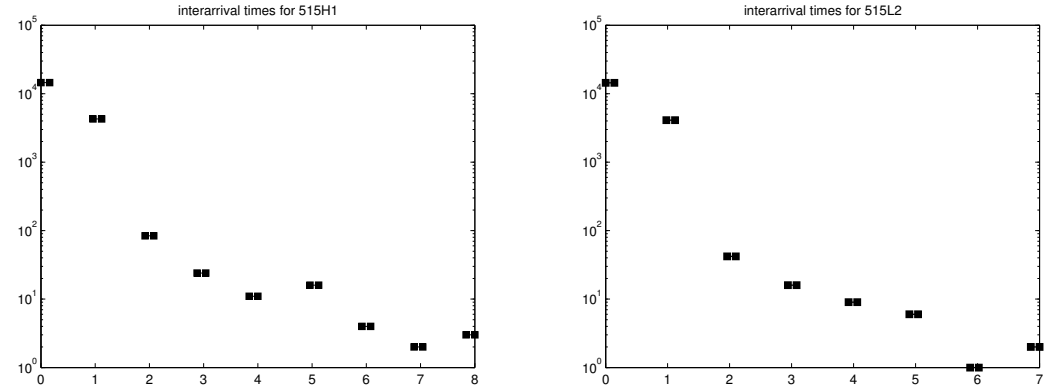


Figure 3.3: Log-histogram of the interarrival times (seconds) for experimental results on one subject given high level (left) and low level (right) alcohol rewards respectively.

an estimate of the variance of $x(\cdot)$, $\hat{\sigma}^2/(1-\hat{\rho}^2)$, are shown in Table 3.2 for the 8th epoch for the two experiments on each of four subjects using an estimator with $\Delta = 2$ seconds. With the exception of Subject 515, larger rewards lead to a larger value of $\hat{\mu}$ which is the constant part of the rate function for button pushing. Larger rewards had a less consistent effect on the time-varying properties of the rate function since in half

Table 3.2: Parameter estimates from experimental data. "High" (H) versus "Low" (L) is the amount of ethanol provided at the completion of the task.

Subject	Reward	$\hat{\rho}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\sigma}^2/(1-\hat{\rho}^2)$
527	H	0.289486	0.0691861	1.60211	0.0755
527	L	0.308521	0.0838568	1.16288	0.0927
548	H	0.289161	0.147671	1.14713	0.161
548	L	0.00054163	0.0113837	0.992117	0.114
502	H	0.0974935	0.101417	1.43138	0.124
502	L	0.0856287	0.210045	1.25287	0.212
515	H	0.852337	0.277011	1.06055	1.013
515	L	0.860755	0.164302	1.28342	0.634

of the cases larger reward lead to a larger value of $\hat{\sigma}^2$ while in the other half it did not and, similarly, in half of the cases larger reward lead to a larger value for the estimated variance of $x(\cdot)$ while in the other half it did not. The $\hat{\mu}$ results are consistent with expectations for these subjects, all of whom are non-treatment-seeking alcoholics.

The standard deviation of the estimation error for each parameter, denoted $\bar{s}$, is the square root of the diagonal of the empirical Fisher information matrix. The results for $\bar{s}$ in the order μ, ρ, σ^2 for 515H (515L) are $[0.0103, 0.0154, 0.0071]$ ($[0.0104, 0.0054, 0031]$), where "H" and "L" indicate the high and low levels of rewards respectively. Because the difference between μ for a low and high reward is roughly $0.2 - 0.5$ while the standard deviation of the estimation error for the low reward case is roughly $0.01 - 0.05$, the probability of the largest symmetric confidence interval that excludes the high reward estimate is essentially 1.

For statistical testing on the data, two different tests are performed: two-population t-test and two-population z-test [5]. In the two population t-test, the variances of the classes are assumed to be unknown but equal, and the samples are the button pushes

from the eighth epoch of high and low reward experiments. The hypotheses are

$$H_0 : \bar{m}_H - \bar{m}_L = 0 \tag{3.27}$$

$$H_1 : \bar{m}_H - \bar{m}_L \neq 0, \tag{3.28}$$

where $\bar{m}$ indicates the mean and H and L indicate the "high" and "low" reward games. The results of this test for the subjects $527, 548, 515, 502$ are $H = 1, H = 1, H = 1, H = 0$ respectively, where $H = 1$ indicates that the null hypothesis is rejected.

For a two-population z-test, variance for each population is known. We calculate the variance by using the output of our algorithm and the Hessian information,

$$\exp\left(2\hat{\mu} + 2\frac{-1}{H_\rho} + \frac{2\widehat{\sigma^2}}{1 - \left(\frac{-1}{H_\rho} + \hat{\rho}^2\right)}\right),$$

where H_ρ and H_μ indicate the relative terms in the Hessian matrix. The hypothesis are the same as Eqs. 3.27 - 3.28. The results of the tests for subjects $527, 548, 515, 502$ are $H = 1, H = 1, H = 1, H = 0$ respectively. Both two-population t-test and two population z-test fail to reject the null hypothesis for subject 502. The p-values for the two tests on subject 502 where the null hypothesis was not rejected are [0.210, 0.104] for t-test and z-test, respectively. Though the value for the z-test is closer, neither is small enough for a size $\alpha = 0.05$ test.

3.6 Computation of the Maximum Likelihood via Markov Chain Monte Carlo Methods

The other method used in evaluating the expectations in Section 3.2 is the Markov Chain Monte Carlo(MCMC) method. As outlined in Ref. [9], MCMC methods use Sequential

Importance Sampling (SIS), where x_k is drawn from an importance density and the relative importance weights are calculated. The importance weights are then normalized, so that they add up to 1, and the expectations and any summary statistics can be computed via the importance weights. This procedure is done as many times as required, the trade off being the computation time.

Our model that we have described in Section 3.1 has a linear Gaussian state equation and observations are distributed Poisson. In such case, by using [9], we can set the importance function for x_k as:

$$\pi(x_k|x_{k-1}, dN_k) = \mathcal{N}(m(x) + x, \Sigma(x)) \tag{3.29}$$

$$\Sigma(x) = -l''(x)^{-1} \tag{3.30}$$

$$m(x) = \Sigma(x) l'(x), \tag{3.31}$$

where $l(x)$ is the log-likelihood function shown in Section 3.3.

Since the observations are Poisson in nature, an approximation to the second derivative of the likelihood can be made, [9], where

$$l''(x_k) = -\exp(\mu + x_k)\Delta - \frac{1}{\sigma^2},$$

and the mode x can be computed by a Newton method where

$$x_{j+1} = x_j - \left[l''(x_{(j)}\right]^{-1} l'(x_{(j)})$$

The moments can be computed as follows:

$$x_{k|k}(j) = \sum_{j=1}^{N} \tilde{w_k}^{(j)} x_k^{(j)} \tag{3.32}$$

$$\sigma_{k|k}^2(j) = (\sum_{j=1}^{N} \tilde{w_k}^{(j)} (x_k^{(j)})^2) - x_{k|k}^2 \tag{3.33}$$

$$W_k = \sum_{j=1}^{K} \sigma_{k|K}^2(j) + x_{k|K}^2(j) \tag{3.34}$$

$$W_{k,k-1} = \sum_{j=1}^{K} \sigma_{k|K}^2(j) * A_k(j) + (x_{k|K}(j) * x_{k|K}(j+1)), \tag{3.35}$$

where the last two are smoothing operations and they are computed $j = K, K-1..1$.

The SIS algorithm requires resampling. Once the importance weights are calculated the effective sample size [9] is calculated. An estimate of the effective sample size is:

$$\widehat{N_{eff}} = \frac{1}{\sum_{i=1}^{N} (\tilde{w}_k^{(i)})^2}.$$

If the calculated N_{eff} is larger than the threshold set then there is no resampling needed. However, if N_{eff} is smaller than the threshold then an index $j(i)$ is sampled with a distribution $Prj(i) = l = \tilde{w_k}^{(l)}$. The weights are then set to $1/N$.

For fixed interval smoothing, it is necessary to calculate W_k and $W_{k,k-1}$ and the algorithm outlined in Ref. [9] is followed. To calculate the new importance weights, the following formula is used:

$$\tilde{w}_{k|n}^{(j)} = \sum_{j=1}^{N} \tilde{w}_{k+1|n}^{(j)} \frac{\tilde{w}_k^{(i)} p(x_{k+1}^{(j)} | x_k^{(j)})}{\sum_{l=1}^{N} \tilde{w}_k^{(l)} p(x_{k+1}^{(j)} | x_k^{(l)})}.$$

3.7 Simulation Results with MCMC Algorithm

The MCMC algorithm has different components and each component was tested to find the optimal result before running our algorithm on the experimental data. One of the

important factors is the number of times the algorithm is allowed to run. The algorithm is run on synthetic data with parameters, $\rho = 0.5, \sigma^2 = 1, \mu = 2, \Delta = 2, x_o = 2$. The Table 3.3 summarizes results, for a MATLAB seed that has been set to 55.

Table 3.3: Simulation results showing the different number of trials and parameter estimates

Number of trials	Parameter	Results
J=100	ρ	0.465
	σ^2	0.02
	μ	0.8583
J=250	ρ	0.5398
	σ^2	0.0334
	μ	0.8233
J=300	ρ	0.5426
	σ^2	0.03
	μ	0.8487

Table 3.3 shows that the parameter estimates do not change significantly between $J = 250$ and $J = 300$. In experimental results, $J = 250$ will be used as the numer of trials.

We have also found that the threshold, $N_threshhold$, is an important factor and it should be based upon the number of trials needed to get meaningful results. In our simulations and experimental data analysis, it is set to $J/4$.

3.8 Experimental Results with MCMC

The estimated dynamical system parameters, ρ, μ,σ^2 are shown in Table 3.4. For these tests, number of trials was set to $J = 250$ and the MATLAB seed was set to a predetermined number, so that the random numbers generated for input generation are the same in all MCMC trials. Since these parameters are estimated J times, the average of all estimates are reported in Table 3.4.

Table 3.4: Parameter estimates from experimental data by using MCMC methods. "High" (H) versus "Low" (L) is the amount of ethanol provided at the completion of the task.

Subject	Reward	$\hat{\rho}$	$\hat{\sigma}^2$	$\hat{\mu}$
527	H	0.70	0.03	1.49
527	L	0.54	0.24	0.83
548	H	0.50	0.01	1.22
548	L	0.30	0.07	0.82
502	H	0.54	1.33	0.27
502	L	0.28	0.83	0.72
515	H	0.87	0.26	0.48
515	L	0.73	0.22	0.66

The parameter estimates vary greatly from the estimates of Section 3.5, however, the difference in parameters between high and low levels of reward is pronounced in this algorithm as well as in the algorithm of Section 3.5. The subjects, with the exception of 515, all resulted in higher μ estimates with a higher reward test. In particular, for subject 515, both algorithms(MCMC and Section 3.5) give a value of $\hat{\rho}$ that is large compared with the value estimated than other subjects where a large ρ implies a long correlation

time in the subject's responses.

CHAPTER 4

NON-LINEAR MODEL ANALYSIS

4.1 Physiologically Based Pharmacokinetic Models

Both the 3-state and 2-state Physiologically Based Pharmacokinetic (PBPK) [24] models were developed at the Indiana Alcohol Research Center, Indiana University School of Medicine. In this work, both the 3-state and 2-state models are presented but the analysis is focused on the 3-state model, which is the more complete model.

The notation for the 3-state PBPK model is as follows:

- $C_{x,y}$: ethanol concentration in a compartment.
- R_x: flow rate in a vein.
- $k_{k,y}$: fraction of the available ethanol that is transported.
- $M_{x,y}$: mass flux of ethanol form x to y.
- M_{gut}: oral ethanol input to the system.
- M_{infuse}: intravenous input to the system.
- F_x: fraction of the input that exists on the exiting edges.
- V_{max}: maximum velocity of conversion of ethanol to inactive metabolites.
- M_{metab}: mass flux from the liver parenchyma out of the system.

Let $u(x)$ be the unit step and $r(x)$ be the unit ramp functions. The three state PBPK model has a state equation for the ethanol concentration in each of the liver ($\mathscr{L}$), the vascular volume ($\mathscr{V}$), and the remainder of the body ($\mathscr{T}$). Each compartment is described by the mass of ethanol in the comparment and the mathemetical model is an

array of first-order non-linear differential equations. The vasculature compartment circulates ethanol through the system whereas the periphery compartment acts as a storage compartment. The liver compartment is where alcohol elimination occurs by enzymes. The differential equations for the model are:

$$\begin{aligned}
\frac{\mathrm{d}\mu_{\mathscr{L}}}{\mathrm{d}t}(t) &= M_{\mathrm{HA},\mathscr{L}}(t) + M_{\mathrm{PV},\mathscr{L}}(t) - M_{\mathscr{L},\mathrm{HV}}(t) - M_{\mathrm{Metab}}(t) && (4.1)\\
\frac{\mathrm{d}\mu_{\mathscr{T}}}{\mathrm{d}t}(t) &= M_{\mathrm{P},\mathscr{T}}(t) - M_{\mathscr{T},\mathrm{P}}(t) && (4.2)\\
\frac{\mathrm{d}\mu_{\mathscr{V}}}{\mathrm{d}t}(t) &= -M_{\mathrm{HA},\mathscr{L}}(t) - M_{\mathrm{PV},\mathscr{L}}(t) + M_{\mathscr{L},\mathrm{HV}}(t) \\
&\quad + M_{\mathscr{T},\mathrm{P}}(t) - M_{\mathrm{P},\mathscr{T}}(t) + M_{\mathrm{Infuse}}(t) + M_{\mathrm{Gut}}(t). && (4.3)
\end{aligned}$$

$M_{\mathrm{Infuse}}(t)$ and $M_{\mathrm{Gut}}(t)$ are external inputs from intravenous infusion and oral intake, respectively. Therefore, it remains only to provide equations in terms of $\mu_{\mathscr{X}}(t)$ for the mass fluxes $M_{\mathrm{Metab}}(t)$, $M_{\mathrm{HA},\mathscr{L}}(t)$, $M_{\mathrm{PV},\mathscr{L}}(t)$, $M_{\mathscr{L},\mathrm{HV}}(t)$, $M_{\mathscr{T},\mathrm{P}}(t)$, and $M_{\mathrm{P},\mathscr{T}}(t)$. From Michaelis-Menten kinetics,

$$M_{\mathrm{Metab}}(t) = V_{\mathscr{L}} V_{\max} \frac{\mu_{\mathscr{L}}(t)/V_{\mathscr{L}}}{K_m + \mu_{\mathscr{L}}(t)/V_{\mathscr{L}}}. \tag{4.4}$$

The remaining terms are

$$\begin{aligned}
M_{\mathrm{P},\mathscr{T}}(t) &= k_{\mathrm{P},\mathscr{T}} R_{\mathrm{A}} (1 - F_{\mathrm{L}}) r \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} - \frac{\mu_{\mathscr{T}}(t)}{V_{\mathscr{T}}} \right) && (4.5)\\
M_{\mathscr{T},\mathrm{P}}(t) &= k_{\mathscr{T},\mathrm{P}} R_{\mathrm{A}} (1 - F_{\mathrm{L}}) r \left(\frac{\mu_{\mathscr{T}}(t)}{V_{\mathscr{T}}} - \frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} \right) && (4.6)\\
M_{\mathrm{HA},\mathscr{L}}(t) &= k_{\mathrm{HA},\mathscr{L}} R_{\mathrm{A}} F_{\mathrm{L}} (1 - F_{\mathrm{PV}}) r \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} - \frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} \right) && (4.7)\\
M_{\mathrm{PV},\mathscr{L}}(t) &= k_{\mathrm{PV},\mathscr{L}} R_{\mathrm{A}} F_{\mathrm{L}} F_{\mathrm{PV}} r \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} + \frac{M_{\mathrm{Gut}}(t)}{R_{\mathrm{A}} F_{\mathrm{L}} F_{\mathrm{PV}}} - \frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} \right) && (4.8)\\
M_{\mathrm{HA}}(t) &= R_{\mathrm{A}} F_{\mathrm{L}} (1 - F_{\mathrm{PV}}) \frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} && (4.9)\\
M_{\mathrm{PV}}^{(2)}(t) &= \frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} R_{\mathrm{A}} F_{\mathrm{L}} F_{\mathrm{PV}} + M_{\mathrm{Gut}}(t) && (4.10)\\
M_{\mathscr{L},\mathrm{HV}}(t) &= k_{\mathscr{L},\mathrm{HV}} R_{\mathrm{A}} F_{\mathrm{L}} r \left(\frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} - C_{\mathrm{HV}}(t) \right). && (4.11)
\end{aligned}$$

Finally, in order to evaluate $C_{\mathrm{HV}}(t)$ it is necessary to account for mass flux from the liver parenchyma to the hepatic vein which is done by

$$\alpha(t) = \frac{1}{R_{\mathrm{A}}F_{\mathrm{L}}}\left(M_{\mathrm{HA}}(t) - M_{\mathrm{HA},\mathscr{L}}(t) + M^{(2)}_{\mathrm{PV}}(t) - M_{\mathrm{PV},\mathscr{L}}(t)\right) \tag{4.12}$$

$$\gamma(t) = \frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} \tag{4.13}$$

$$C_{\mathrm{HV}}(t) = \begin{cases} \frac{1}{1+k_{\mathscr{L},\mathrm{HV}}}\left(\alpha(t) + k_{\mathscr{L},\mathrm{HV}}\gamma(t)\right) & \alpha(t) < \gamma(t) \\ \alpha(t) & \alpha(t) > \gamma(t) \end{cases}. \tag{4.14}$$

The parameters that need to be identified for individual subjects are $k_{x,y}$, R_A, $V_{\mathscr{V}}$, $V_{\mathscr{L}}$, $V_{\mathscr{T}}$ and V_{max} while using nominal values for the other constant, F_x, has been satisfactory.

The two state model lacks a variable for the mass of ethanol in the liver (i.e., $\mu_{\mathscr{V}}(t)$), because the volume of the liver (i.e., $V_{\mathscr{L}}$) has been difficult to estimate. Instead, the concentration of ethanol in the liver compartment is assumed to be the concentration entering the liver. This assumption can lead to negative concentration values under certain conditions. In summmary, the resulting equations are

$$\frac{\mathrm{d}\mu_{\mathscr{T}}}{\mathrm{d}t}(t) = M_{\mathrm{P},\mathscr{T}}(t) - M_{\mathscr{T},\mathrm{P}}(t) \tag{4.15}$$

$$\begin{aligned}\frac{\mathrm{d}\mu_{\mathscr{V}}}{\mathrm{d}t}(t) = {} & -\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}}R_{\mathrm{A}}F_{\mathrm{L}} + M_{\mathscr{T},\mathrm{P}}(t) - M_{\mathrm{P},\mathscr{T}}(t) \\ & + r\left[\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}}R_{\mathrm{A}}F_{\mathrm{L}} + M_{\mathrm{Gut}}(t) - M_{\mathrm{Metab}}(t)\right] \\ & + M_{\mathrm{Infuse}}(t);\end{aligned} \tag{4.16}$$

the new equations

$$C_{\mathscr{L}}(t) = \frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} + \frac{M_{\mathrm{Gut}}(t)}{R_{\mathrm{A}}F_{\mathrm{L}}} \tag{4.17}$$

$$M_{\mathrm{Metab}}(t) = M_{\max}\frac{C_{\mathscr{L}}(t)}{K_m + C_{\mathscr{L}}(t)}; \tag{4.18}$$

$M_{\mathrm{P},\mathscr{T}}(t)$ (Eq. 4.5), and $M_{\mathscr{T},\mathrm{P}}(t)$ (Eq. 4.6); and the two external inputs $M_{\mathrm{Gut}}(t)$ and $M_{\mathrm{Infuse}}(t)$. The parameters that need to be identified for individual subjects are $k_{x,y}$,

R_A, $V_{\mathscr{V}}$, $V_{\mathscr{L}}$, $V_{\mathscr{T}}$ and V_{max} while using nominal values for the other constant, F_x, has been satisfactory.

4.2 The measurement model

The time at which a measurement is taken is assumed to be known exactly. The value of the measurement is, however, not perfect and the fact that some measurements are incorrectly performed must be accounted for. Both of these aspects are combined in a Gaussian mixture model. We assume that whether a measurement is or is not correctly performed is independent from measurement to measurement. Let $x(t)$ be the state of the PPBK model. Let h be a matrix such that $h^T x(t)$ is the blood ethanol concentration in the PBPK model. Let t_i and y_i be the time and value of the ith measurement. Let i.i.d. stand for independent and identically distributed and let the Gaussian probability density function (pdf) with mean m and covariance Σ evaluated at value x be denoted by $N(m,\Sigma)(x)$. With probability q_i the measurement is successful, in which case the measurement error is Gaussian with mean μ_1 (which is probably 0) and variance σ_1^2: $y_i = h^T x(t_i) + \mu_1 + \sigma_1 w_i$ where w_i is i.i.d. $N(0,1)$. On the other hand, with probability $1-q_i$, the measurement is unsuccessful, in which case the measurement error is Gaussian with mean μ_2 and variance σ_2^2: $y_i = h^T x(t_i) + \mu_2 + \sigma_2 w_i$ where w_i is i.i.d. $N(0,1)$. The value of μ_2 is possibly negative because an unsuccessful measurement is often due to measuring the ethanol concentration in air that is not end-expiratory air and therefore has a lower ethanol concentration. The variances probably satisfy $\sigma_1^2 < \sigma_2^2$.

The model of the preceding paragraph can be summarized in several ways. Focusing on probabilities, the error between y_i and $h^T x(t_i)$ is described by a Gaussian mixture pdf:

$$p(y_i - h^T x(t_i)) = q_i N(\mu_2, \sigma_1^2)(y_i - h^T x(t_i)) + (1-q_i) N(\mu_2, \sigma_2^2)(y_i - h^T x(t_i)). \quad (4.19)$$

Alternatively, focusing on conditional probabilities, the conditional probability distribution on the measurement given the true value is also a Gaussian mixture pdf:

$$p(y_i|h^T x(t_i)) = q_i N(h^T x(t_i) + \mu_1, \sigma_1^2)(y_i) + (1-q_i)N(h^T x(t_i) + \mu_2, \sigma_2^2)(y_i). \quad (4.20)$$

Instead of probabilities, focus on labels. In particular, let z_i be a label that takes value 0 if the ith measurement is successful and value 1 otherwise. Therefore, the probability mass function (pmf) on z_i is

$$p(z_i) = \begin{cases} q_i, & z_i = 1 \\ 1-q_i, & z_i = 2 \end{cases}. \quad (4.21)$$

Then

$$y_i = h^T x(t_i) + \mu_{z_i} + \sigma_{z_i} w_i \quad (4.22)$$

where w_i are i.i.d. $N(0,1)$ Alternatively, focusing on conditional probabilities, the conditional probability distribution on the measurement given the true value and the value of the label is Gaussian:

$$p(y_i|z_i, h^T x(t_i)) = N(h^T x(t_i) + \mu_{z_i}, \sigma_{z_i}^2)(y_i). \quad (4.23)$$

Two different statistical cases are considered. In the first case, the probability that the ith measurement is incorrectly performed is not known and is the same for every measurement. Therefore q_i is independent of i and is unknown and to be estimated. In the second case, the probability is known but the known value is not necessarily the same for every measurement. Therefore q_i depends on i but has a known value. We focus on the case where q_i is independent of i and unknown.

4.3 Maximum likelihood estimation of the model parameters

The unknown parameters in the PBPK model are denoted by θ_x. The unknown parameters in the measurement model are denoted by θ_y and the largest set is $\theta_y = (q, \mu_1, \sigma_1, \mu_2, \sigma_2)$ where q is the probability of a correct measurement and μ_1 and σ_1^2 (μ_2 and σ_2^2) are the mean and variance of a correctly performed (incorrectly performed) measurement. It may be that μ_1 is assumed to be zero. Let $\theta = (\theta_x, \theta_y)$. Let $x(\cdot;\theta_x)$ be the state trajectory of the PBPK model with parameter vector θ_x and a known input. Let $h \in \mathbb{R}^n$ (n is 2 or 3) be a known vector such that $h^T x(\cdot;\theta_x)$ is the blood ethanol concentration trajectory. The ith measurement is denoted by y_i and occurs at a time that is denoted by t_i. The measurement equation is

$$y_i = h^T x(t_i;\theta_x) + \sigma(z_i;\theta_y) w_i + \mu(z_i;\theta_y) \tag{4.24}$$

where μ is the mean and σ^2 is the variance of the additive noise and $z_\cdot$ and $w_\cdot$ are the sequences describing the class label for correctly versus incorrectly performed measurements and the additive noise, respectively. The sequences $z_\cdot$ and $w_\cdot$ are independent. The elements of the sequence $z_\cdot$ are in $\{1,2\}$. The sequence $z_\cdot$ is independent and identically distributed (i.i.d.) with $P(z_i = 1) = q$. The sequence $w_\cdot$ is i.i.d. $N(0,1)$. The mean and variance functions are defined by $\mu(z_i,\theta_y) = \mu_{z_i}$ and $\sigma^2(z_i,\theta_y) = \sigma_{z_i}^2$, respectively. Let the number of measurements be denoted by N_y and define $y = (y_1, \ldots, y_{N_y})$ and $z = (z_1, \ldots, z_{N_y})$.

The goal is a maximum likelihood estimate of θ (denoted by $\hat{\theta}$), i.e.,

$$\hat{\theta} = \arg\max_{\theta} p(y|\theta). \tag{4.25}$$

Because the problem would be simplified if the class labels $z_\cdot$ were known, we use an expectation maximization (EM) algorithm [21] where the hidden variables are exactly

the z_i variables. An EM algorithm is an iterative algorithm that starts from a user-provided initial condition for the parameter values and, at each iteration, updates the entire set of parameter values in a manner such that the likelihood is non-decreasing. Updating only a subset of parameter values at any particular iteration gives a broad class of algorithms called generalized EM algorithms [21] which can offer advantages. In the problem of interest here, alternating between updating θ_x and updating θ_y is an attractive approach because the update with respect to θ_y is simple. The update for θ_y is described in Section 4.3 and the more complicated update for θ_x is described in Sections 4.4 and 4.5. Finally, the complete algorithm is summarized in Section 4.5.

Update of θ_y

From Eq. 4.24,

$$y_i - h^T x(t_i;\theta_x) = \sigma(z_i;\theta_y)w_i + \mu(z_i;\theta_y) \tag{4.26}$$

which, when θ_x is known, is exactly the problem of maximum likelihood estimation of the parameters in a Gaussian mixture model [21, Section 2.7, pp. 68–74][26, Eqs. 4.5–4.9, pp. 217–218][4]. Let

$$\tilde{y}_i = y_i - h^T x(t_i;\theta_x). \tag{4.27}$$

Then the solution is

$$
\begin{aligned}
p(\tilde{y}_i|z_i,\theta_y) &= N(\mu(z_i;\theta_y),\sigma^2(z_i;\theta_y))(\tilde{y}_i) && (4.28)\\
p(z_i|\tilde{y}_i,\theta_y) &= \frac{p(z_i,\tilde{y}_i,\theta_y)}{p(\tilde{y}_i,\theta_y)} && (4.29)\\
&= \frac{p(z_i,\tilde{y}_i,\theta_y)}{\sum_{z_i'} p(z_i',\tilde{y}_i,\theta_y)} && (4.30)\\
&= \frac{p(\tilde{y}_i|z_i,\theta_y)p(z_i|\theta_y)p(\theta_y)}{\sum_{z_i'} p(\tilde{y}_i|z_i',\theta_y)p(z_i'|\theta_y)p(\theta_y)} && (4.31)\\
&= \frac{p(\tilde{y}_i|z_i,\theta_y)p(z_i|\theta_y)}{\sum_{z_i'} p(\tilde{y}_i|z_i',\theta_y)p(z_i'|\theta_y)} && (4.32)\\
&= \frac{p(\tilde{y}_i|z_i,\theta_y)q_{z_i}}{\sum_{z_i'} p(\tilde{y}_i|z_i',\theta_y)q_{z_i'}} && (4.33)\\
q_z &= \frac{1}{N_y}\sum_{i=1}^{N_y} p(z|\tilde{y}_i,{}_0\theta_y) && (4.34)\\
\mu_z &= \frac{1}{\sum_{i=1}^{N_y} p(z|\tilde{y}_i,{}_0\theta_y)}\sum_{i=1}^{N_y}\tilde{y}_i p(z|\tilde{y}_i,{}_0\theta_y) && (4.35)\\
\sigma_z^2 &= \frac{1}{\sum_{i=1}^{N_y} p(z|\tilde{y}_i,{}_0\theta_y)}\sum_{i=1}^{N_y}(\tilde{y}_i-\mu_z)^2 p(z|\tilde{y}_i,{}_0\theta_y) && (4.36)
\end{aligned}
$$

where $z \in \{1,2\}$; q_z, μ_z, and σ_z^2 are the new values of the θ_y parameters at the end of the iteration; $\theta_{x,0}$ are the old values of the parameters at the start of the iteration.

4.4 Update of θ_x: The Q function

In this section the algorithm for a joint update of θ_x and θ_y is derived. Then the algorithm for an update of just θ_x is the joint algorithm with the θ_y parameter held constant. Each iteration of an EM algorithm, e.g., the nth iteration, involves two steps: The first step is an expectation step, compute

$$Q(\theta|\hat{\theta}_{n-1},y) = \int_z \ln[p(y,z|\theta)]\, p(z|\hat{\theta}_{n-1},y)\mathrm{d}z$$

where the integral is over the unknown values of the hidden variables, and the second step is a maximization step, compute

$$\hat{\theta}_n = \arg\max_{\theta} Q(\theta|\hat{\theta}_{n-1}, y). \tag{4.37}$$

The idea in the expectation step is that the "additional measurements", which in fact are not available, are averaged out to yield an averaged cost. Then the idea in the maximization step is that the next value of $\hat{\theta}$ is picked by maximizing this averaged cost. In the problem of this thesis, the integral in the expectation step is actually a sum because the z_i variables are binary.

Standard conditional probability density function (pdf) calculations imply that

$$Q(\theta|\theta_0, y) = \int_z \ln\left[p(y,z|\theta)\right] p(z|\theta_0, y)\mathrm{d}z = \frac{1}{p(y|\theta_0)} \int_z \ln\left[p(y|z,\theta)p(z|\theta)\right] p(y|z,\theta_0)p(z|\theta_0)\mathrm{d}z. \tag{4.38}$$

Using the independence assumptions implies

$$\begin{aligned} Q(\theta|\theta_0, y) &= \frac{1}{p(y|\theta_0)} \int_z \ln\left\{ \left[\prod_{i=1}^{N_y} p(y_i|z_i,\theta)\right] \left[\prod_{i=1}^{N_y} p(z_i|\theta)\right] \right\} \times \\ &\times \left[\prod_{i=1}^{N_y} p(y_i|z_i,\theta_0)\right] \left[\prod_{i=1}^{N_y} p(z_i|\theta_0)\right] \mathrm{d}z. \end{aligned} \tag{4.39}$$

Using the logarithm to transform a product into a sum implies

$$\begin{aligned} Q(\theta|\theta_0, y) &= \frac{1}{p(y|\theta_0)} \sum_{i=1}^{N_y} \int_z \ln\left[p(y_i|z_i,\theta)p(z_i|\theta)\right] p(y_i|z_i,\theta_0)p(z_i|\theta_0) \times \\ &\times \left[\prod_{i'=1, i'\neq i}^{N_y} p(y_{i'}|z_{i'},\theta_0)p(z_{i'}|\theta_0)\right] \mathrm{d}z. \end{aligned} \tag{4.40}$$

The integral over z is a product of integrals over z_i which implies

$$\begin{aligned} Q(\theta|\theta_0, y) &= \frac{1}{p(y|\theta_0)} \sum_{i=1}^{N_y} \left\{ \left[\int_{z_i} \ln\left[p(y_i|z_i,\theta)p(z_i|\theta)\right] p(y_i|z_i,\theta_0)p(z_i|\theta_0)\mathrm{d}z_i\right] \times \right. \\ &\left. \times \prod_{i'=1, i'\neq i}^{N_y} \left[\int_{z_{i'}} p(y_{i'}|z_{i'},\theta_0)p(z_{i'}|\theta_0)\mathrm{d}z_{i'}\right] \right\}. \end{aligned} \tag{4.41}$$

Define

$$Q_i(\theta|\theta_0,y) = \frac{\int_{z_i} \ln\left[p(y_i|z_i,\theta)p(z_i|\theta)\right] p(y_i|z_i,\theta_0)p(z_i|\theta_0)\mathrm{d}z_i}{\int_{z_i} p(y_i|z_i,\theta_0)p(z_i|\theta_0)\mathrm{d}z_i}. \tag{4.42}$$

Using Eq. 4.42 in Eq. 4.41 implies

$$Q(\theta|\theta_0,y) = \sum_{i=1}^{N_y} Q_i(\theta|\theta_0,y). \tag{4.43}$$

In order to do a MAP instead of ML estimate, it is necessary to add the term $\ln p(\theta)$ to $Q(\theta|\theta_0,y)$.

An explicit form for Q_i is given by two numerator and one denominator terms as follows;

$$\begin{aligned} Qnum1 &= \left[-\frac{1}{2}\ln(2\pi) - \ln(\sigma_1) - \frac{1}{2}\frac{(y_i - h^T x(t_i;\theta_x) - \mu_1)^2}{\sigma_1^2} + \ln(q)\right] \times \\ &\times \quad N\left(h^T x(t_i;\theta_{x,0}) + \mu_{1,0}, \sigma_{1,0}^2\right)(y_i) q_0 \end{aligned} \tag{4.44}$$

$$\begin{aligned} Qnum2 &= \left[-\frac{1}{2}\ln(2\pi) - \ln(\sigma_2) - \frac{1}{2}\frac{(y_i - h^T x(t_i;\theta_x) - \mu_2)^2}{\sigma_2^2} + \ln(1-q)\right] \times \\ &\times \quad N\left(h^T x(t_i;\theta_{x,0}) + \mu_{2,0}, \sigma_{2,0}^2\right)(y_i)(1-q_0) \end{aligned} \tag{4.45}$$

Qden=

$$\left(h^T x(t_i;\theta_{x,0}) + \mu_{1,0}, \sigma_{1,0}^2\right)(y_i) q_0 + N\left(h^T x(t_i;\theta_{x,0}) + \mu_{2,0}, \sigma_{2,0}^2\right)(y_i)(1-q_0) \tag{4.46}$$

$$Q(\theta|\theta_0,y) = \frac{(Qnum1 + Qnum2)}{Qden}. \tag{4.47}$$

Maximization of Q with respect to θ_x requires maximization of the sum of the Q_i with respect to θ_x. There are additive terms that are not a function of θ_x so this is equivalent to maximizing

$$\tilde{Q}(\theta|\theta_0,y) = \sum_{i=1}^{N_y} \tilde{Q}_i(\theta|\theta_0,y) \tag{4.48}$$

with respect to θ_x where

$$\tilde{Q}_i(\theta|\theta_0,y) = \frac{\begin{array}{c} -\frac{1}{2}\frac{(y_i-h^T x(t_i;\theta_x)-\mu_1)^2}{\sigma_1^2} N\left(h^T x(t_i;\theta_{x,0})+\mu_{1,0},\sigma_{1,0}^2\right)(y_i) q_0 \\ -\frac{1}{2}\frac{(y_i-h^T x(t_i;\theta_x)-\mu_2)^2}{\sigma_2^2} N\left(h^T x(t_i;\theta_{x,0})+\mu_{2,0},\sigma_{2,0}^2\right)(y_i)(1-q_0) \end{array}}{N\left(h^T x(t_i;\theta_{x,0})+\mu_{1,0},\sigma_{1,0}^2\right)(y_i) q_0 + N\left(h^T x(t_i;\theta_{x,0})+\mu_{2,0},\sigma_{2,0}^2\right)(y_i)(1-q_0)} \tag{4.49}$$

where the only dependence on θ_x is in the two quadratice forms in the numerator.

4.5 Update of θ_x: Optimization of $\tilde{Q}$ with respect to θ_x

A variety of numerical optimization methods can be used to maximize Eqs. 4.48 and 4.49 with respect to θ_x when θ_y is assumed to be known. The derivatives of $x(t;\theta_x)$ with respect to the components of the parameter vector θ_x can be found by solving differential equations. Use the notation of Ref. [3, Appendix A.5, pp. 664–666]. In particular, if $f:\mathbb{R}^{n_x}\to\mathbb{R}$ then $\nabla_x f = (\partial f/\partial x_1,\cdots,\partial f/\partial x_{n_x})^T \in \mathbb{R}^{n_x}$ while if $f:\mathbb{R}^{n_x}\to\mathbb{R}^m$ then $\nabla_x f = (\nabla_x f_1|\ldots|\nabla_x f_m)\in\mathbb{R}^{n_x\times m}$. In this notation, the chain rule takes the following form: if $f:\mathbb{R}^k\to\mathbb{R}^m$, $g:\mathbb{R}^m\to\mathbb{R}^n$, and $h:\mathbb{R}^k\to\mathbb{R}^n$, $h(x)=g(f(x))$ then $\nabla h(x)=\nabla f(x)\nabla g(f(x))$. The PBPK models (Section 4.1) are of the following form: Let $\theta\in\mathbb{R}^{n_\theta}$, $x:\mathbb{R}\times\mathbb{R}^{n_\theta}\to\mathbb{R}^{n_x}$, $x_0:\mathbb{R}^{n_\theta}\to\mathbb{R}^{n_x}$, and $f:\mathbb{R}^{n_x}\times\mathbb{R}\times\mathbb{R}^{n_\theta}\to\mathbb{R}^{n_x}$. Then

$$\frac{\mathrm{d}x_\theta}{\mathrm{d}t}(t) = f(x_\theta(t),t,\theta) \tag{4.50}$$

$$x_\theta(0) = x_0(\theta). \tag{4.51}$$

Assuming mixed partial derivatives do not depend on the order of the derivatives and the validity of chain rule, it follows that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\left[\nabla_\theta x_\theta(t)\right] &= \nabla_\theta\left[\frac{\mathrm{d}}{\mathrm{d}t}x_\theta(t)\right] && (4.52)\\
&= \nabla_\theta\left[f(x_\theta(t),t,\theta)\right] && (4.53)\\
&= \left[\nabla_\theta x_\theta(t)\right]\left[\nabla_x f(x_\theta(t),t,\theta)\right]+\nabla_\theta f(x_\theta(t),t,\theta) && (4.54)\\
\nabla_\theta x_\theta(0) &= \nabla_\theta x_0(\theta). && (4.55)
\end{aligned}$$

Eq. 4.54 is a matrix-valued ordinary differential equation with matrix-valued initial condition given by Eq. 4.55. The right hand side of Eq. 4.54 depends on $x_\theta(t)$ so Eq. 4.54 (with its initial condition Eq. 4.55) must be solved simultaneously with Eq. 4.50 (with its initial condition Eq. 4.51). Higher order derivatives can be computed by the same method, although the answer cannot be written as simply in matrix-vector notation.

In our optimization, we use an algorithm that uses both the cost and the gradient of the cost where the cost is evaluated by solving for the Q function and the gradient is evaulated by computing the first derivatives of the state equations. To do that, it is sufficient to evaluate the first derivatives of the mass flux equations and use the additive property of the derivatives. Doing so, we need to keep in mind that the derivative of a ramp function is a unit step and the derivative of a unit-step is the impulse response. The full list of the first order derivatives can be found in the Appendix sections of this thesis.

The cost-gradient optimization is used for minimization so in our algorithm we use it to minimize $-Q$, which is equivalent to maximizing Q.

Computation of derivatives and using a cost-gradient based algorithm makes the problem more complicated, resulting in a $(3+3*6)=21$ state vector instead of a 3 state vector.

Algorithm

The initial condition on the PBPK parameters (θ_x) is set by the morphometrics estimate [20] and the initial condition on the measurement parameters (θ_y) is set by Indiana School of Medicine.. Then the algorithm iterates until convergence which is defined to be $\left|(\theta_x^i - \theta_x^{i-1})/\theta_x^{(i-1)}\right| < 0.0001$. Pseudocode is given in Algorithm 2.

Algorithm 2: **The generalized EM algorithm**

set the initial condition for θ_x from the morphometrics estimate [20].

set the initial condition for θ_y by collaborators.

while not converged **do**

update θ_x.

update θ_y

end while

4.6 Performance

The performance of the estimator can be studied before any data is collected to describe the ensemble average performance (so-called *a priori* performance) and after a specific set of data is collected to describe performance on that particular set of data (so-called *a posteriori* performance). In both cases, performance depends on the true values of the parameters and a natural source of estimates is morphometrics [20]. Symbolic formulas depend on a formula for the log likelihood function. From Section 4.2, the likelihood function is

$$p(y|\theta) = \prod_{i=1}^{N_y} \sum_{z\in\{1,2\}} q_z N(h^T x(t_i;\theta_x) + \mu_z, \sigma_z^2)(y_i) \tag{4.56}$$

where, to simplify notation, $q_1 = q$ and $q_2 = 1 - q$. This is a Gaussian mixture pdf where the mean has both a contribution from the PBPK model ($h^T x(t_i; \theta_x)$) and a contribution from the measurement model (μ_z).

A priori performance

The Cramér-Rao bound limits the performance of any unbiased estimator by the inequality

$$E[(\theta - \hat{\theta})(\theta - \hat{\theta})^T | \theta] \geq J^{-1} \tag{4.57}$$

where θ is the true value of the parameter vector, $\hat{\theta}$ is the estimated value of the parameter vector, and J is the Fisher Information matrix defined by

$$J = -E_y \left[\frac{\partial^2}{\partial \theta^2} \ln p(y|\theta) | \theta \right]. \tag{4.58}$$

While derivatives of the Gaussian mixture pdf are straightforward to compute (Section A), derivatives of the log pdf (Section A) are not because the Gaussian mixture pdf is not an exponential pdf. Therefore it is not possible to symbolically compute the expectation of the term $\frac{\partial^2}{\partial \theta^2} \ln p(y|\theta)$.

An alternative is to compute the expectation of the term $\frac{\partial^2}{\partial \theta^2} \ln p(y|\theta)$ by Monte Carlo: (1) A value of the parameters θ is selected. (2) The trajectory $x(\cdot; \theta_x)$ and its first and second derivatives with respect to θ_x is computed. (3) A number of trials for Monte Carlo, denoted by N_{MC}, is selected. (4)(a) For each trial indexed by m, a realization of $w_\cdot$ and $z_\cdot$ is computed. (4)(b) The second derivatives of the log likelihood are evaluated and sums of m of these derivatives are updated.

The Cramér-Rao bound is a bound on the performance of any unbiased estimator. However, the particular estimator described in this paper might be biased and may not achieve the bound. Monte Carlo calculations can also provide the ensemble average

performance of this particular estimator: (1) A value of the parameters θ is selected. (2) The trajectory $x(\cdot;\theta_x)$ is computed. (3) A number of trials for Monte Carlo, denoted by N_{MC}, is selected. (4)(a) For each trial indexed by m, a realization of $w_\cdot$ and $z_\cdot$ is computed. (4)(b) Using the measurement equation (Eq. 4.24), synthetic data $y_\cdot$ is computed. (4)(c) The estimation algorithm of Section 4.5 is used to compute estimates $\hat{\theta}_m$. (4)(d) Measures of performance are computed, e.g., $\theta - \hat{\theta}_m$ and $(\theta - \hat{\theta}_m)(\theta - \hat{\theta}_m)^T$, and sums over m of these measures of performance are updated.

A posteriori performance

This is the performance case that is important when the investigator is trying to evaluate whether a just-concluded experiment provided sufficiently precise parameters for a particular subject such that the parameters can be used to design further infusion protocols for that particular subject. To determine performance based on an individual trajectory, standard results in maximum likelihood estimation are used [10]. These results provide an approximation for the estimator error covariance matrix. Let y be the vector of data and θ be the vector of unknown parameters. Let the estimate of θ, which is a function of y, be denoted by $\hat{\theta}(y)$. Let the Hessian of the log likelihood function, the matrix of mixed second-order partial derivatives of the log likelihood function, be denoted by $H(\theta)$ with i,jth element defined by $\partial^2 \ln p(y|\theta)/\partial\theta_i\partial\theta_j$ where $p(y|\theta)$ is the conditional probability density function on the data y given the unknown parameters θ. Let θ_* be the true value of the parameters. The key result [10] is that the estimation error, $\hat{\theta}(y) - \theta_*$, is approximately Gaussian distributed with mean vector $\mathbf{0}$ and covariance matrix $-[H(\hat{\theta}(y))]^{-1}$. Formulas for the components of H are given by Eq. A.9.

Another approach to evaluating a posteriori performance is to run Monte Carlo calculations, using the parameter estimates provided by the ML estimator as if they

were the true parameters. New synthetic trajectories are created by using the values of θ_x^{ML}. By using θ_y^{ML}, new synthetic measurements are created by using the parameters as parameters of a Gaussian mixture noise and adding the noise to the trajectory path. Every iteration of a Monte Carlo run uses a different set of pseudo random variables, hence the measurements are different from iteration to iteration. Then the expectation-maximization algorithm is run by using the θ_x^{ML} and θ_y^{ML} as initial conditions. After a previously set number of iterations, the mean error is computed by $\bar{\theta}_x = \frac{1}{N}\sum_{i=1}^{N} \theta_x^{MC}(i) - \theta_x^{ML}$ and the error covariance matrix is computed by $cov(\theta_x) = \frac{1}{N}\sum_{i=1}^{N}(\theta_x^{MC}(i) - \bar{\theta}_x)(\theta_x^{MC}(i) - \bar{\theta}_x)^T$. The algorithm for such an approach is shown in Algorithm 3.

Algorithm 3: Hessian calculation via Monte Carlo Methods

set θ_x to $\theta_{x,ML}$

set θ_y to $\theta_{y,ML}$

set MC to number of Monte Carlo trials

calculate the $x(t;\theta_x)$ trajectory

while $MC > 0$ **do**

  create synthetic data using θ_y

  while not converged **do**

    update θ_x

    update θ_y

  end while

  $MC \leftarrow MC$-1

end while

calculate mean error

calculate error covariance matrix

4.7 Numerical results on estimation and estimation performance

In this section, the ideas presented in Section 4.6 will be used. For an a priori performance estimate Monte Carlo analysis will be used and for an a posteriori performance analysis the Hessian matrix will be computed. In this section the parameters and results are presented in the following format:

- θ_x=$k_{x,y}$,R_A,V_V,V_L,V_T,V_{max}.
- θ_y=q,μ_1,μ_2,σ_1^2,σ_2^2.

The optimization algorithm was run on synthetic data created with the following parameters:

- θ_x=$[0.3, 58.9, 106.23, 12.00, 264.71, 10.599]$
- θ_y=$[0.8, 0, -1, 0.01, 0.1]$

These parameters are the realistic values for this application [24]. Summary of the results of the experimental runs are shown in tables 4.1 and 4.2:

θ_x	$k_{x,y}$	R_A	V_V	V_L	V_T	V_{max}
10% change in $k_{x,y}$	0.3013	58.6219	107.1474	11.9915	266.8351	10.0868
10% change in V_L	0.3011	57.1343	106.0129	12.4335	278.2924	10.8152
$q = 0.6$	0.3142	51.6817	105.0074	10.6718	315.0049	11.0975

Table 4.1: Results for θ_x

An important result is that, contrary to what was originally thought, estimation of the liver volume did not cause any problems to the numerical optimization. It can also been seen that even without a good initial condition on the liver volume, we can still get

θ_y	q	μ_1	μ_2	σ_1^2	σ_2^2
10% change in $k_{x,y}$	0.6434	-0.5570	-0.3956	0.2146	0.0043
10% change in V_L	0.9574	-0.0087	-0.2660	0.0106	0.0003
$q = 0.6$	0.5062	-0.0689	-0.9195	0.0259	0.5366

Table 4.2: Results for θ_y

a good estimate of all the parameters. Another realization is that when the probability of a good measurement is lower, i.e. $q = 0.6$, the variance of a bad measurement is quite high($q = 0.5366$ rather than < 0.0001).

Using the estimated parameters, one can also generate plots of the three state trajectories and compare. The plots of the trajectories are shown in Figure 4.7.

Even though the parameter estimates were different than the initial parameters used, the state trajectories do not differ much. Most of the difference can be seen towards the end of the experiment, when the subjects are sobering up. This is an important result, since most of the phenotyping can occur while the subjects are sobering up. These results give us confidence in the expectation maximization algorithm and it can be used with real data.

In order to evaluate the performance, two approaches were taken. As discussed in Section 4.6, Hessian matrix approach or Monte Carlo simulations are used for maximum likelihood estimates.The performance of the estimator can be studied before any data is collected to describe the ensemble average performance (so-called *a priori* performance) and after a specific set of data is collected to describe performance on that particular set of data (so-called *a posteriori* performance). In both cases, performance depends on the true values of the parameters and a natural source of estimates is morphometrics [20]. Symbolic formulas depend on a formula for the log likelihood function.

From Section 4.2, the likelihood function is

$$p(y|\theta) = \prod_{i=1}^{N_y} \sum_{z\in\{1,2\}} q_z N(h^T x(t_i;\theta_x) + \mu_z, \sigma_z^2)(y_i) \tag{4.59}$$

where, to simplify notation, $q_1 = q$ and $q_2 = 1-q$. This is a Gaussian mixture pdf where the mean has both a contribution from the PBPK model ($h^T x(t_i;\theta_x)$) and a contribution from the measurement model (μ_z). Thus the two approaches are;

1. Realizing that the Hessian matrix is the second derivative of the log-likelihood evaluated at the maximum likelihood estimates and using finite difference equations on the first derivative of the log-likelihood.
2. Running Monte Carlo simulations using the maximum likelihood estimates as the tru parameter values.

In the first approach, due to the high non-linearity of the problem,the second derivative of the log-likelihood did not produce positive semi-definite matrices.

An example of a covariance matrix computed by the second method is shown below in table 4.3.

	$k_{x,y}$	R_A	V_V	V_L	V_T	V_{max}
$k_{x,y}$	0.0002	-0.0023	0.0045	-0.0007	0.0064	-0.0000
R_A	-0.0023	1.2707	1.3116	0.2717	-0.8772	-0.2602
V_V	0.0045	1.3116	3.6165	0.2768	1.9463	-0.4635
V_L	-0.0007	0.2717	0.2768	0.1843	-0.8096	-0.0831
V_T	0.0064	-0.8772	1.9463	-0.8096	13.8585	0.1171
V_{max}	-0.0000	-0.2602	-0.4635	-0.0831	0.1171	0.1248

Table 4.3: Covariance matrix computed using Monte Carlo simulations

4.8 Designing optimal inputs

Within limits set by safety and the performance of the infusion pump, the input used to excite the system can be optimized in order to get the best estimates of the parameters in the system. When designing the input, we focus on the estimator performance for θ_x, not θ_y since θ_x values are the goal of the entire experiment.

M_{Infuse} is the mass flow out of the infusion pump. The infusion pump has the following constraints:

- $M_{\text{min}}^{\text{pump}} \leq M_{\text{Infuse}}(t) \leq M_{\text{max}}^{\text{pump}}$, where $M_{\text{min}}^{\text{pump}}$ is $4mL/s$ and $M_{\text{max}}^{\text{pump}}$ is $1996mL/s$.
- $M_{\text{Infuse}}(t)$ is piece-wise constant with transitions every $30s$.
- The infused liquid is 6% ethanol by volume.

Safety requires the following additional constraints:

- $M_{\text{min}}^{\text{safety}} < M_{\text{Infuse}}(t)$ so that the intravenous line is continually in use.
- $M_{\text{Infuse}}(t) \leq M_{\text{max}}^{\text{safety}}$.
- $\int_t M_{\text{Infuse}}(t)\text{d}t \leq D_{\text{total}}$ where D_{total} is the maximum total dose, which is $120mg/deciliter$.

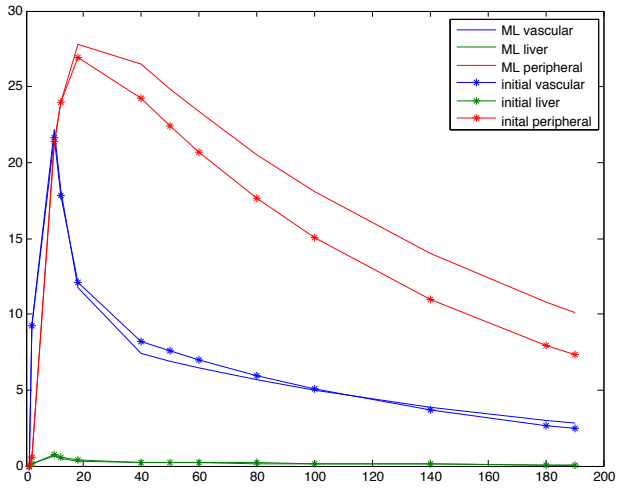
Because of the fixed duration piecewise-constant character of the pump,it is natural to think of the input as a sequence of characters from an alphabet where there is only one global constraint, which is the constraint on the total dose.

The goal of designing the experiment is to design an input which will yield an accurate estimate of θ_x. For any particular input, the quality of the estimate is measured

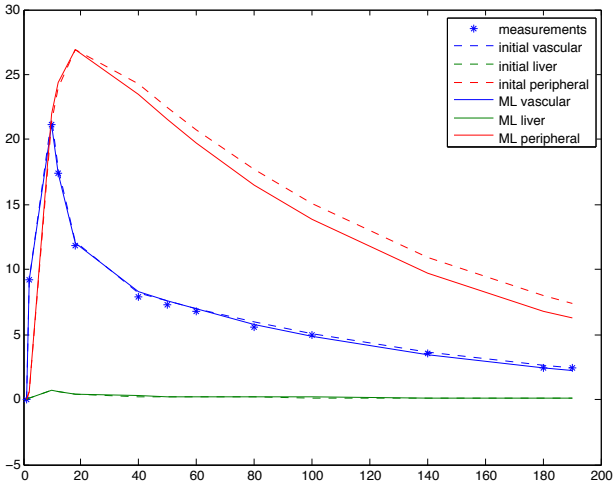
either by the sum of the variances for the elements of θ_x provided either by the Cramér-Rao bound, which applies to any unbiased estimate, or the Monte Carlo simulation of the estimator (both described in Section 4.6).

With the cost function of the previous paragraph and the idea of the input as a sequence of characters from an alphabet, a natural optimization technique is a genetic algorithm. In particular, the algorithm given at the end of Section 4.5, i.e. Algorithm 1 is used.

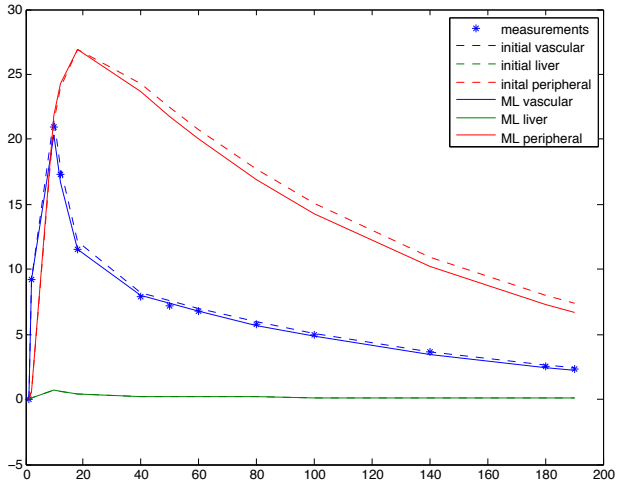
The trace of the Hessian is treated as the cost/fitness function and the inputs are treated as the individuals in a population. We started with an input based on [24] and modified it at each run to create different ”individuals” in the population. The genetic algorithm was then able to choose the ”fittest individual”.



(a)



(b)



(c)

Figure 4.1: (a) $q = 0.6$ (b) V_L is changed by 10% (c) $k_{x,y}$ is changed by 10%.

CHAPTER 5

CONCLUSION

Ethanol is a naturally produced drug used by humans for thousands of years because of its psychoactive properties. Ethanol can be devastating if used excessively. As was emphasised before, of people who use ethanol, nearly eight percent will become addicted during the course of their life, and about a third of them will die of complications attributable to the addiction.

It is extremely difficult to measure alcohol's psychoactive effects on brain. It is hard to measure brain alcohol concentration and PBPK is used to compute it from inputs and/or it is easier easier to measure concentration in the blood. It is also hard to quantify human behaviour; O'Connor and colleagues use video games for this purpose.

A major theme in alcoholism research is the influence of genetics on risk of addiction, e.g., [30]. Difficulty of phenotyping human use of alcohol in the community has shifted genetic studies to controlled laboratory environment. There are two different components to phenotyping subjects in laboratory settings: determining the ethanol dose to the brain and quantitatively phenotyping the subject's response.

O'Connor and colleagues have developed several experiments, that are essentially video games where the reward is alcohol, based on a progressive work paradigm in order to measure quantitatively phenotype on aspect of the subjects's response specifically, how much effort a human subject is willing to invest in order to get alcohol. In a progressive work game, the first dose of alcohol (given by intravenous infusion) requires relatively little work but the amount of work required to get successive doses progressively increases. Using the ideas of [24], the dose can be normalized to achieve brain alcohol concentration changes that are the same for all subjects independent of age, sex,

weight, *etc*.

In order to control the brain dosage of alcohol, physiologically-based pharmacokinetic models (PBPK) are used. In previous work [24] the authors describe ordinary differential equation models with two and three states which have different number of parameters. These models have been widely used with parameter values that are linear transformations of morphometrics (e.g., gender, age, height,weight) [20]. It is important to note that the models do not represent populations of subjects but rather individual subjects. This greatly constrains the complexity of the model and the type and amount of data that is available to identify the parameters in the model.

The pharmacokinetic data used to determine the parameters in the PBPK model comes from the following type of experiment: A solution of ethanol in normal saline is injected by intravenous infusion. The volume flow of the infusion is known. Periodically, at intervals of roughly two minutes, arterial ethanol concentration is measured by performing a breath-analyser measurement. A correctly performed measurement requires the subject to produce end-expiratory air. Especially when the subject is drunk, not every measurement is correctly performed. When a measurement is correctly performed, it is well documented that the measurement is equivalent to a arterial ethanol concentration measurement, which is one of the variables in the PBPK model [24].

The work done in this thesis focuses on two problems. The first problem(Chapter 3) is to develop a generative model (containing parameters with unknown values) of a simple game involving a progressive work paradigm along with the associated point-process signal processing that allows system identification of the model. The system developed is demonstrated on human subject data space.The same human subject playing the game under different circumstances, e.g., with and without a psychoactive drug, is assigned different parameter values.

The second problem (Chapter 4) is to design optimal inputs for the experiments used to determine the parameters in the PBPK model from which the brain dose of alcohol is computed. Mathematically, the model is constructed from non-linear ordinary differential equations. These equations are solved and optimized, by using their gradient, to formulate and refine parameter identification and control strategies. The Hessian information is then used to design an optimal input to the system.

5.1 The Analysis of Progressive-work Phenotyping Experiment

In light of the estimator performance described in Section 3.4, the differences in dynamical system parameters μ, ρ, and σ^2 described in Section 3.5 (Table 3.2) may be sufficient to allow the dynamical system parameter estimates to act as features in pattern recognition and clustering algorithms. In particular, $\hat{\rho}$ ranges over an order of magnitude which implies large differences in the characteristic time over which the x_k process, and therefore the λ_k rate process, is strongly correlated.

An attractive characteristic of this approach is that it provides information on the temporal dynamics of the rate function, e.g., $\hat{\rho}$. In different types of experiments, alcoholics are known to have different temporal character to their responses and we hope that this will be apparent in the parameter estimates.

An attractive feature of an ML estimator is that standard theory provides an estimate of the covariance of the difference between the parameter estimates and the true parameter values [10]. This estimate requires computation of the Hessian of the log likelihood at the parameter values that maximize the log likelihood. While this computation is not currently implemented in the system described in this paper, it will be a key component of statistical tests for whether differences in parameter values are significant. In

the future we will also investigate alternative ideas to compute the moments defined in Section 3.3, including particle filter ideas.

The values of μ, ρ, and σ^2 summarize the data: μ describes the average rate of button pushing, ρ describes how rapidly the instantaneous rate of button pushing changes with respect to time, and σ^2 describes the size of the changes. Potentially these features, or similar features from a more complex model, will be useful as the input to classifiers for distinguishing subjects who use alcohol in different ways and, through such classification, aid the selection of more appropriate therapy. Similarly, comparing these features when a subject is taking versus not taking a drug may be useful in the development of the drug and/or in the selection of an appropriate drug for therapy.

Using preliminary data, the resulting estimator computes parameter estimates with standard devations that are substantially smaller than the estimated values. Furthermore, standard hypothesis testing methods indicate that the performance of 3 out of 4 subjects is statistically significantly different in the "high" versus "low" reward data, indicating that his formulation has promise for detecting changes in the behaviour of subjects when the conditions change by a realistic amount.

5.2 The Design of Optimal Inputs for Determining the Parameters in PBPK Models

For the nonlinear model analysis, simulated data was used. However, the most important contribution of this algorithm is the automation that it provides. The good and the bad measurements are taken into account in one algorithm, without the need for a human input. The process of deciding if a measurement is good or bad and proceeding with

analysis is automated and has become faster.

The plots in Fig 4.1 can be thought of as a capacitor charging and discharging, which falls in line with our previous analogy that these PBPK models can be viewed as electrical circuits.

Monte Carlo algorithm computes errors, from which sample error covariance matrix is evaluated. Using such an algorithm ensures positivity of the error covariance matrix, which we were not able to achieve via computing the second order derivatives of the log-likelihood. We believe this is due to the imperfect numerical solutions to the differential equations. The Hessian matrix, evaluated from the sample error covariance matrix, is then used to design optimal inputs via the genetic algorithm.

APPENDIX A

RESULTS RELATED TO GAUSSIAN PDFS

The Gaussian pdf is

$$N(\mu,\sigma^2)(x) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right). \tag{A.1}$$

The various derivatives of $N(\mu,\sigma^2)(x)$ are

$$\frac{\partial}{\partial\mu}N(\mu,\sigma^2)(x) = N(\mu,\sigma^2)(x)\frac{1}{\sigma}\frac{x-\mu}{\sigma} \tag{A.2}$$

$$\frac{\partial^2}{\partial\mu^2}N(\mu,\sigma^2)(x) = N(\mu,\sigma^2)(x)\frac{1}{\sigma^2}\left[\frac{(x-\mu)^2}{\sigma^2}-1\right] \tag{A.3}$$

$$\frac{\partial}{\partial(\sigma^2)}N(\mu,\sigma^2)(x) = N(\mu,\sigma^2)(x)\frac{1}{2\sigma^2}\left[\frac{(x-\mu)^2}{\sigma^2}-1\right] \tag{A.4}$$

$$\frac{\partial^2}{\partial(\sigma^2)^2}N(\mu,\sigma^2)(x) = N(\mu,\sigma^2)(x)\frac{1}{4\sigma^4}\left\{\left[\frac{(x-\mu)^2}{\sigma^2}\right]^2 - 6\frac{(x-\mu)^2}{\sigma^2}+3\right\} \tag{A.5}$$

$$\frac{\partial^2}{\partial(\sigma^2)\partial\mu}N(\mu,\sigma^2)(x) = N(\mu,\sigma^2)(x)\frac{1}{2\sigma^3}\left\{\left[\frac{(x-\mu)^2}{\sigma^2}\right]^2 - 3\right\}\frac{x-\mu}{\sigma}. \tag{A.6}$$

Let $n(\mu,\sigma^2)(x) = \ln N(\mu,\sigma^2)(x)$. The various derivatives of $n(\mu,\sigma^2)(x)$ are available from Eqs. A.2–A.6 since

$$\frac{\partial}{\partial\theta}n(\mu,\sigma^2)(x) = \frac{\partial}{\partial\theta}\ln N(\mu,\sigma^2)(x) = \frac{1}{N(\mu,\sigma^2)(x)}\frac{\partial}{\partial\theta}N(\mu,\sigma^2)(x). \tag{A.7}$$

The Gaussian mixture with N_z classes has pdf

$$p(x|\theta) = \sum_{z=1}^{N_z}\alpha_z N(\mu_z,\sigma_z^2)(x) \tag{A.8}$$

where $\phi = (\alpha_1,\mu_1,\sigma_1^2,\ldots,\alpha_{N_z},\mu_{N_z},\sigma_{N_z}^2)$, $\alpha_z > 0$ for all $z \in \{1,\ldots,N_z\}$, and $\sum_{z=1}^{N_z}\alpha_z = 1$. Due to the additive structure of Eq. A.8, many of the derivatives of $p(x|\theta)$ are zero, in particular, Let $\rho(x|\theta) = \ln p(x|\theta)$. The various derivatives of $\rho(x|\theta)$ can be computed from

$$\frac{\partial}{\partial\theta}\rho(x|\theta) = \frac{\partial}{\partial\theta}\ln p(x|\theta) = \frac{1}{p(x|\theta)}\frac{\partial}{\partial\theta}p(x|\theta) \tag{A.9}$$

and Eqs. A.8–**??**.

APPENDIX B

STATEMENT OF MODEL

$$
\begin{aligned}
y_i &= h^T x(t_i;\theta_x) + \sigma(z_i;\theta_y) w_i + \mu(z_i;\theta_y) \\
&\quad w_\cdot \text{ and } z_\cdot \text{ are independent sequences} \\
h &\in \mathbb{R}^n, \text{ known} \\
\theta_x &\doteq \text{PBPK model parameter vector} \\
x(\cdot;\theta_x) &\doteq \text{PBPK model state trajectory with parameter vector } \theta_x \text{ and known input} \\
\theta_y &\doteq (q, \mu_1, \sigma_1, \mu_2, \sigma_2) \quad \text{class probability and mean and variance for each class} \\
z_i &\doteq \text{class labels} \\
z_i &\in \{1,2\} \\
z_i &\sim \text{i.i.d. with } P(z_i = 1) = q \\
\sigma(z_i,\theta_y) &\doteq \sigma_{z_i} \quad (\text{standard deviation of the noise}) \\
\mu(z_i,\theta_y) &\doteq \mu_{z_i} \quad (\text{mean of the noise}) \\
w_i &\sim \text{i.i.d. } N(0,1) \\
\theta &\doteq (\theta_x, \theta_y) \\
y &\doteq (y_1, \ldots, y_{N_y}) \\
z &\doteq (z_1, \ldots, z_{N_y}).
\end{aligned}
$$

APPENDIX C

DERIVING THE LIKELIHOOD

$$
\begin{aligned}
p(y|z,\theta) &= \prod_{i=i}^{N_y} N\left(h^T x(t_i;\theta_x) + \mu(z_i;\theta_y), \sigma^2(z_i;\theta_y)\right)(y_i) \\
p(z|\theta_y) &= p(z|\theta) \\
&= \prod_{i=i}^{N_y} [q\delta_{z_i,1} + (1-q)\delta_{z_i,2}] \\
p(y,z|\theta) &= \frac{p(y,z,\theta)}{p(\theta)} \\
&= \frac{p(y|z,\theta)p(z|\theta)p(\theta)}{p(\theta)} \\
&= p(y|z,\theta)p(z|\theta) \\
p(y|\theta) &= \sum_z p(y,z|\theta) \\
&= \sum_z p(y|z,\theta)p(z|\theta) \\
&= \sum_z \left[\prod_{i=i}^{N_y} N\left(h^T x(t_i;\theta_x) + \mu(z_i;\theta_y), \sigma^2(z_i;\theta_y)\right)(y_i)\right]\left[\prod_{i=i}^{N_y} [q\delta_{z_i,1} + (1-q)\delta_{z_i,2}]\right] \\
&= \sum_z \left[\prod_{i=i}^{N_y} N\left(h^T x(t_i;\theta_x) + \mu(z_i;\theta_y), \sigma^2(z_i;\theta_y)\right)(y_i)\,[q\delta_{z_i,1} + (1-q)\delta_{z_i,2}]\right] \\
&= \left[\prod_{i=1}^{N_y} \sum_{z_i\in\{1,2\}}\right]\left[\prod_{i=i}^{N_y} N\left(h^T x(t_i;\theta_x) + \mu(z_i;\theta_y), \sigma^2(z_i;\theta_y)\right)(y_i)\,[q\delta_{z_i,1} + (1-q)\delta_{z_i,2}]\right] \\
&= \prod_{i=i}^{N_y} \sum_{z_i\in\{1,2\}} \left\{N\left(h^T x(t_i;\theta_x) + \mu(z_i;\theta_y), \sigma^2(z_i;\theta_y)\right)(y_i)\,[q\delta_{z_i,1} + (1-q)\delta_{z_i,2}]\right\}.
\end{aligned}
$$

APPENDIX D

DETAILS OF THE DERIVATION OF Q

Since

$$\begin{aligned}
p(y,z|\theta) &= \frac{p(y,z,\theta)}{p(\theta)} \\
&= \frac{p(y|z,\theta)p(z|\theta)p(\theta)}{p(\theta)} \\
&= p(y|z,\theta)p(z|\theta) \qquad \text{(D.1)} \\
p(z|\theta_0,y) &= \frac{p(z,\theta_0,y)}{p(y,\theta_0)} \\
&= \frac{p(y|z,\theta_0)p(z|\theta_0)p(\theta_0)}{p(y|\theta_0)p(\theta_0)} \\
&= \frac{p(y|z,\theta_0)p(z|\theta_0)}{p(y|\theta_0)} \qquad \text{(D.2)} \\
p(y|\theta) &= \int_z p(y,z|\theta)\mathrm{d}z \\
&= \int_z \frac{p(y,z,\theta)}{p(\theta)}\mathrm{d}z \\
&= \int_z \frac{p(y|z,\theta)p(z|\theta)p(\theta)}{p(\theta)}\mathrm{d}z \\
&= \int_z p(y|z,\theta)p(z|\theta)\mathrm{d}z \\
&= \int_z \left[\prod_{i=1}^{N_y} p(y_i|z_i,\theta)\right]\left[\prod_{i=1}^{N_y} p(z_i|\theta)\right]\mathrm{d}z \\
&= \prod_{i=1}^{N_y}\left[\int_{z_i} p(y_i|z_i,\theta)p(z_i)\mathrm{d}z_i\right] \qquad \text{(D.3)} \\
p(y_i|\theta) &= \int_{z_i} p(y_i,z_i|\theta)\mathrm{d}z_i \\
&= \int_{z_i} p(y_i,z_i,\theta)/p(\theta)\mathrm{d}z_i \\
&= \int_{z_i} p(y_i|z_i,\theta)p(z_i|\theta)p(\theta)/p(\theta)\mathrm{d}z_i \\
&= \int_{z_i} p(y_i|z_i,\theta)p(z_i|\theta)\mathrm{d}z_i \qquad \text{(D.4)}
\end{aligned}$$

$$
\begin{aligned}
Q(\theta|\theta_0,y) &= \int_z \ln\left[p(y,z|\theta)\right] p(z|\theta_0,y)\mathrm{d}z \\
&= \int_z \ln\left[p(y|z,\theta)p(z|\theta)\right] \frac{p(y|z,\theta_0)p(z|\theta_0)}{p(y|\theta_0)}\mathrm{d}z \\
&\quad \text{by Eqs. D.1 and D.2} \\
&= \frac{1}{p(y|\theta_0)} \int_z \ln\left[p(y|z,\theta)p(z|\theta)\right] p(y|z,\theta_0)p(z|\theta_0)\mathrm{d}z \\
&= \frac{1}{p(y|\theta_0)} \int_z \ln\left\{ \left[\prod_{i=1}^{N_y} p(y_i|z_i,\theta)\right] \left[\prod_{i=1}^{N_y} p(z_i|\theta)\right] \right\} \times \\
&\times \left[\prod_{i=1}^{N_y} p(y_i|z_i,\theta_0)\right] p(z|\theta_0)\mathrm{d}z \\
&= \frac{1}{p(y|\theta_0)} \int_z \ln\left\{ \left[\prod_{i=1}^{N_y} p(y_i|z_i,\theta)\right] \left[\prod_{i=1}^{N_y} p(z_i|\theta)\right] \right\} \times \\
&\times \left[\prod_{i=1}^{N_y} p(y_i|z_i,\theta_0)\right] \left[\prod_{i=1}^{N_y} p(z_i|\theta_0)\right] \mathrm{d}z \\
&= \frac{1}{p(y|\theta_0)} \int_z \left\{ \sum_{i=1}^{N_y} \ln\left[p(y_i|z_i,\theta)p(z_i|\theta)\right] \right\} \times \\
&\times \left[\prod_{i'=1}^{N_y} p(y_{i'}|z_{i'},\theta_0)p(z_{i'}|\theta_0)\right] \mathrm{d}z \\
&= \frac{1}{p(y|\theta_0)} \sum_{i=1}^{N_y} \int_z \ln\left[p(y_i|z_i,\theta)p(z_i|\theta)\right] \times \\
&\times \left[\prod_{i'=1}^{N_y} p(y_{i'}|z_{i'},\theta_0)p(z_{i'}|\theta_0)\right] \mathrm{d}z \\
&= \frac{1}{p(y|\theta_0)} \sum_{i=1}^{N_y} \int_z \ln\left[p(y_i|z_i,\theta)p(z_i|\theta)\right] p(y_i|z_i,\theta_0)p(z_i|\theta_0) \times \\
&\times \left[\prod_{i'=1,i'\neq i}^{N_y} p(y_{i'}|z_{i'},\theta_0)p(z_{i'}|\theta_0)\right] \mathrm{d}z \\
&= \frac{1}{p(y|\theta_0)} \sum_{i=1}^{N_y} \left\{ \left[\int_{z_i} \ln\left[p(y_i|z_i,\theta)p(z_i|\theta)\right] p(y_i|z_i,\theta_0)p(z_i|\theta_0)\mathrm{d}z_i\right] \times \right. \\
&\quad \left. \times \prod_{i'=1,i'\neq i}^{N_y} \left[\int_{z_{i'}} p(y_{i'}|z_{i'},\theta_0)p(z_{i'}|\theta_0)\mathrm{d}z_{i'}\right] \right\}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{p(y|\theta_0)} \sum_{i=1}^{N_y} \left\{ \left[\int_{z_i} \ln\left[p(y_i|z_i,\theta)p(z_i|\theta)\right] p(y_i|z_i,\theta_0)p(z_i|\theta_0)\mathrm{d}z_i \right] \times \right. \\
&\quad \left. \times \frac{\prod_{i'=1}^{N_y} \int_{z_{i'}} p(y_{i'}|z_{i'},\theta_0)p(z_{i'}|\theta_0)\mathrm{d}z_{i'}}{\int_{z_i} p(y_i|z_i,\theta_0)p(z_i|\theta_0)\mathrm{d}z_i} \right\} \\
&= \frac{1}{p(y|\theta_0)} \sum_{i=1}^{N_y} \left\{ \left[\int_{z_i} \ln\left[p(y_i|z_i,\theta)p(z_i|\theta)\right] p(y_i|z_i,\theta_0)p(z_i|\theta_0)\mathrm{d}z_i \right] \times \right. \\
&\quad \left. \times \frac{p(y|\theta_0)}{\int_{z_i} p(y_i|z_i,\theta_0)p(z_i|\theta_0)\mathrm{d}z_i} \right\} \quad \text{by Eq. D.3} \\
&= \sum_{i=1}^{N_y} \left\{ \left[\int_{z_i} \ln\left[p(y_i|z_i,\theta)p(z_i|\theta)\right] p(y_i|z_i,\theta_0)p(z_i|\theta_0)\mathrm{d}z_i \right] \times \right. \\
&\quad \left. \times \frac{1}{\int_{z_i} p(y_i|z_i,\theta_0)p(z_i|\theta_0)\mathrm{d}z_i} \right\} \\
&= \sum_{i=1}^{N_y} \frac{1}{p(y_i|\theta_0)} \int_{z_i} \ln\left[p(y_i|z_i,\theta)p(z_i|\theta)\right] p(y_i|z_i,\theta_0)p(z_i|\theta_0)\mathrm{d}z_i
\end{aligned}
$$

by Eq. D.4 with θ replaced by θ_0.

$Q_i(\theta|\theta_0,y)$ is defined in Eq. 4.42. Using this definition gives the formula for $Q(\theta|\theta_0,y)$ shown in Eq. 4.43. In order to do a MAP instead of ML estimate, it is necessary to add the term $\ln p(\theta)$ to $Q(\theta|\theta_0,y)$.

In order to compute Q_i, the following calculations are necessary:

$$
\begin{aligned}
Q_i(\theta|\theta_0,y) &= \frac{\int_{z_i} \ln\left[p(y_i|z_i,\theta)p(z_i|\theta)\right] p(y_i|z_i,\theta_0)p(z_i|\theta_0)\mathrm{d}z_i}{\int_{z_i} p(y_i|z_i,\theta_0)p(z_i|\theta_0)\mathrm{d}z_i} \\
&= \frac{\sum_{z_i\in\{1,2\}} \begin{array}{l} \ln\left[N\left(h^T x(t_i;\theta_x)+\mu(z_i;\theta_y),\sigma^2(z_i;\theta_y)\right)(y_i)\left[q\delta_{z_i,1}+(1-q)\delta_{z_i,2}\right]\right] \times \\ \times N\left(h^T x(t_i;\theta_{x,0})+\mu(z_i;\theta_{y,0}),\sigma^2(z_i;\theta_{y,0})\right)(y_i)\left[q_0\delta_{z_i,1}+(1-q_0)\delta_{z_i,2}\right] \end{array}}{\sum_{z_i\in\{1,2\}} N\left(h^T x(t_i;\theta_{x,0})+\mu(z_i;\theta_{y,0}),\sigma^2(z_i;\theta_{y,0})\right)(y_i)\left[q_0\delta_{z_i,1}+(1-q_0)\delta_{z_i,2}\right]} \\
&= \frac{\begin{array}{l} \ln\left[N\left(h^T x(t_i;\theta_x)+\mu_1,\sigma_1^2\right)(y_i)q\right] N\left(h^T x(t_i;\theta_{x,0})+\mu_{1,0},\sigma_{1,0}^2\right)(y_i)q_0 \\ +\ln\left[N\left(h^T x(t_i;\theta_x)+\mu_2,\sigma_2^2\right)(y_i)(1-q)\right] N\left(h^T x(t_i;\theta_{x,0})+\mu_{2,0},\sigma_{2,0}^2\right)(y_i)(1-q_0) \end{array}}{N\left(h^T x(t_i;\theta_{x,0})+\mu_{1,0},\sigma_{1,0}^2\right)(y_i)q_0 + N\left(h^T x(t_i;\theta_{x,0})+\mu_{2,0},\sigma_{2,0}^2\right)(y_i)(1-q_0)}
\end{aligned}
$$

where the notation is

$$\begin{aligned}
\theta &\doteq (\theta_x, \theta_y) \\
\theta_0 &\doteq (\theta_{x,0}, \theta_{y,0}) \\
\theta_y &\doteq (q, \mu_1, \sigma_1, \mu_2, \sigma_2) \\
\theta_{y,0} &\doteq (q_0, \mu_{1,0}, \sigma_{1,0}, \mu_{2,0}, \sigma_{2,0}).
\end{aligned}$$

Since

$$\begin{aligned}
N(m,\sigma^2)(y) &\doteq \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\frac{(y-m)^2}{\sigma^2}\right) \\
\ln N(m,\sigma^2)(y) &= -\frac{1}{2}\ln(2\pi) - \ln(\sigma) - \frac{1}{2}\frac{(y-m)^2}{\sigma^2} \\
\ln\left(N(m,\sigma^2)(y)q\right) &= -\frac{1}{2}\ln(2\pi) - \ln(\sigma) - \frac{1}{2}\frac{(y-m)^2}{\sigma^2} + \ln(q)
\end{aligned}$$

it follows that

$$Q_i(\theta|\theta_0,y) = \frac{Q_i(\theta|\theta_0,y)_{num1} + Q_i(\theta|\theta_0,y)_{num2}}{Q_i(\theta|\theta_0,y)_{denom}}$$

where

$$\begin{aligned}
Q_i(\theta|\theta_0,y)_{num1} &= \left[-\frac{1}{2}\ln(2\pi) - \ln(\sigma_1) - \frac{1}{2}\frac{(y_i - h^T x(t_i;\theta_x) - \mu_1)^2}{\sigma_1^2} + \ln(q)\right] \times \\
&\times \; N\left(h^T x(t_i;\theta_{x,0}) + \mu_{1,0}, \sigma_{1,0}^2\right)(y_i) q_0
\end{aligned}$$

$$\begin{aligned}
Q_i(\theta|\theta_0,y)_{num2} &= \left[-\frac{1}{2}\ln(2\pi) - \ln(\sigma_2) - \frac{1}{2}\frac{(y_i - h^T x(t_i;\theta_x) - \mu_2)^2}{\sigma_2^2} + \ln(1-q)\right] \times \\
&\times \; N\left(h^T x(t_i;\theta_{x,0}) + \mu_{2,0}, \sigma_{2,0}^2\right)(y_i)(1-q_0)
\end{aligned}$$

$$\begin{aligned}
Q_i(\theta|\theta_0,y)_{denom} &= N\left(h^T x(t_i;\theta_{x,0}) + \mu_{1,0}, \sigma_{1,0}^2\right)(y_i) q_0 \\
&+ \; N\left(h^T x(t_i;\theta_{x,0}) + \mu_{2,0}, \sigma_{2,0}^2\right)(y_i)(1-q_0)
\end{aligned}$$

APPENDIX E

PROPAGATION OF SECOND DERIVATIVES

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\partial}{\partial(\theta_x)_j}\nabla_\theta x_\theta(t)\right] \\
&= \frac{\partial}{\partial(\theta_x)_j}\frac{\mathrm{d}}{\mathrm{d}t}\left[\nabla_\theta x_\theta(t)\right] && \text{(E.1)}\\
&= \frac{\partial}{\partial(\theta_x)_j}\left\{\left[\nabla_\theta x_\theta(t)\right]\left[\nabla_x f(x_\theta(t),t,\theta)\right]+\nabla_\theta f(x_\theta(t),t,\theta)\right\} && \text{(E.2)}\\
&= \left[\frac{\partial}{\partial(\theta_x)_j}\nabla_\theta x_\theta(t)\right]\left[\nabla_x f(x_\theta(t),t,\theta)\right]+\left[\nabla_\theta x_\theta(t)\right]\left[\frac{\partial}{\partial(\theta_x)_j}\nabla_x f(x_\theta(t),t,\theta)\right] \\
&\quad+\frac{\partial}{\partial(\theta_x)_j}\nabla_\theta f(x_\theta(t),t,\theta). && \text{(E.3)}
\end{aligned}
$$

On the second and third terms, it is necessary to account for direct dependence of f on θ but also indirect dependence of f on θ via x_θ. The notation being using is not really adequate but basically,

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\partial}{\partial(\theta_x)_j}\nabla_\theta x_\theta(t)\right] \\
&= \left[\frac{\partial}{\partial(\theta_x)_j}\nabla_\theta x_\theta(t)\right]\left[\nabla_x f(x_\theta(t),t,\theta)\right] \\
&\quad+\left[\nabla_\theta x_\theta(t)\right]\left[\sum_{k=1}^{n_x}\left(\frac{\partial}{\partial(x_\theta(t))_k}\nabla_x f(x_\theta(t),t,\theta)\right)\frac{\partial(x_\theta(t))_k}{\partial(\theta_x)_j}+\sum_{k=1}^{n_\theta}\left(\frac{\partial}{\partial(\theta_x)_k}\nabla_x f(x_\theta(t),t,\theta)\right)\frac{\partial(\theta_x)_k}{\partial(\theta_x)_j}\right] \\
&\quad+\left[\sum_{k=1}^{n_x}\left(\frac{\partial}{\partial(x_\theta(t))_k}\nabla_\theta f(x_\theta(t),t,\theta)\right)\frac{\partial(x_\theta(t))_k}{\partial(\theta_x)_j}+\sum_{k=1}^{n_\theta}\left(\frac{\partial}{\partial(\theta_x)_k}\nabla_\theta f(x_\theta(t),t,\theta)\right)\frac{\partial(\theta_x)_k}{\partial(\theta_x)_j}\right] && \text{(E.4)}\\
&= \left[\frac{\partial}{\partial(\theta_x)_j}\nabla_\theta x_\theta(t)\right]\left[\nabla_x f(x_\theta(t),t,\theta)\right] \\
&\quad+\left[\nabla_\theta x_\theta(t)\right]\left[\sum_{k=1}^{n_x}\left(\frac{\partial}{\partial(x_\theta(t))_k}\nabla_x f(x_\theta(t),t,\theta)\right)\frac{\partial(x_\theta(t))_k}{\partial(\theta_x)_j}+\left(\frac{\partial}{\partial(\theta_x)_j}\nabla_x f(x_\theta(t),t,\theta)\right)\right] \\
&\quad+\left[\sum_{k=1}^{n_x}\left(\frac{\partial}{\partial(x_\theta(t))_k}\nabla_\theta f(x_\theta(t),t,\theta)\right)\frac{\partial(x_\theta(t))_k}{\partial(\theta_x)_j}+\left(\frac{\partial}{\partial(\theta_x)_j}\nabla_\theta f(x_\theta(t),t,\theta)\right)\right]. && \text{(E.5)}
\end{aligned}
$$

The initial condition is

$$
\frac{\partial}{\partial(\theta_x)_j}\nabla_\theta x_\theta(0)=\frac{\partial}{\partial(\theta_x)_j}\nabla_\theta x_0(\theta). \qquad \text{(E.6)}
$$

APPENDIX F

DERIVATIVES

In order to compute

$$\frac{\partial \mu_{\mathscr{X}}}{\partial (\theta_x)_i} \tag{F.1}$$

it is necessary and sufficient to compute

$$\frac{\partial M}{\partial (\theta_x)_i} \tag{F.2}$$

The parameter vector is denoted by θ_x (the subscript x is because these parameters influence the PBPK model, not the measurement model). The ith component of θ_x is denoted by $(\theta_x)_i$.

F.0.1 $M_{\text{Metab}}(t)$ (Eq. 4.4)

$$
\begin{aligned}
M_{\text{Metab}}(t) &= V_{\mathscr{L}} V_{\max} \frac{\mu_{\mathscr{L}}(t)/V_{\mathscr{L}}}{K_m + \mu_{\mathscr{L}}(t)/V_{\mathscr{L}}} \\
&= V_{\max} \frac{\mu_{\mathscr{L}}(t)}{K_m + \mu_{\mathscr{L}}(t)/V_{\mathscr{L}}} \\
&= V_{\max} \mu_{\mathscr{L}}(t) \left(K_m + \mu_{\mathscr{L}}(t)/V_{\mathscr{L}}\right)^{-1} \\
\frac{\partial M_{\text{Metab}}}{\partial V_{\mathscr{L}}}(t) &= V_{\max} \frac{\partial \mu_{\mathscr{L}}}{\partial V_{\mathscr{L}}}(t) \left(K_m + \mu_{\mathscr{L}}(t)/V_{\mathscr{L}}\right)^{-1} \\
&\quad - V_{\max} \mu_{\mathscr{L}}(t) \left(K_m + \mu_{\mathscr{L}}(t)/V_{\mathscr{L}}\right)^{-2} \frac{\partial \mu_{\mathscr{L}}(t)/V_{\mathscr{L}}}{\partial V_{\mathscr{L}}} \\
&= V_{\max} \frac{\partial \mu_{\mathscr{L}}}{\partial V_{\mathscr{L}}}(t) \left(K_m + \mu_{\mathscr{L}}(t)/V_{\mathscr{L}}\right)^{-1} \\
&\quad - V_{\max} \mu_{\mathscr{L}}(t) \left(K_m + \mu_{\mathscr{L}}(t)/V_{\mathscr{L}}\right)^{-2} \frac{V_{\mathscr{L}} \frac{\partial \mu_{\mathscr{L}}(t)}{\partial V_{\mathscr{L}}} - \mu_{\mathscr{L}}(t) \frac{\partial V_{\mathscr{L}}}{\partial V_{\mathscr{L}}}}{V_{\mathscr{L}}^2} \\
&= V_{\max} \frac{\partial \mu_{\mathscr{L}}}{\partial V_{\mathscr{L}}}(t) \left(K_m + \mu_{\mathscr{L}}(t)/V_{\mathscr{L}}\right)^{-1} \\
&\quad - V_{\max} \mu_{\mathscr{L}}(t) \left(K_m + \mu_{\mathscr{L}}(t)/V_{\mathscr{L}}\right)^{-2} \frac{V_{\mathscr{L}} \frac{\partial \mu_{\mathscr{L}}(t)}{\partial V_{\mathscr{L}}} - \mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}^2} \\
\frac{\partial M_{\text{Metab}}}{\partial V_{\max}}(t) &= \mu_{\mathscr{L}}(t) \left(K_m + \mu_{\mathscr{L}}(t)/V_{\mathscr{L}}\right)^{-1} \\
&\quad + V_{\max} \frac{\partial \mu_{\mathscr{L}}}{\partial V_{\max}}(t) \left(K_m + \mu_{\mathscr{L}}(t)/V_{\mathscr{L}}\right)^{-1} \\
&\quad - V_{\max} \mu_{\mathscr{L}}(t) \left(K_m + \mu_{\mathscr{L}}(t)/V_{\mathscr{L}}\right)^{-2} \frac{1}{V_{\mathscr{L}}} \frac{\partial \mu_{\mathscr{L}}}{\partial V_{\max}}(t) \\
\frac{\partial M_{\text{Metab}}}{\partial (\theta_x)_i}(t) &= V_{\max} \frac{\partial \mu_{\mathscr{L}}}{\partial (\theta_x)_i}(t) \left(K_m + \mu_{\mathscr{L}}(t)/V_{\mathscr{L}}\right)^{-1} \\
&\quad - V_{\max} \mu_{\mathscr{L}}(t) \left(K_m + \mu_{\mathscr{L}}(t)/V_{\mathscr{L}}\right)^{-2} \frac{1}{\partial V_{\mathscr{L}}} \frac{\partial \mu_{\mathscr{L}}}{\partial (\theta_x)_i}(t)
\end{aligned}
$$

where Eq. F.3 applies for $(\theta_x)_i \in \{V_{\mathscr{T}}, V_{\mathscr{V}}, k, R_{\text{A}}\}$.

F.0.2 $M_{\mathrm{P},\mathscr{T}}(t)$ (Eq. 4.5)

$$
\begin{aligned}
M_{\mathrm{P},\mathscr{T}}(t) &= k_{\mathrm{P},\mathscr{T}} R_{\mathrm{A}} (1-F_{\mathrm{L}}) r \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} - \frac{\mu_{\mathscr{T}}(t)}{V_{\mathscr{T}}} \right) \\
\frac{\partial M_{\mathrm{P},\mathscr{T}}}{\partial R_{\mathrm{A}}}(t) &= k_{\mathrm{P},\mathscr{T}} (1-F_{\mathrm{L}}) r \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} - \frac{\mu_{\mathscr{T}}(t)}{V_{\mathscr{T}}} \right) \\
&\quad + k_{\mathrm{P},\mathscr{T}} R_{\mathrm{A}} (1-F_{\mathrm{L}}) u \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} - \frac{\mu_{\mathscr{T}}(t)}{V_{\mathscr{T}}} \right) \left(\frac{1}{V_{\mathscr{V}}} \frac{\partial \mu_{\mathscr{V}}}{\partial R_{\mathrm{A}}}(t) - \frac{1}{V_{\mathscr{T}}} \frac{\partial \mu_{\mathscr{T}}}{\partial R_{\mathrm{A}}}(t) \right) \\
\frac{\partial M_{\mathrm{P},\mathscr{T}}}{\partial k_{\mathrm{P},\mathscr{T}}}(t) &= R_{\mathrm{A}} (1-F_{\mathrm{L}}) r \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} - \frac{\mu_{\mathscr{T}}(t)}{V_{\mathscr{T}}} \right) \\
&\quad + k_{\mathrm{P},\mathscr{T}} R_{\mathrm{A}} (1-F_{\mathrm{L}}) u \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} - \frac{\mu_{\mathscr{T}}(t)}{V_{\mathscr{T}}} \right) \left(\frac{1}{V_{\mathscr{V}}} \frac{\partial \mu_{\mathscr{V}}}{\partial k_{\mathrm{P},\mathscr{T}}}(t) - \frac{1}{V_{\mathscr{T}}} \frac{\partial \mu_{\mathscr{T}}}{\partial k_{\mathrm{P},\mathscr{T}}}(t) \right) \\
\frac{\partial M_{\mathrm{P},\mathscr{T}}}{\partial V_{\mathscr{V}}}(t) &= k_{\mathrm{P},\mathscr{T}} R_{\mathrm{A}} (1-F_{\mathrm{L}}) u \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} - \frac{\mu_{\mathscr{T}}(t)}{V_{\mathscr{T}}} \right) \left(\frac{1}{V_{\mathscr{V}}} \frac{\partial \mu_{\mathscr{V}}}{\partial V_{\mathscr{V}}}(t) - \frac{1}{V_{\mathscr{V}}^2} \mu_{\mathscr{V}}(t) - \frac{1}{V_{\mathscr{T}}} \frac{\partial \mu_{\mathscr{T}}}{\partial V_{\mathscr{V}}}(t) \right) \\
\frac{\partial M_{\mathrm{P},\mathscr{T}}}{\partial V_{\mathscr{T}}}(t) &= k_{\mathrm{P},\mathscr{T}} R_{\mathrm{A}} (1-F_{\mathrm{L}}) u \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} - \frac{\mu_{\mathscr{T}}(t)}{V_{\mathscr{T}}} \right) \left(\frac{1}{V_{\mathscr{V}}} \frac{\partial \mu_{\mathscr{V}}}{\partial V_{\mathscr{T}}}(t) - \frac{1}{V_{\mathscr{T}}} \frac{\partial \mu_{\mathscr{T}}}{\partial V_{\mathscr{T}}}(t) + \frac{1}{V_{\mathscr{T}}^2} \mu_{\mathscr{T}}(t) \right) \\
\frac{\partial M_{\mathrm{P},\mathscr{T}}}{\partial (\theta_x)_i}(t) &= k_{\mathrm{P},\mathscr{T}} R_{\mathrm{A}} (1-F_{\mathrm{L}}) u \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} - \frac{\mu_{\mathscr{T}}(t)}{V_{\mathscr{T}}} \right) \left(\frac{1}{V_{\mathscr{V}}} \frac{\partial \mu_{\mathscr{V}}}{\partial (\theta_x)_i}(t) - \frac{1}{V_{\mathscr{T}}} \frac{\partial \mu_{\mathscr{T}}}{\partial (\theta_x)_i}(t) \right)
\end{aligned}
$$

for $(\theta_x)_i \in \{V_{\mathscr{L}}, V_{\max}\}$.

F.0.3 $M_{\mathscr{T},\mathrm{P}}(t)$ (Eq. 4.6)

$$
\begin{aligned}
M_{\mathscr{T},\mathrm{P}}(t) &= k_{\mathscr{T},\mathrm{P}} R_{\mathrm{A}} (1-F_{\mathrm{L}}) r \left(\frac{\mu_{\mathscr{T}}(t)}{V_{\mathscr{T}}} - \frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} \right) \\
\frac{\partial M_{\mathscr{T},\mathrm{P}}}{\partial R_{\mathrm{A}}}(t) &= k_{\mathscr{T},\mathrm{P}} (1-F_{\mathrm{L}}) r \left(\frac{\mu_{\mathscr{T}}(t)}{V_{\mathscr{T}}} - \frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} \right) \\
&\quad + k_{\mathscr{T},\mathrm{P}} R_{\mathrm{A}} (1-F_{\mathrm{L}}) u \left(\frac{\mu_{\mathscr{T}}(t)}{V_{\mathscr{T}}} - \frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} \right) \left(\frac{1}{V_{\mathscr{T}}} \frac{\partial \mu_{\mathscr{T}}}{\partial R_{\mathrm{A}}}(t) - \frac{1}{V_{\mathscr{V}}} \frac{\partial \mu_{\mathscr{V}}}{\partial R_{\mathrm{A}}}(t) \right) \\
\frac{\partial M_{\mathscr{T},\mathrm{P}}}{\partial k_{\mathscr{T},\mathrm{P}}}(t) &= R_{\mathrm{A}} (1-F_{\mathrm{L}}) r \left(\frac{\mu_{\mathscr{T}}(t)}{V_{\mathscr{T}}} - \frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} \right) \\
&\quad + k_{\mathscr{T},\mathrm{P}} R_{\mathrm{A}} (1-F_{\mathrm{L}}) u \left(\frac{\mu_{\mathscr{T}}(t)}{V_{\mathscr{T}}} - \frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} \right) \left(\frac{1}{V_{\mathscr{T}}} \frac{\partial \mu_{\mathscr{T}}}{\partial k_{\mathscr{T},\mathrm{P}}}(t) - \frac{1}{V_{\mathscr{V}}} \frac{\partial \mu_{\mathscr{V}}}{\partial k_{\mathscr{T},\mathrm{P}}}(t) \right) \\
\frac{\partial M_{\mathscr{T},\mathrm{P}}}{\partial V_{\mathscr{T}}}(t) &= k_{\mathscr{T},\mathrm{P}} R_{\mathrm{A}} (1-F_{\mathrm{L}}) u \left(\frac{\mu_{\mathscr{T}}(t)}{V_{\mathscr{T}}} - \frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} \right) \left(\frac{1}{V_{\mathscr{T}}} \frac{\partial \mu_{\mathscr{T}}}{\partial V_{\mathscr{T}}}(t) - \frac{1}{V_{\mathscr{T}}^2} \mu_{\mathscr{T}}(t) - \frac{1}{V_{\mathscr{V}}} \frac{\partial \mu_{\mathscr{V}}}{\partial V_{\mathscr{T}}}(t) \right) \\
\frac{\partial M_{\mathscr{T},\mathrm{P}}}{\partial V_{\mathscr{V}}}(t) &= k_{\mathscr{T},\mathrm{P}} R_{\mathrm{A}} (1-F_{\mathrm{L}}) u \left(\frac{\mu_{\mathscr{T}}(t)}{V_{\mathscr{T}}} - \frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} \right) \left(\frac{1}{V_{\mathscr{T}}} \frac{\partial \mu_{\mathscr{T}}}{\partial V_{\mathscr{V}}}(t) - \frac{1}{V_{\mathscr{V}}} \frac{\partial \mu_{\mathscr{V}}}{\partial V_{\mathscr{V}}}(t) + \frac{1}{V_{\mathscr{V}}^2} \mu_{\mathscr{V}}(t) \right) \\
\frac{\partial M_{\mathscr{T},\mathrm{P}}}{\partial (\theta_x)_i}(t) &= k_{\mathscr{T},\mathrm{P}} R_{\mathrm{A}} (1-F_{\mathrm{L}}) u \left(\frac{\mu_{\mathscr{T}}(t)}{V_{\mathscr{T}}} - \frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} \right) \left(\frac{1}{V_{\mathscr{T}}} \frac{\partial \mu_{\mathscr{T}}}{\partial (\theta_x)_i}(t) - \frac{1}{V_{\mathscr{V}}} \frac{\partial \mu_{\mathscr{V}}}{\partial (\theta_x)_i}(t) \right)
\end{aligned}
$$

for $(\theta_x)_i \in \{V_{\mathscr{L}}, V_{\max}\}$.

F.0.4 $M_{\mathrm{HA},\mathscr{L}}(t)$ (Eq. 4.7)

$$
\begin{aligned}
M_{\mathrm{HA},\mathscr{L}}(t) &= k_{\mathrm{HA},\mathscr{L}} R_{\mathrm{A}} F_{\mathrm{L}} (1-F_{\mathrm{PV}}) r \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} - \frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} \right) \\
\frac{\partial M_{\mathrm{HA},\mathscr{L}}}{\partial R_{\mathrm{A}}}(t) &= k_{\mathrm{HA},\mathscr{L}} F_{\mathrm{L}} (1-F_{\mathrm{PV}}) r \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} - \frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} \right) \\
&\quad + k_{\mathrm{HA},\mathscr{L}} R_{\mathrm{A}} F_{\mathrm{L}} (1-F_{\mathrm{PV}}) u \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} - \frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} \right) \left(\frac{1}{V_{\mathscr{V}}} \frac{\partial \mu_{\mathscr{V}}}{\partial R_{\mathrm{A}}}(t) - \frac{1}{V_{\mathscr{L}}} \frac{\partial \mu_{\mathscr{L}}}{\partial R_{\mathrm{A}}}(t) \right) \\
\frac{\partial M_{\mathrm{HA},\mathscr{L}}}{\partial k_{\mathrm{HA},\mathscr{L}}}(t) &= R_{\mathrm{A}} F_{\mathrm{L}} (1-F_{\mathrm{PV}}) r \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} - \frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} \right) \\
&\quad + k_{\mathrm{HA},\mathscr{L}} R_{\mathrm{A}} F_{\mathrm{L}} (1-F_{\mathrm{PV}}) u \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} - \frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} \right) \left(\frac{1}{V_{\mathscr{V}}} \frac{\partial \mu_{\mathscr{V}}}{\partial k_{\mathrm{HA},\mathscr{L}}}(t) - \frac{1}{V_{\mathscr{L}}} \frac{\partial \mu_{\mathscr{L}}}{\partial k_{\mathrm{HA},\mathscr{L}}}(t) \right) \\
\frac{\partial M_{\mathrm{HA},\mathscr{L}}}{\partial V_{\mathscr{V}}}(t) &= k_{\mathrm{HA},\mathscr{L}} R_{\mathrm{A}} F_{\mathrm{L}} (1-F_{\mathrm{PV}}) \times \\
&\quad \times \; u \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} - \frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} \right) \left(\frac{1}{V_{\mathscr{V}}} \frac{\partial \mu_{\mathscr{V}}}{\partial V_{\mathscr{V}}}(t) - \frac{1}{V_{\mathscr{V}}^2} \mu_{\mathscr{V}}(t) - \frac{1}{V_{\mathscr{L}}} \frac{\partial \mu_{\mathscr{L}}}{\partial V_{\mathscr{V}}}(t) \right) \\
\frac{\partial M_{\mathrm{HA},\mathscr{L}}}{\partial V_{\mathscr{L}}}(t) &= k_{\mathrm{HA},\mathscr{L}} R_{\mathrm{A}} F_{\mathrm{L}} (1-F_{\mathrm{PV}}) \times \\
&\quad \times \; u \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} - \frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} \right) \left(\frac{1}{V_{\mathscr{V}}} \frac{\partial \mu_{\mathscr{V}}}{\partial V_{\mathscr{L}}}(t) - \frac{1}{V_{\mathscr{L}}} \frac{\partial \mu_{\mathscr{L}}}{\partial V_{\mathscr{L}}}(t) + \frac{1}{V_{\mathscr{L}}^2} \mu_{\mathscr{L}}(t) \right) \\
\frac{\partial M_{\mathrm{HA},\mathscr{L}}}{\partial (\theta_x)_i}(t) &= k_{\mathrm{HA},\mathscr{L}} R_{\mathrm{A}} F_{\mathrm{L}} (1-F_{\mathrm{PV}}) u \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} - \frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} \right) \left(\frac{1}{V_{\mathscr{V}}} \frac{\partial \mu_{\mathscr{V}}}{\partial (\theta_x)_i}(t) - \frac{1}{V_{\mathscr{L}}} \frac{\partial \mu_{\mathscr{L}}}{\partial (\theta_x)_i}(t) \right)
\end{aligned}
$$

for $(\theta_x)_i \in \{V_{\mathscr{T}}, V_{\max}\}$.

F.0.5 $M_{\mathrm{PV},\mathscr{L}}(t)$ (Eq. 4.8)

$$
\begin{aligned}
M_{\mathrm{PV},\mathscr{L}}(t) &= k_{\mathrm{PV},\mathscr{L}} R_{\mathrm{A}} F_{\mathrm{L}} F_{\mathrm{PV}} r \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} + \frac{M_{\mathrm{Gut}}(t)}{R_{\mathrm{A}} F_{\mathrm{L}} F_{\mathrm{PV}}} - \frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} \right) \\
\frac{\partial M_{\mathrm{PV},\mathscr{L}}}{\partial R_{\mathrm{A}}}(t) &= k_{\mathrm{PV},\mathscr{L}} F_{\mathrm{L}} F_{\mathrm{PV}} r \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} + \frac{M_{\mathrm{Gut}}(t)}{R_{\mathrm{A}} F_{\mathrm{L}} F_{\mathrm{PV}}} - \frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} \right) \\
&\quad + k_{\mathrm{PV},\mathscr{L}} R_{\mathrm{A}} F_{\mathrm{L}} F_{\mathrm{PV}} u \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} + \frac{M_{\mathrm{Gut}}(t)}{R_{\mathrm{A}} F_{\mathrm{L}} F_{\mathrm{PV}}} - \frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} \right) \times \\
&\quad \times \left(\frac{1}{V_{\mathscr{V}}} \frac{\partial \mu_{\mathscr{V}}}{\partial R_{\mathrm{A}}}(t) - \frac{M_{\mathrm{Gut}}(t)}{R_{\mathrm{A}}^2 F_{\mathrm{L}} F_{\mathrm{PV}}} - \frac{1}{V_{\mathscr{L}}} \frac{\partial \mu_{\mathscr{L}}}{\partial R_{\mathrm{A}}}(t) \right) \\
\frac{\partial M_{\mathrm{PV},\mathscr{L}}}{\partial k_{\mathrm{PV},\mathscr{L}}}(t) &= R_{\mathrm{A}} F_{\mathrm{L}} F_{\mathrm{PV}} r \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} + \frac{M_{\mathrm{Gut}}(t)}{R_{\mathrm{A}} F_{\mathrm{L}} F_{\mathrm{PV}}} - \frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} \right) \\
&\quad + k_{\mathrm{PV},\mathscr{L}} R_{\mathrm{A}} F_{\mathrm{L}} F_{\mathrm{PV}} u \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} + \frac{M_{\mathrm{Gut}}(t)}{R_{\mathrm{A}} F_{\mathrm{L}} F_{\mathrm{PV}}} - \frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} \right) \times \\
&\quad \times \left(\frac{1}{V_{\mathscr{V}}} \frac{\partial \mu_{\mathscr{V}}}{\partial k_{\mathrm{PV},\mathscr{L}}}(t) - \frac{1}{V_{\mathscr{L}}} \frac{\partial \mu_{\mathscr{L}}}{\partial k_{\mathrm{PV},\mathscr{L}}}(t) \right) \\
\frac{\partial M_{\mathrm{PV},\mathscr{L}}}{\partial V_{\mathscr{V}}}(t) &= k_{\mathrm{PV},\mathscr{L}} R_{\mathrm{A}} F_{\mathrm{L}} F_{\mathrm{PV}} u \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} + \frac{M_{\mathrm{Gut}}(t)}{R_{\mathrm{A}} F_{\mathrm{L}} F_{\mathrm{PV}}} - \frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} \right) \times \\
&\quad \times \left(\frac{1}{V_{\mathscr{V}}} \frac{\partial \mu_{\mathscr{V}}}{\partial V_{\mathscr{V}}}(t) - \frac{1}{V_{\mathscr{V}}^2} \mu_{\mathscr{V}}(t) - \frac{1}{V_{\mathscr{L}}} \frac{\partial \mu_{\mathscr{L}}}{\partial V_{\mathscr{V}}}(t) \right) \\
\frac{\partial M_{\mathrm{PV},\mathscr{L}}}{\partial V_{\mathscr{L}}}(t) &= k_{\mathrm{PV},\mathscr{L}} R_{\mathrm{A}} F_{\mathrm{L}} F_{\mathrm{PV}} u \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} + \frac{M_{\mathrm{Gut}}(t)}{R_{\mathrm{A}} F_{\mathrm{L}} F_{\mathrm{PV}}} - \frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} \right) \times \\
&\quad \times \left(\frac{1}{V_{\mathscr{V}}} \frac{\partial \mu_{\mathscr{V}}}{\partial V_{\mathscr{L}}}(t) - \frac{1}{V_{\mathscr{L}}} \frac{\partial \mu_{\mathscr{L}}}{\partial V_{\mathscr{L}}}(t) + \frac{1}{V_{\mathscr{L}}^2} \mu_{\mathscr{L}}(t) \right) \\
\frac{\partial M_{\mathrm{PV},\mathscr{L}}}{\partial (\theta_x)_i}(t) &= k_{\mathrm{PV},\mathscr{L}} R_{\mathrm{A}} F_{\mathrm{L}} F_{\mathrm{PV}} u \left(\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} + \frac{M_{\mathrm{Gut}}(t)}{R_{\mathrm{A}} F_{\mathrm{L}} F_{\mathrm{PV}}} - \frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} \right) \times \\
&\quad \times \left(\frac{1}{V_{\mathscr{V}}} \frac{\partial \mu_{\mathscr{V}}}{\partial (\theta_x)_i}(t) - \frac{1}{V_{\mathscr{L}}} \frac{\partial \mu_{\mathscr{L}}}{\partial (\theta_x)_i}(t) \right)
\end{aligned}
$$

for $(\theta_x)_i \in \{V_{\mathscr{T}}, V_{\max}\}$.

F.0.6 $M_{\mathrm{HA}}(t)$ (Eq. 4.9)

$$\begin{aligned}
M_{\mathrm{HA}}(t) &= R_{\mathrm{A}}F_{\mathrm{L}}(1-F_{\mathrm{PV}})\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} \\
\frac{\partial M_{\mathrm{HA}}}{\partial R_{\mathrm{A}}}(t) &= F_{\mathrm{L}}(1-F_{\mathrm{PV}})\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} + R_{\mathrm{A}}F_{\mathrm{L}}(1-F_{\mathrm{PV}})\frac{1}{V_{\mathscr{V}}}\frac{\partial \mu_{\mathscr{V}}}{\partial R_{\mathrm{A}}}(t) \\
\frac{\partial M_{\mathrm{HA}}}{\partial V_{\mathscr{V}}}(t) &= R_{\mathrm{A}}F_{\mathrm{L}}(1-F_{\mathrm{PV}})\left(\frac{1}{V_{\mathscr{V}}}\frac{\partial \mu_{\mathscr{V}}}{\partial V_{\mathscr{V}}}(t) - \frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}^2}\right) \\
\frac{\partial M_{\mathrm{HA}}}{\partial (\theta_x)_i}(t) &= R_{\mathrm{A}}F_{\mathrm{L}}(1-F_{\mathrm{PV}})\frac{1}{V_{\mathscr{V}}}\frac{\partial \mu_{\mathscr{V}}}{\partial (\theta_x)_i}(t)
\end{aligned}$$

for $(\theta_x)_i \in \{V_{\mathscr{T}}, V_{\mathscr{L}}, V_{\max}, k\}$.

F.0.7 $M_{\mathrm{PV}}^{(2)}(t)$ (Eq. 4.10)

$$\begin{aligned}
M_{\mathrm{PV}}^{(2)}(t) &= \frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}}R_{\mathrm{A}}F_{\mathrm{L}}F_{\mathrm{PV}} + M_{\mathrm{Gut}}(t) \\
\frac{\partial M_{\mathrm{PV}}^{(2)}}{\partial R_{\mathrm{A}}}(t) &= F_{\mathrm{L}}F_{\mathrm{PV}}\frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}} + R_{\mathrm{A}}F_{\mathrm{L}}F_{\mathrm{PV}}\frac{1}{V_{\mathscr{V}}}\frac{\partial \mu_{\mathscr{V}}}{\partial R_{\mathrm{A}}}(t) \\
\frac{\partial M_{\mathrm{PV}}^{(2)}}{\partial V_{\mathscr{V}}}(t) &= R_{\mathrm{A}}F_{\mathrm{L}}F_{\mathrm{PV}}\left(\frac{1}{V_{\mathscr{V}}}\frac{\partial \mu_{\mathscr{V}}}{\partial V_{\mathscr{V}}}(t) - \frac{\mu_{\mathscr{V}}(t)}{V_{\mathscr{V}}^2}\right) \\
\frac{\partial M_{\mathrm{PV}}^{(2)}}{\partial (\theta_x)_i}(t) &= R_{\mathrm{A}}F_{\mathrm{L}}F_{\mathrm{PV}}\frac{1}{V_{\mathscr{V}}}\frac{\partial \mu_{\mathscr{V}}}{\partial (\theta_x)_i}(t)
\end{aligned}$$

for $(\theta_x)_i \in \{V_{\mathscr{T}}, V_{\mathscr{L}}, V_{\max}, k\}$.

F.0.8 $M_{\mathscr{L},\mathrm{HV}}(t)$ (Eq. 4.11 and Eqs. 4.12, 4.13, and 4.14)

$M_{\mathscr{L},\mathrm{HV}}(t)$ **(Eq. 4.11)**

$$
\begin{aligned}
M_{\mathscr{L},\mathrm{HV}}(t) &= k_{\mathscr{L},\mathrm{HV}} R_{\mathrm{A}} F_{\mathrm{L}} r \left(\frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} - C_{\mathrm{HV}}(t) \right) \\
\frac{\partial M_{\mathscr{L},\mathrm{HV}}}{\partial R_{\mathrm{A}}}(t) &= k_{\mathscr{L},\mathrm{HV}} F_{\mathrm{L}} r \left(\frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} - C_{\mathrm{HV}}(t) \right) \\
&\quad + k_{\mathscr{L},\mathrm{HV}} R_{\mathrm{A}} F_{\mathrm{L}} u \left(\frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} - C_{\mathrm{HV}}(t) \right) \left(\frac{1}{V_{\mathscr{L}}} \frac{\partial \mu_{\mathscr{L}}}{\partial R_{\mathrm{A}}}(t) - \frac{\partial C_{\mathrm{HV}}}{\partial R_{\mathrm{A}}}(t) \right) \\
\frac{\partial M_{\mathscr{L},\mathrm{HV}}}{\partial k_{\mathscr{L},\mathrm{HV}}}(t) &= R_{\mathrm{A}} F_{\mathrm{L}} r \left(\frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} - C_{\mathrm{HV}}(t) \right) \\
&\quad + k_{\mathscr{L},\mathrm{HV}} R_{\mathrm{A}} F_{\mathrm{L}} u \left(\frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} - C_{\mathrm{HV}}(t) \right) \left(\frac{1}{V_{\mathscr{L}}} \frac{\partial \mu_{\mathscr{L}}}{\partial k_{\mathscr{L},\mathrm{HV}}}(t) - \frac{\partial C_{\mathrm{HV}}}{\partial k_{\mathscr{L},\mathrm{HV}}}(t) \right) \\
\frac{\partial M_{\mathscr{L},\mathrm{HV}}}{\partial V_{\mathscr{L}}}(t) &= k_{\mathscr{L},\mathrm{HV}} R_{\mathrm{A}} F_{\mathrm{L}} u \left(\frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} - C_{\mathrm{HV}}(t) \right) \left(\frac{1}{V_{\mathscr{L}}} \frac{\partial \mu_{\mathscr{L}}}{\partial V_{\mathscr{L}}}(t) - \frac{1}{V_{\mathscr{L}}^2} \mu_{\mathscr{L}}(t) - \frac{\partial C_{\mathrm{HV}}}{\partial V_{\mathscr{L}}}(t) \right) \\
\\
\frac{\partial M_{\mathscr{L},\mathrm{HV}}}{\partial (\theta_x)_i}(t) &= k_{\mathscr{L},\mathrm{HV}} R_{\mathrm{A}} F_{\mathrm{L}} u \left(\frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} - C_{\mathrm{HV}}(t) \right) \left(\frac{1}{V_{\mathscr{L}}} \frac{\partial \mu_{\mathscr{L}}}{\partial (\theta_x)_i}(t) - \frac{\partial C_{\mathrm{HV}}}{\partial (\theta_x)_i}(t) \right)
\end{aligned}
$$

for $(\theta_x)_i \in \{V_{\mathscr{T}}, V_{\mathscr{V}}, V_{\max}\}$.

$C_{\mathrm{HV}}(t)$ **(Eq. 4.14)**

$$
\begin{aligned}
C_{\mathrm{HV}}(t) &= \begin{cases} \frac{1}{1+k_{\mathscr{L},\mathrm{HV}}}\left(\alpha(t)+k_{\mathscr{L},\mathrm{HV}}\gamma(t)\right) & \alpha(t)<\gamma(t) \\ \alpha(t) & \alpha(t)>\gamma(t) \end{cases} \\
&= \alpha(t)+\left[\frac{1}{1+k_{\mathscr{L},\mathrm{HV}}}\left(\alpha(t)+k_{\mathscr{L},\mathrm{HV}}\gamma(t)\right)-\alpha(t)\right]u(\gamma(t)-\alpha(t))
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial C_{\mathrm{HV}}}{\partial k_{\mathscr{L},\mathrm{HV}}}(t) \\
&= \frac{\partial \alpha}{\partial k_{\mathscr{L},\mathrm{HV}}}(t)+\left[-\frac{1}{(1+k_{\mathscr{L},\mathrm{HV}})^2}\left(\alpha(t)+k_{\mathscr{L},\mathrm{HV}}\gamma(t)\right)\right. \\
&\quad \left.+\frac{1}{1+k_{\mathscr{L},\mathrm{HV}}}\left(\frac{\partial \alpha}{\partial k_{\mathscr{L},\mathrm{HV}}}(t)+k_{\mathscr{L},\mathrm{HV}}\frac{\partial \gamma}{\partial k_{\mathscr{L},\mathrm{HV}}}(t)+\gamma(t)\right)-\frac{\partial \alpha}{\partial k_{\mathscr{L},\mathrm{HV}}}(t)\right]u(\gamma(t)-\alpha(t)) \\
&\quad +\left[\frac{1}{1+k_{\mathscr{L},\mathrm{HV}}}\left(\alpha(t)+k_{\mathscr{L},\mathrm{HV}}\gamma(t)\right)-\alpha(t)\right]\delta(\gamma(t)-\alpha(t))\left(\frac{\partial \gamma}{\partial k_{\mathscr{L},\mathrm{HV}}}(t)-\frac{\partial \alpha}{\partial k_{\mathscr{L},\mathrm{HV}}}(t)\right)
\end{aligned}
$$

When $\gamma(t)=\alpha(t)$, the δ function has its singularity and

$$
\frac{1}{1+k_{\mathscr{L},\mathrm{HV}}}\left(\alpha(t)+k_{\mathscr{L},\mathrm{HV}}\gamma(t)\right)-\alpha(t)=0. \tag{F.3}
$$

Since $0\delta(0)=0$, it follows that

$$
\begin{aligned}
&\frac{\partial C_{\mathrm{HV}}}{\partial k_{\mathscr{L},\mathrm{HV}}}(t) \\
&= \frac{\partial \alpha}{\partial k_{\mathscr{L},\mathrm{HV}}}(t)+\left[-\frac{1}{(1+k_{\mathscr{L},\mathrm{HV}})^2}\left(\alpha(t)+k_{\mathscr{L},\mathrm{HV}}\gamma(t)\right)\right. \\
&\quad \left.+\frac{1}{1+k_{\mathscr{L},\mathrm{HV}}}\left(\frac{\partial \alpha}{\partial k_{\mathscr{L},\mathrm{HV}}}(t)+k_{\mathscr{L},\mathrm{HV}}\frac{\partial \gamma}{\partial k_{\mathscr{L},\mathrm{HV}}}(t)+\gamma(t)\right)-\frac{\partial \alpha}{\partial k_{\mathscr{L},\mathrm{HV}}}(t)\right]u(\gamma(t)-\alpha(t))
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial C_{\mathrm{HV}}}{\partial (\theta_x)_i}(t) \\
&= \frac{\partial \alpha}{\partial (\theta_x)_i}(t)+\left[\frac{1}{1+k_{\mathscr{L},\mathrm{HV}}}\left(\frac{\partial \alpha}{\partial (\theta_x)_i}(t)+k_{\mathscr{L},\mathrm{HV}}\frac{\partial \gamma}{\partial (\theta_x)_i}(t)\right)-\frac{\partial \alpha}{\partial (\theta_x)_i}(t)\right]u(\gamma(t)-\alpha(t)) \\
&\quad +\left[\frac{1}{1+k_{\mathscr{L},\mathrm{HV}}}\left(\alpha(t)+k_{\mathscr{L},\mathrm{HV}}\gamma(t)\right)-\alpha(t)\right]\delta(\gamma(t)-\alpha(t))\left(\frac{\partial \gamma}{\partial (\theta_x)_i}(t)-\frac{\partial \alpha}{\partial (\theta_x)_i}(t)\right)
\end{aligned}
$$

for $(\theta_x)_i \in \{V_{\mathscr{T}}, V_{\mathscr{V}}, V_{\mathscr{L}}, R_{\mathrm{A}}, V_{\max}\}$. Again, since $0\delta(0) = 0$, it follows that

$$\begin{aligned}
&\frac{\partial C_{\mathrm{HV}}}{\partial (\theta_x)_i}(t) \\
&\quad = \quad \frac{\partial \alpha}{\partial (\theta_x)_i}(t) + \left[\frac{1}{1 + k_{\mathscr{L},\mathrm{HV}}} \left(\frac{\partial \alpha}{\partial (\theta_x)_i}(t) + k_{\mathscr{L},\mathrm{HV}} \frac{\partial \gamma}{\partial (\theta_x)_i}(t)\right) - \frac{\partial \alpha}{\partial (\theta_x)_i}(t)\right] u(\gamma(t) - \alpha(t))
\end{aligned}$$

for $(\theta_x)_i \in \{V_{\mathscr{T}}, V_{\mathscr{V}}, V_{\mathscr{L}}, R_{\mathrm{A}}, V_{\max}\}$.

$\alpha(t)$ (Eq. 4.12)

$$\begin{aligned}
\alpha(t) &= \frac{1}{R_{\mathrm{A}} F_{\mathrm{L}}} \left(M_{\mathrm{HA}}(t) - M_{\mathrm{HA},\mathscr{L}}(t) + M^{(2)}_{\mathrm{PV}}(t) - M_{\mathrm{PV},\mathscr{L}}(t)\right) \\
\frac{\partial \alpha}{\partial R_{\mathrm{A}}}(t) &= -\frac{1}{R_{\mathrm{A}}^2 F_{\mathrm{L}}} \left(M_{\mathrm{HA}}(t) - M_{\mathrm{HA},\mathscr{L}}(t) + M^{(2)}_{\mathrm{PV}}(t) - M_{\mathrm{PV},\mathscr{L}}(t)\right) \\
&\quad + \frac{1}{R_{\mathrm{A}} F_{\mathrm{L}}} \left(\frac{\partial M_{\mathrm{HA}}}{\partial R_{\mathrm{A}}}(t) - \frac{\partial M_{\mathrm{HA},\mathscr{L}}}{\partial R_{\mathrm{A}}}(t) + \frac{\partial M^{(2)}_{\mathrm{PV}}}{\partial R_{\mathrm{A}}}(t) - \frac{\partial M_{\mathrm{PV},\mathscr{L}}}{\partial R_{\mathrm{A}}}(t)\right) \\
\frac{\partial \alpha}{\partial (\theta_x)_i}(t) &= \frac{1}{R_{\mathrm{A}} F_{\mathrm{L}}} \left(\frac{\partial M_{\mathrm{HA}}}{\partial (\theta_x)_i}(t) - \frac{\partial M_{\mathrm{HA},\mathscr{L}}}{\partial (\theta_x)_i}(t) + \frac{\partial M^{(2)}_{\mathrm{PV}}}{\partial (\theta_x)_i}(t) - \frac{\partial M_{\mathrm{PV},\mathscr{L}}}{\partial (\theta_x)_i}(t)\right)
\end{aligned}$$

for $(\theta_x)_i \in \{V_{\mathscr{T}}, V_{\mathscr{V}}, V_{\mathscr{L}}, V_{\max}, k\}$.

$\gamma(t)$ (Eq. 4.13)

$$\begin{aligned}
\gamma(t) &= \frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}} \\
\frac{\partial \gamma}{\partial V_{\mathscr{L}}}(t) &= -\frac{\mu_{\mathscr{L}}(t)}{V_{\mathscr{L}}^2} + \frac{1}{V_{\mathscr{L}}} \frac{\partial \mu_{\mathscr{L}}}{\partial V_{\mathscr{L}}}(t) \\
\frac{\partial \gamma}{\partial (\theta_x)_i}(t) &= \frac{1}{V_{\mathscr{L}}} \frac{\partial \mu_{\mathscr{L}}}{\partial (\theta_x)_i}(t)
\end{aligned}$$

for $(\theta_x)_i \in \{V_{\mathscr{T}}, V_{\mathscr{V}}, R_{\mathrm{A}}, V_{\max}, k\}$.

BIBLIOGRAPHY

[1] Brian D. O. Anderson and John B. Moore. *Optimal Filtering*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1979.

[2] J.E. Baker. Adaptive selection methods for genetic algorithms. *Proceedings of the 1st international Conference on Genetic Algorithms*, 1985.

[3] Dimitri P. Bertsekas. *Nonlinear Programming*. Athena Scientific, P.O. Box 391, Belmont, MA 02178-9998, 2 edition, 1999.

[4] Jeff A. Bilmes. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. Technical Report TR-97-021, Department of Electrical Engineering and Computer Science, University of California at Berkeley, April 1998.

[5] George Casella and Roger L. Berger. *Statistical Inference*. Duxbury, Wadsworth Group, Pacific Grove, CA, 2nd edition, 2002.

[6] L. Cornelius D. Whitmire and P. Whitmire. Monte carlo simulation of an ethanol pharmacokinetic model. *Alcohol Clin. Exp. Res*, 26(10):1484–1493, 2002.

[7] Piet De Jong and Murray J. Mackinnon. Covariances for smoothed estimates in state space models. *Biometrika*, 75(3):601–602, 1988.

[8] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Royal Stat. Soc. B*, 39:1–38, 1977.

[9] Arnaut Doucet, Simon Godsill, and Andrieu Christophe. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing*, 10:197–208, 2000.

[10] Bradley Efron and David V. Hinkley. Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika*, 65(3):457–487, December 1978.

[11] R.A. Fisher. Two new properties of mathematical likelihood. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Characte*, 144, 1934.

[12] R.A. Fisher. *Statistical Methods for Research Workers*. Edinburgh Oliver and Boyd, 7 edition, 1938.

[13] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison Wesley Publishing Company, 1989.

[14] Simon. Haykin. *Kalman Filtering and Neural Networks*. Wiley-Interscience.

[15] Ran Andr C.M. Schagen F. Heij, Christiaan. *Introduction to Mathemetical Systems Theory*. Birkhuser Basel.

[16] J.H. Holland. Adaptation in natural and artificial system. *MIT Press*, 1975.

[17] R. E. Lindstrom J. L. Rheingold and P. K. Wilkinson. A new blood- ow pharmacokinetic model for ethanol. *J. Pharmacokinet. Biopharm.s*, 9:261–278, 1981.

[18] Nicholas T. Longford. *Studying Human Populations*.

[19] G. P. Danko J. Kramer J. Godinez K. K. Bucholz J. I. Nurnberger Jr. M. A. Schuckit, T. L. Smith and V. Hesselbrock. Prospective evaluation of the four dsm-iv criteria for alcohol abuse in a large population. *American Journal of Psychiatry*, 162(2):350–360, 2005.

[20] V. A. Ramchandani M. H. Plawecki, R. A. DeCarlo and S. OConnor. Improved transformation of morphometric measurements for a priori parameter estimation in a physiologically based pharmacokinetic model of ethanol.

[21] Geoffrey J. McLachlan and Thriyambakam Krishnan. *The EM Algorithm and Extensions*. Wiley-Interscience, 1997.

[22] Jerry M. Mendel. *Lessons in Estimation Theory for Signal Processing, Communications, and Control*. Prentice-Hall, 2 edition, 1995.

[23] H.D. Paykin. A. alcohol dependence and abuse diagnoses: Concurrent validity in a nationally respresentative sample. *Alcohol.: Clinical Exp. Res*, 23(1):144–150, 1999.

[24] Martin H. Plawecki, Jae-Joon Han, Peter C. Doerschuk, Vijay Ramchandani, and Sean J. O'Connor. Physiologically-based pharmacokinetic (PBPK) models for ethanol. *IEEE Trans. Biomed. Eng.*, 55(12):2691–2700, December 2008. PMID: 19126448.

[25] Vijay A. Ramchandani and Sean O'Connor. Studying alcohol elimination using the alcohol clamp method. *Alcohol Research and Health*, 29, 2006.

[26] Richard A. Redner and Homer F. Walker. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, 26(2):195–239, April 1984.

[27] Bangalore N. Roopesh, Madhavi Rangaswamy, Chella Karnarajan, et al. Priming deficiency in male subjects at risk for alcoholism: The N4 during a lexical decision task. *Alcoholism: Clinical & Experimental Research*, 33(12):2027–2036, December 2009.

[28] Anne C. Smith and Emery N. Brown. Estimating a state-space model from point process observations. *Neural Computation*, 15:965–991, 2003.

[29] Donald L. Snyder and Michael I. Miller. *Random Point Processes in Time and Space*. Springer-Verlag, New York, 2nd edition, 1991.

[30] Rainer Spanagel, Dusan Bartsch, Bendikt Brors, et al. An integrated genome research network for studying the genetics of alcohol addiction. *Addiction Biology*, 15(4):369–379, October 2010.