1 Introduction

Z. Chen

1.1 A brief overview of state space analysis

1.1.1 Mathematical background

In probability theory and statistics, a random variable is a variable whose value is subject to variations due to chance. A random variable can take on a set of possible different values. The mathematical function describing the possible values of a random variable and their associated probabilities is known as a probability distribution. Random variables can be discrete, that is, taking any of a specified finite or countable list of values, endowed with a probability mass function (pmf); or continuous, taking any numerical value in an interval or collection of intervals, via a probability density function (pdf); or a mixture of both types. From either the pmf or the pdf, one can characterize the cumulant or moment statistics of the random variables, such as the mean, variance, covariance, skewness and kurtosis.

To represent the evolution of the random variable over time, a random process is further introduced. A stochastic process is a collection of random values, which is the probabilistic counterpart to a deterministic process. Whereas the deterministic process is governed by an ordinary differential equation, there is some indeterminacy in the stochastic process: given the identical initial condition, the evolution of the process may vary due to the presence of noise. In discrete time, a stochastic process involves a sequence of random variables and the time series associated with these random variables. A stochastic process is said strictly stationary if the joint probability distribution does not change when shifted in time; whereas a stochastic process is said wide-sense stationary (WSS) if its first moment and covariance statistics do not vary with respect to time. Any strictly stationary process which has a mean and a covariance is also WSS.

A Markov chain (or Markov process), named after Russian mathematician Andrey Markov (1856–1922), is a random process that undergoes transitions from one state to another on a state space. The Markov chain is memoryless: namely, the next state depends only on the current state and not on the sequence of events that preceded it. This specific kind of memoryless property is called a Markovian property.

1.1.2 Unobserved variables and stochastic dynamical systems

A random variable of a system is either observed (or measured) or unobserved (or latent). In the context of a dynamical system, the unobserved variable is termed as the state variable, and the observation is a form of time series. In general, the stochastic dynamical systems can be written as two equations: state equation and observation equation. For simplicity, let us start with an example of a linear stochastic dynamical system.

1.1.2.1 State equation

Assume that an n-dimensional hidden state process $\mathbf{x}_{t+1} \in \mathbb{R}^n$ follows a first-order Markovian dynamics, as it only depends on the previous state at time t and is corrupted by a state noise process $\mathbf{n}_{t+1}$ (which can be either correlated or uncorrelated between individual components)

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{n}_{t+1}, \tag{1.1}$$

where $\mathbf{A}$ is an $n \times n$ state-transition matrix. The state equation describes the state space evolution of a stochastic dynamical system. Equation (1.1) defines a first-order vector autoregressive (AR) process.

1.1.2.2 Observation equation

In a simple form, the m-dimensional observation $\mathbf{y}_t \in \mathbb{R}^m$ is subject to a linear transformation of the hidden state $\mathbf{x}_t$ and is further corrupted by a measurement noise process $\mathbf{v}_t$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{v}_t. \tag{1.2}$$

A stochastic dynamical system with the form of state and observation equations is also called the state space model (SSM).

Note 1: If the noise processes $\mathbf{n}_t$ and $\mathbf{v}_t$ are both Gaussian with zero mean and respective covariance matrices $\mathbf{Q}$ and $\mathbf{R}$, then $\mathbf{y}_t$ will be also a Gaussian process (GP). The GP is a specific stochastic process whose realizations consist of random variables that follow a Gaussian distribution. Moreover, every finite collection of those random variables has a multivariate Gaussian distribution. From the operator theory and nonparametric Bayesian viewpoint, the GP can be viewed as an infinite-dimensional generalization of the multivariate Gaussian distribution.

Note 2: A linear SSM may have a time-varying driving input or control input $\mathbf{u}_t \in \mathbb{R}^r$. For instance,

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_{t+1} + \mathbf{n}_{t+1}, \tag{1.3}$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{D}\mathbf{u}_t + \mathbf{v}_t. \tag{1.4}$$

In the control setup, this is equivalent to a standard linear quadratic Gaussian (LQG) control system (Bertsekas 2005).

Note 3: In a general setup, the SSM is characterized by two probability distributions

$$\mathbf{x}_{t+1}\big|\mathbf{x}_t \sim P(\mathbf{x}_{t+1}|\mathbf{x}_t), \tag{1.5}$$

$$\mathbf{y}_t\big|\mathbf{x}_t \sim P(\mathbf{y}_t|\mathbf{x}_t). \tag{1.6}$$

where the first equation specifies a state transition probability distribution, and the second equation specifies the conditional probability distribution of the observations given the state variable. These two probability distributions are often associated with a parametric or nonparametric statistical model, characterized by a specific parameter θ. For instance, $P(\mathbf{y}_t|\mathbf{x}_t)$ can be defined by the family of generalized linear model (GLM) (McCullagh & Nelder 1989; Fahrmeir & Tutz 2001).

1.1.3 Observability, controllability and stability

To estimate unobserved state variables of stochastic dynamical systems, it is important to understand the conditions of *observability* and *controllability* in systems theory (Kalman 1960). Specifically, in order to infer the latent variables of the dynamical system under observations, the system must be observable; in order to maneuver the dynamical system under control input, the system must be controllable. In linear discrete-time system, observability and controllability are linked to the rank of certain matrices.

For the above discrete-time linear Gaussian system (equations 1.1 and 1.2), it is observable if and only if the rank of the following $(nm) \times n$ observability matrix is equal to the dimensionality of the state, n:

$$\text{rank}\begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \mathbf{CA}^2 \\ \vdots \\ \mathbf{CA}^{n-1} \end{bmatrix} = n.$$

For the above discrete-time linear Gaussian system with control input (equation 1.4), it is controllable if and only if the rank of the following $(nr) \times n$ controllability matrix is equal to the dimensionality of the state, n:

$$\text{rank}\left[\mathbf{B} \ \vdots \ \mathbf{AB} \ \vdots \ \cdots \ \vdots \ \mathbf{A}^{n-1}\mathbf{B}\right] = n.$$

In addition, to assure the stability of the linear system, the eigenvalues of the transition matrix $\mathbf{A}$ have to be within the unit circle, namely $0 < |\lambda_i| < 1$ $(i = 1, \ldots, n)$.

It should be emphasized that for nonlinear stochastic systems, however, all of these conditions no longer hold.

1.1.4 Bayes' rule

The foundation of Bayesian estimation is given by Bayes' rule, which consists of two rules: *product rule* and *sum rule* (Bernardo & Smith 1994). Bayes' rule provides a way to compute the conditional, joint and marginal probabilities. Specifically, if we let X and Y be two continuous random variables, the conditional probability $p(X|Y)$ is written as

$$p(X|Y) = \frac{p(X,Y)}{p(Y)} = \frac{p(Y|X)p(X)}{\int p(Y|X)p(X)dX}. \tag{1.7}$$

If $X = \{X_i\}$ is discrete, then equation (1.7) is rewritten as

$$p(X_i|Y) = \frac{p(X_i, Y)}{p(Y)} = \frac{p(Y|X_i)p(X_i)}{\sum_j p(Y|X_j)p(X_j)}. \tag{1.8}$$

In Bayesian language, $p(Y|X), p(X)$ and $p(X|Y)$ are referred to as the likelihood, prior and posterior, respectively. The Bayesian machinery consists of three types of basic operations: *normalization, marginalization* and *expectation*, all of which involve integration. Except for very few scenarios (i.e., Gaussianity), most integrations are analytically intractable when dealing with high-dimensional problems.

1.1.5 Recursive Bayesian estimation

Two fundamental goals in time series analysis are estimation (i.e., filtering or smoothing of the present or past) and prediction (of the future). The variables of estimation or prediction interest are often related to the observations that may be corrupted by noise. The SSM provides a generic framework for analyzing time series data, with any form of filtering, smoothing or prediction. The objective of state space analysis is to compute the optimal estimate of the hidden state given the observed data, which can be derived by a recursive form of Bayes' rule.

Without loss of generality, let $\mathbf{x}_t$ denote the state at discrete time t and $\mathbf{y}_{0:t}$ denote the cumulative observations up to time t. The filtered posterior distribution of the state, conditional on the observations $\mathbf{y}_{0:t}$, bears a form of recursive Bayesian estimation

$$\begin{aligned} p(\mathbf{x}_t|\mathbf{y}_{0:t}) &= \frac{p(\mathbf{x}_t)p(\mathbf{y}_{0:t}|\mathbf{x}_t)}{p(\mathbf{y}_{0:t})} = \frac{p(\mathbf{x}_t)p(\mathbf{y}_t, \mathbf{y}_{0:t-1}|\mathbf{x}_t)}{p(\mathbf{y}_t, \mathbf{y}_{0:t-1})} \\ &= \frac{p(\mathbf{x}_t)p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{y}_{0:t-1})p(\mathbf{y}_{0:t-1}|\mathbf{x}_t)}{p(\mathbf{y}_t|\mathbf{y}_{0:t-1})p(\mathbf{y}_{0:t-1})} \\ &= \frac{p(\mathbf{x}_t)p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{y}_{0:t-1})p(\mathbf{x}_t|\mathbf{y}_{0:t-1})p(\mathbf{y}_{0:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{0:t-1})p(\mathbf{y}_{0:t-1})p(\mathbf{x}_t)} \\ &= \frac{p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{y}_{0:t-1})p(\mathbf{x}_t|\mathbf{y}_{0:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{0:t-1})} = \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{0:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{0:t-1})}, \end{aligned} \tag{1.9}$$

where the first four steps are derived from Bayes' rule, and the last equality of equation (1.9) assumes the conditional independence between the observations. The one-step state prediction, also known as the *Chapman–Kolmogorov equation*, is given by

$$p(\mathbf{x}_{t+1}|\mathbf{y}_{0:t}) = \int p(\mathbf{x}_{t+1}|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{0:t})d\mathbf{x}_t, \tag{1.10}$$

where the probability distribution (or density) $p(\mathbf{x}_{t+1}|\mathbf{x}_t)$ describes a state transition equation, and the probability distribution (or density) $p(\mathbf{y}_t|\mathbf{x}_t)$ is the observation equation. Together equations (1.9) and (1.10) provide the fundamental relations to conduct state space analyses. The above formulation of recursive Bayesian estimation holds for both continuous and discrete variables, for either $\mathbf{x}$ or $\mathbf{y}$ or both. When the state variable is discrete and countable (in which we use S_t to replace $\mathbf{x}_t$), the SSM is also referred to as a finite-state HMM, with associated $p(S_t|S_{t-1})$ and $p(\mathbf{y}_t|S_t)$.

When the hidden state consists of both continuous and discrete variables, the SSM is referred to as a switching SSM, with associated $p(\mathbf{x}_t|\mathbf{x}_{t-1}, S_t)$ and $p(\mathbf{y}_t|\mathbf{x}_t, S_t)$ (Barber 2012; Ghahramani 1998). In this case, the inference or prediction involves multiple integrals or summations. For example, the prediction equation (1.10) will be rewritten as

$$p(\mathbf{x}_{t+1}|\mathbf{y}_{0:t}, S_{0:t}) = \int \sum_{S_t} p(\mathbf{x}_{t+1}|\mathbf{x}_t, S_{t+1})p(S_{t+1}|S_t)p(\mathbf{x}_t|\mathbf{y}_{0:t}, S_{0:t})d\mathbf{x}_t. \tag{1.11}$$

Extensions of the SSM, such as higher-order Markovian dependence or factorial dependence of Markov chains (Ghahramani & Jordan 1997; Saul & Jordan 1999), will not be discussed here.

1.1.6 Two illustrated examples

Example 1: Motor neuroprosthetics. A motor neuroprosthetic device, or brain machine interface (BMI), is a machine that takes the signal inputs (such as the spike activity or local field potentials) from certain areas of the brain (such as primary motor or premotor cortex), extracts and transforms the information into overt device control such that it reflects the intension of the user's brain; or a computational machine turning thoughts into action (Hatsopoulos & Donoghue 2009). Some neurons from the primary motor cortex (M1) encode the kinematic information of the movement, such as the direction, position and velocity.

Let $\mathbf{x}_t = [\boldsymbol{M}_t^\top, \boldsymbol{V}_t^\top]^\top$ denote the augmented state vector that consists of three-dimensional (3D) hand position and velocity vectors (where the superscript $^\top$ denotes a vector or matrix transpose operator); let $y_{t,c}$ denote the spike count observation at time t from the c-th neuron, with a tuning function λ_c. A discrete-time SSM for kinematics and spiking activity can be formulated as follows (Brockwell *et al.* 2007)

$$\mathbf{x}_t = \begin{bmatrix} \boldsymbol{M}_t \\ \boldsymbol{V}_t \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{3\times 3} & \Delta t\mathbf{I}_{3\times 3} \\ \mathbf{0} & 0.98\mathbf{I}_{3\times 3} \end{bmatrix} \mathbf{x}_{t-1} + \begin{bmatrix} \mathbf{0} \\ \boldsymbol{e}_t \end{bmatrix}, \tag{1.12}$$

$$y_{t,c} \sim \texttt{Poisson}\Big(\lambda_c\big(\boldsymbol{M}_{\text{t-lag}}, \boldsymbol{V}_{\text{t-lag}}, \sigma_c \epsilon_{t,c}\big)\Big), \tag{1.13}$$

where $\Delta t = 10\,\text{ms}$ denotes the temporal bin size, $\boldsymbol{e}_t$ denotes zero-mean white Gaussian noise with a 3×3 diagonal covariance matrix: diag$\{0.009, 0.009, 0.009\}$, $\epsilon_{t,c}$ is a collection of independent standard Gaussian random variables, and σ_c is a scaling constant. The time lag varies between -250 and $+250$ ms. Depending on the modeling need, the M1 neuronal tuning curve λ_c may have a general function form with respect to the position, velocity, or direction (Georgopoulos *et al.* 1986; Hatsopoulos *et al.* 2007). Figure 1.1 shows some spike rasters of two motor neurons during a 3D reach-to-grasp task (Saleh *et al.* 2012).

Example 2: Sleep-stage scoring. A typical 8-hour night sleep for healthy human adults consists of four or five sleep cycles; each cycle lasts approximately 90 minutes

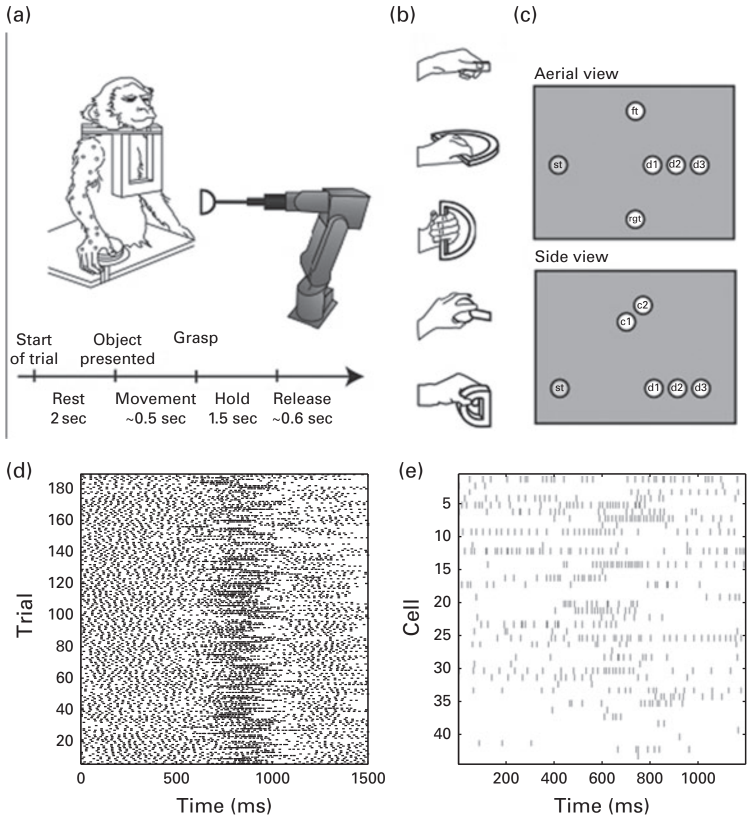

Figure 1.1 (a–c) Motor neuroprosthetics for a 3D reach-to-grasp task for five objects at seven locations. (d) Spike raster of one M1 neuron recorded from multiple trials. (e) Simultaneous recordings of M1 neuronal population spike data from one trial (data courtesy of Professor Nichos Hatsopoulos).

and comprises several different stages: light sleep, deep sleep (slow wave sleep) and rapid eye movement (REM) sleep (Mahowald & Schenck 2005). To diagnose sleep problems, all-night polysomnographic (PSG) recordings including EEG, electrooculogram (EOG) and electromyogram (EMG), are collected from the patient and scored by human experts. However, visual sleep scoring is a time-consuming and highly subjective process. In contrast, computational algorithms can be developed to automatically score the sleep stages (Wake, NREM stages 1 through 3, and REM). Each sleep stage is associated with specific physiological features in EEG, EOG and EMG. In state space analysis, for example, one can design an HMM or hidden semi-Markov model (HSMM) to classify five discrete states that correspond to distinct sleep stages. The state transition can be displayed by a constrained Markov chain (Figure 1.2), which is described by the invariant transition probability $P_{ij} = \Pr(S_t = j|S_{t-1} = i)$ and self-transition probability $P_{ii} = \Pr(S_t = i|S_{t-1} = i)$. The observation probability $P(\mathbf{y}_t|S_t = i)$ describes the observed likelihood conditional on a specific sleep stage.

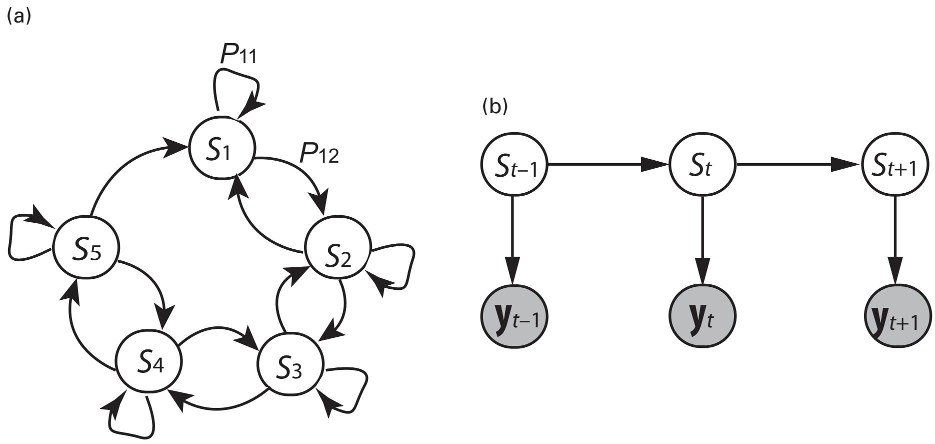

Figure 1.2 (a) A five-state Markov chain that describes the sleep stage transition. S_1: Wake; S_2: NREM stage 1; S_3: NREM stage 2; S_4: NREM stage 3; S_5: REM stage. (b) Graphical illustration of a SSM or HMM. The circles denote random variables, and the observed variable is in gray. The arrows indicate the probabilistic dependency.

1.2 Inference and learning

In statistics, a likelihood is a function of the parameters of a statistical model. The likelihood of a set of parameter values, θ, given observations, is equal to the probability of those observed data conditional on those parameter values, that is $\mathcal{L}(\theta|D) = P(D|\theta)$. In the case of SSM, the complete data consists of observed and latent variables accumulated within a time interval $(0, T]$; and the complete data likelihood is specified by the joint probability distribution of these variables. In the example of the linear Gaussian SSM (equations 1.1 and 1.2), the likelihood function is

$$\begin{aligned} p(X, Y|\theta) = & \frac{1}{(2\pi)^{\frac{n}{2}}|\mathbf{Q}|^{\frac{1}{2}}} \exp\left\{\sum_{t=1}^{T-1} - \frac{(\mathbf{x}_{t+1} - \mathbf{A}\mathbf{x}_t)^\top \mathbf{Q}^{-1}(\mathbf{x}_{t+1} - \mathbf{A}\mathbf{x}_t)}{2}\right\} \\ & + \frac{1}{(2\pi)^{\frac{m}{2}}|\mathbf{R}|^{\frac{1}{2}}} \exp\left\{\sum_{t=1}^{T} - \frac{(\mathbf{y}_t - \mathbf{C}\mathbf{x}_t)^\top \mathbf{R}^{-1}(\mathbf{y}_t - \mathbf{C}\mathbf{x}_t)}{2}\right\}, \end{aligned} \tag{1.14}$$

where the augmented variable $\theta = \{\mathbf{A}, \mathbf{C}, \mathbf{Q}, \mathbf{R}, \mathbf{x}_0\}$ includes the initial state condition and parameters that fully characterize the linear Gaussian SSM.

There are two fundamental approaches to the inference problem: the likelihood approach and the Bayesian approach. The likelihood approach computes a point estimate by maximizing the likelihood function and represents the uncertainty of estimate via confidence intervals (Pawitan 2001). The maximum likelihood estimate (m.l.e.) is asymptotically consistent, normal and efficient, and it is invariant to reparameterization (i.e., functional invariance). By setting $\frac{\partial \mathcal{L}}{\partial \theta} = 0$, one can derive the m.l.e. for the unknown parameter θ, which has the property

$$\theta_{\text{m.l.e.}} \sim \mathcal{N}(\theta, \mathbf{\Sigma}), \tag{1.15}$$

where $\mathcal{N}(\theta, \mathbf{\Sigma})$ denotes a multivariate Gaussian distribution with mean θ and covariance matrix $\mathbf{\Sigma}$; and the covariance is also related to the negative inverse of the Fisher information matrix, i.e., $\mathbf{\Sigma} = -\left[\frac{\partial \mathcal{L}}{\partial\theta\partial\theta^\top}\right]^{-1}$. In many cases there is no closed-form solution to $\frac{\partial \mathcal{L}}{\partial\theta} = 0$, and one has to rely on iterative optimization procedures to obtain either the global or local m.l.e. optimum.

It shall be cautioned that the m.l.e. may suffer from overfitting; namely, there is no constraint imposed on the parameter space for the solution. Depending on the sample size and model complexity, such extra freedom used in model fitting may not necessarily lead to a good predictive performance on unseen data (i.e., poor generalization); therefore, regularization and model selection is required in statistical data analyses (Murphy 2012). In contrast, the Bayesian philosophy lets data speak for themselves and models the unknowns (parameters, latent variables and missing data) and uncertainties (which are not necessarily random) with probabilities or probability densities. The Bayesian approach computes the full posterior of the unknowns based on the rules of probability theory; the Bayesian approach can resolve the overfitting problem in a principled way (Bernardo & Smith 1994; Gelman *et al.* 2004; Barber 2012).

Consider a state and parameter (joint) estimation problem. Bayesian inference aims to infer the joint posterior of the state and the parameter using Bayes' rule,

$$\begin{aligned} p(X,\theta|Y) \approx p(X|Y)p(\theta|Y) &= \frac{p(Y|X,\theta)p(X)p(\theta)}{p(Y)} \\ &= \frac{p(Y|X,\theta)p(X)p(\theta)}{\int\int p(Y|X,\theta)p(X)p(\theta)dXd\theta}, \end{aligned} \tag{1.16}$$

where the first equation assumes a factorial form of the posterior for the state and the parameter (first-stage approximation), and $p(X)$ and $p(\theta)$ denote the prior distributions for the state and parameter, respectively. The denominator of equation (1.16) is a normalizing constant known as the *partition function*. When dealing with a prediction problem for unseen data Y^*, we compute the posterior predictive distribution

$$p(Y^*|Y) = \int\int p(Y^*|Y,\theta,X)p(X|Y)p(\theta|Y)dXd\theta, \tag{1.17}$$

or its expected mean

$$\hat{Y}^* = \mathbb{E}_{p(Y^*|Y)}[Y^*] = \int\int\int Y^* p(Y^*|Y,\theta,X)p(X|Y)p(\theta|Y)dXd\theta dY^*. \tag{1.18}$$

Alternatively, Bayesian inference may optimize an alternate criterion, such as the marginal likelihood (also known as "evidence") $p(Y)$,

$$p(Y) = \int\int p(Y|X,\theta)p(X)p(\theta)dXd\theta. \tag{1.19}$$

Methods for likelihood or Bayesian inference will be covered in many chapters of this book, especially in Chapters 2 and 6.

1.3 Applications in neuroscience and medicine

In the literature, numerous applications of SSM to dynamic analyses of neuroscience and clinical data have been found (Chen *et al.* 2010; Paninski *et al.* 2010), which cover a wide range of neural and clinical data, such as EEG, MEG, ECoG (electrocorticography), fMRI, NIRS (near-infrared spectroscopy), calcium imaging, DTI (diffusion tensor imaging), ECG (electrocardiogram) and other physiological signals. According to the nature of the task, we summarize the representative applications in the following categories:

- *Inverse problems*: Applications include solving EEG or MEG inverse problems (Galka *et al.* 2004; Lamus *et al.* 2012), deconvolving fMRI time series (Penny *et al.* 2005) and deconvolving spike trains from calcium imaging (Vogelstein *et al.* 2009, 2010). As an example, Chapter 3 will present a detailed study of MEG source reconstruction problem.
- *Population neuronal decoding of ensemble spikes*: Applications include decoding the movement kinematics from nonhuman primate M1 neurons in neural prosthetics (Brockwell *et al.* 2004; Srinivasan *et al.* 2006, 2007; Yu *et al.* 2007; Kulkarni & Paninski 2007; Wu *et al.* 2006, 2009), or goal-directed movement control (Srinivasan & Brown 2007; Shanechi *et al.* 2012, 2013), or decoding rat's spatial location from hippocampal ensemble spike trains (Brown *et al.* 1998; Barbieri *et al.* 2004; Ergun *et al.* 2007). In human primate studies, Truccolo and colleagues applied the first point-process state space analysis to decode M1 neuronal spike trains recorded in patients with tetraplegia (Truccolo *et al.* 2008). As examples, Chapters 8 and 9 will present studies on neural decoding of rodent hippocampal and primate M1 neurons, respectively.
- *Analysis of single neuronal plasticity or dynamics*: Applications include tracking the receptive field of rat hippocampal neurons in navigation (Brown *et al.* 2001) and analyzing between-trial monkey hippocampal neuronal dynamics during associative learning experiments (Czanner *et al.* 2008). As an example, Chapter 7 will present a study of inferring latent stepping and ramping models of single neuronal dynamics in decision making.
- *Identification of the state of neuronal ensembles*: Applications include detecting stimulus-driven cortical state during behavior (Jones *et al.* 2007; Kemere *et al.* 2008) or detecting intrinsic cortical up/down states during slow wave sleep (Chen *et al.* 2009). Chapter 6 will discuss methods for estimating the state and model parameters of SSM for spike trains data.
- *Assessment of learning behavior of experimental subjects*: Applications include characterizing dynamic behavioral responses in neuroscience experiments (Smith *et al.* 2004, 2005, 2007; Prerau *et al.* 2009). Chapter 10 will revisit some of those examples.
- *Data smoothing and visualization*: Data smoothing and high-dimensional data visualization has become an increasingly important topic in neuroscience. The SSM has provided a powerful framework to characterize the temporal structure

of time series, such as the smoothness and sparsity (Yu *et al.* 2009; Cunningham & Yu 2014; Ba *et al.* 2014).

- *Classification, prediction and diagnosis of clinical data*: An important direction for eHealthCare is clinical data mining, such as data classification, prediction, prognosis or diagnosis (Kennedy & Turley 2011; Liu *et al.* 2012; van der Heijden *et al.* 2014). As examples, Chapters 11 and 12 will present studies of applying advanced SSMs for analyzing real-life physiological data collected from intensive care unit (ICU).
- *Clinical monitoring*: Applications of SSM for monitoring physiological states of patients in laboratory or ICU is an important research topic in clinical practice. As examples, Chapters 13 through 15 will present studies of clinical applications in cardiovascular modeling and monitoring, modeling and control of medical coma, and physiological signal quality monitoring.

References

Ba, D., Babadi, B., Purdon, P. L. & Brown, E. N. (2014). Harmonic pursuit: robust spectral estimation by iteratively re-weighted least squares. *Proceedings of the National Academy of Sciences USA* **111**(50), 5336–5345.

Barber, D. (2012). *Bayesian Reasoning and Machine Learning*, Cambridge: Cambridge University Press.

Barbieri, R., Frank, L. M., Nguyen, D. P., Quirk, M. C., Solo, V., Wilson, M. A. & Brown, E. N. (2004). Dynamic analyses of information encoding in neural ensembles. *Neural Computation* **16**(2), 277–307.

Bernardo, J. & Smith, A. F. M. (1994). *Bayesian Theory*, New York: Wiley.

Bertsekas, D. (2005). *Dynamic Programming and Optimal Control*, Boston, MA: Athena Scientific.

Brockwell, A. E., Kass, R. E. & Schwartz, A. B. (2007). Statistical signal processing and the motor cortex. *Proceedings of the IEEE* **95**(5), 881–898.

Brockwell, A. E., Rojas, A. L. & Kass, R. E. (2004). Recursive Bayesian decoding of motor cortical signals by particle filtering. *Journal of Neurophysiology* **91**(4), 1899–1907.

Brown, E. N., Frank, L. M., Tang, D., Quirk, M. C. & Wilson, M. A. (1998). A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *Journal of Neuroscience* **18**, 7411–7425.

Brown, E. N., Ngyuen, D. P., Frank, L. M., Wilson, M. A. & Solo, V. (2001). An analysis of neural receptive field plasticity by point process adaptive filtering. *Proceedings of National Academy of Sciences USA* **98**, 12261–12266.

Chen, Z., Barbieri, R. & Brown, E. N. (2010). State-space modeling of neural spike train and behavioral data. In K. Oweiss, ed., *Statistical Signal Processing for Neuroscience and Neurotechnology*, Amsterdom: Elsevier, pp. 175–218.

Chen, Z., Vijayan, S., Barbieri, R., Wilson, M. A. & Brown, E. N. (2009). Discrete- and continuous-time probabilistic models and algorithms for inferring neuronal UP and DOWN states. *Neural Computation* **21**, 1797–1862.