Diffusion Approximation for Bayesian Markov Chains

Michael Duff DUFF@GATSBY.UCL.AC.UK
Gatsby Computational Neuroscience Unit, University College London, WC1N 3AR England

Abstract

Given a Markov chain with uncertain transition probabilities modelled in a Bayesian way, we investigate a technique for analytically approximating the mean transition frequency counts over a finite horizon. Conventional techniques for addressing this problem either require the enumeration of a set of generalized process "hyperstates" whose cardinality grows exponentially with the terminal horizon, or are limited to the two-state case and expressed in terms of hypergeometric series. Our approach makes use of a diffusion approximation technique for modelling the evolution of information state components of the hyperstate process. Interest in this problem stems from a consideration of the policy evaluation step of policy iteration algorithms applied to Markov decision processes with uncertain transition probabilities.

1. Introduction

Given a Markov decision process (MDP) with expressed prior uncertainties in the process transition probabilities, consider the problem of computing a policy that optimizes expected total (finite-horizon) reward. Such a policy would effectively address the "exploration-versus-exploitation tradeoff" faced by an agent that seeks to optimize total reinforcement obtained over the entire duration of its interaction with an uncertain world.

A policy-iteration-based approach for solving this problem would interleave "policy-improvement" steps with "policy-evaluation" steps, which compute or approximate the value of a fixed prospective policy.

If one considers an MDP with a known reward structure, then the expected total reward over some finite horizon under a fixed policy is just the inner product between the expected number of `state→action→state` transitions of each kind with their associated expected immediate rewards. Questions about value in this case reduce to questions about the distribution of transition frequency counts. These considerations lead to the focus of this paper—the estimation of the mean transition-frequency counts associated with a Markov chain (an MDP governed by a fixed policy), with unknown transition probabilities (whose uncertainty distributions are updated in a Bayesian way), over a finite time-horizon.

This is a challenging problem in itself with a long history,[1] and though in this paper we do not explicitly pursue the more ambitious goal of developing full-blown policy-iteration methods for computing *optimal* value functions and policies for uncertain MDP's, one can envision using the approximation techniques described here as a basis for constructing effective static stochastic policies or receding-horizon approaches that repeatedly compute such policies.

Our analysis begins by examining the dynamics of information-state. For each physical Markov chain state, we view parameters describing uncertainty in transition probabilities associated with arcs leading from the state as that state's *information state component.* Information state components change when transitions from their corresponding physical state are observed, and on this time-scale, the dynamics of each information state component forms an embedded Markov chain. The dynamics of each embedded Markov chain is non-stationary but smoothly-varying. We introduce a diffusion model for the embedded dynamics, which yields a simple analytic approximation for describing the flow of information-state density.

[1] This problem was addressed by students of Ronald Howard's at MIT in the 1960's (Silver, 1963; Martin, 1967); however, their methods either entail an enumeration of process "hyperstates," (a set whose cardinality grows exponentially with the planning horizon), or else express the solution analytically in terms of hypergeometric series (for the case of two states only).

Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), Washington DC, 2003.

The analysis of embedded information state component dynamics may be carried out independently for each physical-state component. In order to account for interdependency between components and statistically model their joint behavior, we incorporate flux constraints into our analysis; *i.e.*, we account for the fact that, roughly, the number of transitions into a physical state must be equal to the number of transitions out of that state. The introduction of flux constraints essentially links the local time-scales of each embedded component information-state process together in a way that accounts for the interdependence of joint information-state components and maintains consistency with the global time-scale that governs the true, overall hyperstate process.

In slightly more concrete terms, we model each embedded information-state component density by a Gaussian process. Incorporating this model into the flux constraints leads to a linear system of equations in which the coefficient matrix has random (normal) elements, and the expected value of the solution to this linear system provides an analytic approximation for the mean number of transition counts. Analysis reduces to finding an approximation for the mean-value of the inverse of a matrix with random elements. In this paper we show how this can be done when the matrix possesses the special structure implied by our problem.

2. Diffusion approximation and solution

2.1. Information drift of a Bernoulli process

Consider first a Bernoulli process with unknown parameter θ. If we use a beta distribution to model the uncertainty in θ, then two parameters, (m_1, m_2), suffice (the parameters are directly related to the number of successes and failures observed)—we choose the beta distribution to model uncertainty because it is a *conjugate family* (Degroot, 1970) for samples from a Bernoulli distribution. This means that if our prior uncertainty is described by a beta distribution specified by a particular pair of parameters, then our uncertainty after observing a Bernoulli process sample is also described by a beta distribution, but with parameter values that are incremented in a simple way. Figure 1 shows the state transition diagram for this elementary process. Each information state determines the transition probabilities leading out of it in a well-defined way.

In preparation for what follows, we submit the information state space to a rotation and translation:

$$t \;=\; m_1 + m_2 - 2$$

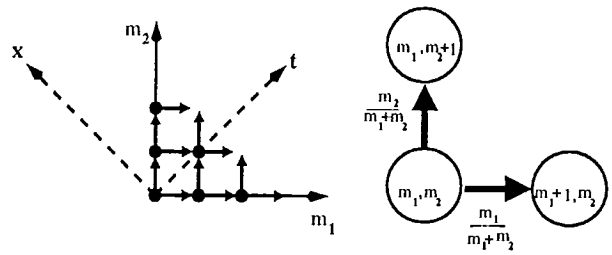

Figure 1. The state transition diagram associated with a Bayes-Bernoulli process. Transition probabilities label arcs in the exploded view. Dashed axes show the affine transformed coordinate system, whose origin coincides with $(m_1, m_2) = (1, 1)$

$$x \;=\; m_2 - m_1.$$

In the new coordinate system, the probability of moving up or down $\Delta x = 1$ in time $\Delta t = 1$ is $\frac{1}{2}(1 + \frac{x}{t+2})$ and $\frac{1}{2}(1 - \frac{x}{t+2})$ respectively. And in analogy to the passage from random walks to Wiener processes, we propose a limiting process in which the probability of moving up or down Δx in time Δt is $\frac{1}{2}(1 + \frac{x}{t+2}\sqrt{\Delta t})$ and $\frac{1}{2}(1 - \frac{x}{t+2}\sqrt{\Delta t})$ respectively. Taking the limit as $\Delta t \to 0$ (with $\Delta x = \sqrt{\Delta t}$) results in a diffusion model with drift $\frac{x}{t+2}$ and diffusion coefficient 1; the density function of $x(t)$ is specified (by Kolmogorov's equations) as soon as these two local quantities have been determined. Alternatively, we may formulate the dynamics of x in terms of the stochastic differential equation (SDE):

$$dx = \frac{x}{t+2}dt + dW_t,$$

where W_t is a Wiener process and the symbolic differential may be rigorously interpreted as an (Ito) integral equation for each sample path.

This is a homogeneous equation with additive noise; it is "linear in the narrow sense" and an application of Ito's formula leads to the closed form solution:

$$x_t = \frac{t+2}{t_0+2}x_{t_0} + (t+2)\int_{t_0}^{t} \frac{1}{s+2}dW_s,$$

which implies that the solution is a Gaussian process. Its mean is $E(x_t) = x_{t_0}\frac{t+2}{t_0+2}$.

The second moment, $E(x_t^2) = P(t)$, may be shown to satisfy $\frac{dP}{dt} = \frac{2}{t+2}P + 1$, which by introducing an integrating factor can be solved: $P(t) = c(t+2)^2 - (t+2)$, where $c = \frac{x_{t_0}^2 + (t_0+2)}{(t_0+2)^2}$.

Though many steps have been omitted—see (Arnold, 1974) and (Kloeden & Platen, 1992) for relevant back-

ground material on second order processes and Ito calculus—the idea we wish to advance is that diffusion models can be developed in this context and that they can yield useful information (namely, estimates for the mean and variance of the information state as a function of time).

2.2. Information drift of a Bayesian Markov chain

The preceeding development has modeled the information state dynamics of an unknown Bernoulli process, but what we are really interested in is a similar model for Markov chains with unknown transition probabilities. Consider Figure 2, which depicts a Markov chain with two physical states and unknown transition probabilities.

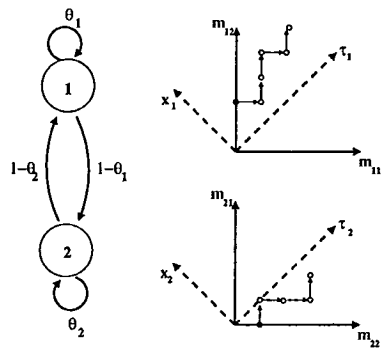

Figure 2. A Markov chain with two physical states and unknown transition probabilities, along with the component information state transition diagrams (the information state trajectory shown corresponds to the following sequence of physical state transitions: $1 \to 1 \to 2 \to 1 \to 2 \to 2 \to 2 \to 1 \to 1 \to 2$).

We may think of each physical state as having associated information state components that describe the uncertainty in the transition probabilities leading from that physical state—these information state components evolve only when the associated physical state experiences a transition, and on a time-scale defined by these particular physical state transitions, the evolution of the corresponding information state components exactly follows that of a Bernoulli process with an unknown parameter. (More generally, for N physical states, sampling is multinomial and the appropriate conjugate family is Dirichlet.) Moreover, the information state component transition diagrams associated with each physical state are identical; all information state components move about upon identical terrain—the only differences between the information state components' dynamics are: (1) they may have differing initial conditions (defined by their priors at starting time), and (2) their "local clocks" may tick at different time-varying rates—the matriculation rate of a given component of information state through its transition diagram is determined by the rate of physical state transitions observed from the information state component's corresponding physical state.

2.3. Flux constraints

In Section 2.1 we proposed a diffusion model for the density function associated with embedded information state components. What we desire is a model for the *joint* density of these components under the constraint that they correspond to physical-state trajectories of a Markov chain. The information-state trajectories must conform to certain constraints of flux—the implied concerted flow into and out of physical states—in order to be consistent with realizable physical state trajectories. Later in this section, we shall argue that these flux constraints imply that reachable information states must lie within regions defined, in part, by certain particular hyperplanes in the joint information state space.

If we are given a set of component information state trajectories, $\left\{x^i(\tau^i), \tau_0^i \leq \tau^i \leq \tau_f^i\right\}_{i=1}^n$ and the initial physical state s_{t_0}, we can, in a straightforward manner construct the corresponding physical state trajectory. We simply follow the information state diffusion $x^{s_{t_0}}$ until it "level-crosses"; *i.e.*, it reaches a boundary line (or hyperplane in higher dimensions) that signifies a change in physical state. For example, in Figure 2, lines defined by constant integral values of m_{12} signify boundaries that, when crossed by the information state process, indicate a physical state transition to `state 2`. We then follow the information state diffusion associated with the new physical state until it level-crosses, and so on.

This procedure constructs sample paths in a left-to-right fashion. An alternative approach starts from an initial set of pre-generated component information state trajectories,[2] which may then be retracted so as to satisfy flux constraints.

2.4. Normal approximation and interpolation

We now consider an approach that may be viewed as using the diffusion model to gauge the distribution of a given information state component if we were to follow

[2]These information state trajectories can be generated independently; *i.e.*, a given trajectory's shape (but not extent or progress as a function of time) is independent of other information state components.

it out to any specified horizon. The slope defined by linear interpolation to the terminal information state is a (normal) random variable, which in turn appears in the coefficient matrix of a linear system of equations describing flux-constraints (and whose solution determines the *joint* values of information state).

Using the special structure of the flux-constraint matrix we can invert the linear flux-constraint system. Finally, we use a Taylor-series estimation technique to approximate the mean values of joint information state, or equivalently, the expected values of the Markov-chain transition frequencies.

In an attempt to make this approach more concrete, we present the steps for the simple case of a Markov chain with two physical states. Let the parameters for the (matrix beta) prior for P be

$$M^o \stackrel{\text{def}}{=} \begin{bmatrix} m_{11}^0 & m_{12}^0 \\ m_{21}^0 & m_{22}^0 \end{bmatrix}.$$

Begin by transforming to (x, τ) coordinates:

$$\begin{bmatrix} \tau_1^0 \\ x_1^0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} m_{11}^0 \\ m_{12}^0 \end{bmatrix} + \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

and

$$\begin{bmatrix} \tau_2^0 \\ x_2^0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} m_{22}^0 \\ m_{21}^0 \end{bmatrix} + \begin{bmatrix} -2 \\ 0 \end{bmatrix}.$$

Introducing a diffusion approximation for each of these components as in Section 2.1 implies that each embedded information state component has a density that is Gaussian with mean

$$E(x_i(\tau_i)) = \frac{x_i(\tau_i^0)}{\tau_i^0 + 2}\tau_i + \frac{2x_i(\tau_i^0)}{\tau_i^0 + 2} \qquad i = 1, 2$$

and second moment

$$E(x_i^2(\tau_i)) = c_i(\tau_i + 2)^2 - (\tau_i + 2) \qquad i = 1, 2$$

where

$$c_i = \frac{x_i^2(\tau_i) + (\tau_i^0 + 2)}{(\tau_i^0 + 2)^2} \qquad i = 1, 2.$$

This implies that the variance of x_i is a quadratic function of τ_i.

Suppressing subscripts for the moment, define the slope, $s(\Delta\tau)$, as the slope of linearly-interpolated trajectory of x defined by its values at τ_0 and $\tau_0 + \Delta\tau$:

$$s(\Delta\tau) \stackrel{\text{def}}{=} \frac{x(\tau_0 + \Delta\tau) - x(\tau_0)}{\Delta\tau}$$

whose expected value is constant:

$$E(s(\Delta\tau)) = \frac{x(\tau_0)}{\tau_0 + 2}.$$

The variance of the slope is

$$Var(s(\Delta\tau)) = \left(\frac{1}{\Delta\tau}\right)^2 Var(x(\tau_0 + \Delta\tau)).$$

Suppose the Markov chain starts in physical state 1. The flux constraints require that observed state transition frequency counts satisfy

$$f_{12} - f_{21} = \delta_{2v},$$

where δ is the Kronecker delta and v is the final state, together with

$$f_{11} + f_{12} + f_{22} + f_{21} = n,$$

where n is the total number of observed transitions.

In matrix form, the flux constraints are

$$\begin{bmatrix} 0 & 1 & 0 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{22} \\ f_{21} \end{bmatrix} = \begin{bmatrix} \delta_{2v} \\ n \end{bmatrix}.$$

(The two right-hand sides correspond to the final state being equal to `state 1` or `state 2`, respectively).

The transition frequency matrix is just the difference of the prior and posterior matrices of matrix beta parameters. To write the flux constraints in terms of (x, τ) parameters we first consider the transformation to matrix beta parameters, and then to transition frequencies; *e.g.*, for components associated with state 1,

$$\begin{bmatrix} m_{11} \\ m_{12} \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \tau_1 \\ x_1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

which implies that

$$\begin{bmatrix} f_{11} \\ f_{12} \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \Delta\tau_1 \\ \Delta x_1 \end{bmatrix}.$$

Thus, in terms of (x, τ) coordinates, the flux constraints may be written as

$$\begin{bmatrix} 0 & 1 & 0 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \frac{1}{2} \begin{bmatrix} 1 & -1 & & \\ 1 & 1 & & \\ & & 1 & -1 \\ & & 1 & 1 \end{bmatrix} \begin{bmatrix} \Delta\tau_1 \\ \Delta x_1 \\ \Delta\tau_2 \\ \Delta x_2 \end{bmatrix} =$$

$$\frac{1}{2}\begin{bmatrix} 1 & 1 & -1 & -1 \\ 2 & 0 & 2 & 0 \end{bmatrix} \begin{bmatrix} \Delta\tau_1 \\ \Delta x_1 \\ \Delta\tau_2 \\ \Delta x_2 \end{bmatrix} = \begin{bmatrix} \delta_{2v} \\ n \end{bmatrix}.$$

Now in time $\Delta\tau_1$, $\Delta x_1 \approx s_1\Delta\tau_1$, and in time $\Delta\tau_2$, $\Delta x_2 \approx s_2\Delta\tau_2$, which make the flux constraints

$$\begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta\tau_1 \\ s_1\Delta\tau_1 \\ \Delta\tau_2 \\ s_2\Delta\tau_2 \end{bmatrix} = \begin{bmatrix} 2\delta_{2v} \\ n \end{bmatrix}$$

or

$$\begin{bmatrix} 1+s_1 & -(1+s_2) \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \Delta\tau_1 \\ \Delta\tau_2 \end{bmatrix} = \begin{bmatrix} 2\delta_{2v} \\ n \end{bmatrix}.$$

The full system (including equations of the form $\Delta x = s\Delta\tau$) is:

$$\begin{bmatrix} 1+s_1 & -(1+s_2) & & \\ 1 & 1 & & \\ s_1 & & -1 & \\ & s_2 & & -1 \end{bmatrix} \begin{bmatrix} \Delta\tau_1 \\ \Delta\tau_2 \\ \Delta x_1 \\ \Delta x_2 \end{bmatrix} = \begin{bmatrix} 2\delta_{2v} \\ n \\ 0 \\ 0 \end{bmatrix}.$$

Call the coefficient matrix A. It has the partitioned form

$$A = \begin{bmatrix} P & 0 \\ D & -I \end{bmatrix}$$

whose inverse is

$$\begin{aligned} A^{-1} &= \begin{bmatrix} P^{-1} & 0 \\ DP^{-1} & -I \end{bmatrix} \\ &= \begin{bmatrix} \frac{\begin{bmatrix} 1 & s_2+1 \\ -1 & s_1+1 \end{bmatrix}}{s_1+s_2+2} & 0 \\ \frac{\begin{bmatrix} s_1 & s_1(s_2+1) \\ -s_2 & s_2(s_1+1) \end{bmatrix}}{s_1+s_2+2} & \begin{bmatrix} -1 & \\ & -1 \end{bmatrix} \end{bmatrix}. \end{aligned}$$

2.5. Taylor-series approximation for mean

We now make use of the fact (see Papoulis, 1991, p. 156, for example) that if a function, $g(x,y)$ is sufficiently smooth near the point (μ_x, μ_y), then the mean μ_g (and variance σ_g^2) of $g(X,Y)$ can be estimated in terms of the mean, variance, and covariance of X and Y:

$$\mu_g \cong g(\mu_x,\mu_y) + \frac{1}{2}\left[\frac{\partial^2 g}{\partial x^2}\sigma_x^2 + 2\frac{\partial^2 g}{\partial x \partial y} r\sigma_x\sigma_y + \frac{\partial^2 g}{\partial y^2}\sigma_y^2\right],$$

where the derivatives of g are evaluated at $x = \mu_x$ and $y = \mu_y$, are r is the correlation coefficient.

In our case, the slopes, s_1 and s_2, are independent, making r zero:

$$\mu_g \cong g(\bar{s}_1,\bar{s}_2) + \frac{1}{2}\left[\frac{\partial^2 g}{\partial s_1^2}\sigma_{s_1}^2 + \frac{\partial^2 g}{\partial s_2^2}\sigma_{s_2}^2\right],$$

Estimates for the mean value of any desired element of A^{-1} can be derived by applying the Taylor-series formula in which the function g assumes the form of the corresponding element's functional form. For example, the second column of A^{-1} is:

$$\begin{bmatrix} \frac{s_2+1}{s_1+s_2+2} \\ \frac{s_1+1}{s_1+s_2+2} \\ \frac{s_1(s_2+1)}{s_1+s_2+2} \\ \frac{s_2(s_1+1)}{s_1+s_2+2} \end{bmatrix}$$

and the Taylor-series estimate for the mean of the third element, for example, in this column is given by

$$\begin{aligned} \mu\left(\tfrac{s_1(s_2+1)}{s_1+s_2+2}\right) \cong\ & \tfrac{\bar{s}_1(\bar{s}_2+1)}{\bar{s}_1+\bar{s}_2+2} \\ & -\tfrac{(\bar{s}_2+1)(\bar{s}_2+2)}{(\bar{s}_1+\bar{s}_2+2)^3}\sigma_{s_1}^2 \\ & -\tfrac{\bar{s}_1(\bar{s}_1+1)}{(\bar{s}_1+\bar{s}_2+2)^3}\sigma_{s_2}^2. \end{aligned}$$

Finally, we may obtain an estimate for joint transition frequencies by substituting values for the slope means, $\bar{s}_i$, and variances, $\sigma_{s_i}^2$, derived previously. Estimates for the mean of the joint distribution over x_i and τ_i can be derived by multiplying the estimate for the mean of A^{-1} by the appropriate right-hand side, yielding an estimate for mean x_i and τ_i. A simple linear transformation translates the result back into the original coordinate system that has an interpretation in terms of transition frequencies.

2.6. Some details

2.6.1. Slope Variance

It can be shown that

$$Var\,(x(\tau)) = \frac{(\tau+2)(\tau-\tau_0)}{\tau_0+2},$$

and hence that

$$\begin{aligned} \sigma_s^2 &= \tfrac{1}{(\tau-\tau_0)^2} Var(x(\tau)) \\ &= \tfrac{(\tau+2)}{(\tau-\tau_0)(\tau_0+2)}, \end{aligned}$$

whose asymptotic value is $\frac{1}{\tau_0+2}$.

For computing estimates for joint transition frequencies, we suggest two possibilities: (1) Use the asymptotic values for slope variances, or (2) Begin by using the asymptotic slope variance values in solving the linear system for mean x and τ, then substitute the mean τ into the formula for σ_s^2 and recompute the solution to the linear system.

2.6.2. Flux constraints

Previously, we stated that the flux constraints require (assuming `state 1` is the starting state) that the transition frequency counts satisfy $f_{12} - f_{21} = \delta_{2v}$, where v

is the final state, together with $f_{11}+f_{12}+f_{22}+f_{21} = n$, where n is the total number of observed transitions.

Adopting the linearly-interpolated model for increments in information state components implies that f_{12} and f_{21} are linear functions of f_{11} and f_{22}, respectively:

$$\begin{aligned} f_{12} &= s_1 f_{11} \\ f_{21} &= s_2 f_{22}, \end{aligned}$$

where the s_i now denote slopes in the transition frequency planes, and we can rewrite the second flux constraint as:

$$\frac{s_1+1}{s_1} f_{12} + \frac{s_2+1}{s_2} f_{21} = n.$$

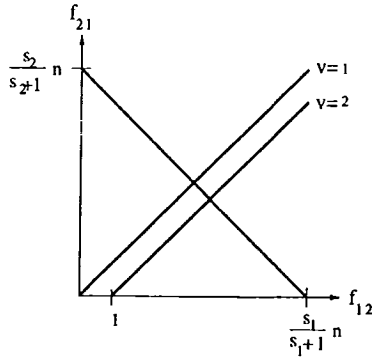

Figure 3. Flux constraint geometry.

Figure 3 sketches the geometric relationships implied by the flux constraints. The first flux constraint defines two hyperplanes, each corresponding to a particular terminal state. The second flux constraint defines a third hyperplane whose intersection with the previously defined hyperplanes determines transition frequency counts that would satisfy the flux constraint equations. In general, the solution to the flux constraint system will be a set of transition frequencies that are real numbers rather than integers. An interpretation of interpolated trajectories in terms of incremental jumps on the information state transition diagram (Figure 1) would give rise to "stairstep" functions. Allowing transition frequencies to take on real values facilitates analysis, but introduces another source of approximation error.

3. Example

3.1. Summary of the analytic procedure

Step 1. Translate from the matrix-beta parameters, which specify prior uncertainty in the unknown transition probabilities, to (x, τ) coordinates:

$$\begin{bmatrix} \tau_1^0 \\ x_1^0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} m_{11}^0 \\ m_{12}^0 \end{bmatrix} + \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

and

$$\begin{bmatrix} \tau_2^0 \\ x_2^0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} m_{22}^0 \\ m_{21}^0 \end{bmatrix} + \begin{bmatrix} -2 \\ 0 \end{bmatrix}.$$

Step 2. An approximation for the expected increments in the values of the information state trajectories, in (x, τ) coordinates, assuming terminal `state 1`, is given by the expected value of the second column of the inverse of the flux constraint matrix, A^{-1}, multiplied by the total number of transitions, n. Utilizing the Taylor-series approximation, the expected values of the elements of the second column of A^{-1} are given by:

$$\begin{aligned}
\mu\left(\tfrac{s_2+1}{s_1+s_2+2}\right) &\cong \tfrac{\bar{s}_2+1}{\bar{s}_1+\bar{s}_2+2} \\
&\quad + \tfrac{\bar{s}_2+1}{(\bar{s}_1+\bar{s}_2+2)^3}\sigma_{s_1}^2 - \tfrac{\bar{s}_1+1}{(\bar{s}_1+\bar{s}_2+2)^3}\sigma_{s_2}^2 \\
\mu\left(\tfrac{s_1+1}{s_1+s_2+2}\right) &\cong \tfrac{\bar{s}_1+1}{\bar{s}_1+\bar{s}_2+2} \\
&\quad - \tfrac{(\bar{s}_2+1)}{(\bar{s}_1+\bar{s}_2+2)^3}\sigma_{s_1}^2 + \tfrac{\bar{s}_1+1}{(\bar{s}_1+\bar{s}_2+2)^3}\sigma_{s_2}^2 \\
\mu\left(\tfrac{s_1(s_2+1)}{s_1+s_2+2}\right) &\cong \tfrac{\bar{s}_1(\bar{s}_2+1)}{\bar{s}_1+\bar{s}_2+2} \\
&\quad - \tfrac{(\bar{s}_2+1)(\bar{s}_2+2)}{(\bar{s}_1+\bar{s}_2+2)^3}\sigma_{s_1}^2 \\
&\quad - \tfrac{\bar{s}_1(\bar{s}_1+1)}{(\bar{s}_1+\bar{s}_2+2)^3}\sigma_{s_2}^2 \\
\mu\left(\tfrac{s_2(s_1+1)}{s_1+s_2+2}\right) &\cong \tfrac{\bar{s}_2(\bar{s}_1+1)}{\bar{s}_1+\bar{s}_2+2} \\
&\quad - \tfrac{\bar{s}_2(\bar{s}_2+1)}{(\bar{s}_1+\bar{s}_2+2)^3}\sigma_{s_1}^2 \\
&\quad - \tfrac{(\bar{s}_1+1)(\bar{s}_1+2)}{(\bar{s}_1+\bar{s}_2+2)^3}\sigma_{s_2}^2,
\end{aligned}$$

where

$$\bar{s}_i = E(s_i) = \frac{x_i(\tau_i^0)}{\tau_i^0 + 2}, \qquad i = 1, 2,$$

and the slope variances, $\sigma_{s_i}^2$, $i = 1, 2$, may be defined by their asymptotic values, $\sigma_{s_i}^2 = \frac{1}{\tau_i^0+2}$, or through an iterative procedure of the sort described previously in Section 2.6.1.

Step 3. Multiply the approximate $\mu(\cdot)$-values from Step 2 by n, the total number of transitions, to get approximate values for the expected increments in (x, τ) coordinates. Perform a simple linear transformation of the result to get an approximation for the expected increments in matrix-beta parameters, which are the expected transition frequencies:

$$E\begin{bmatrix} f_{11} \\ f_{12} \\ f_{22} \\ f_{21} \end{bmatrix} \cong \frac{1}{2}\begin{bmatrix} 1 & -1 & & \\ 1 & 1 & & \\ & & 1 & -1 \\ & & 1 & 1 \end{bmatrix}\begin{bmatrix} \Delta\tau_1 \\ \Delta x_1 \\ \Delta\tau_2 \\ \Delta x_2 \end{bmatrix}.$$

Finally, one may take the inner product of the approximate expected transition frequencies with the corresponding expected immediate rewards to form an approximation for the value of the Bayes-adaptive Markov reward process.

3.2. Example (2 physical states, horizon=25)

As a simple test of our analytic approach we consider a two-state Markov chain whose transition probability uncertainty is characterized by a matrix-beta prior with parameters:

$$M^o \equiv \begin{bmatrix} m^0_{11} & m^0_{12} \\ m^0_{21} & m^0_{22} \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}.$$

Over the course of 25 state transitions (starting from `state 1`) a Monte-Carlo estimate (which may be regarded as "ground truth") for the mean transition frequencies yields:

$$\overline{f}_{MC} = \begin{bmatrix} 3.05 & 4.47 \\ 3.75 & 13.7 \end{bmatrix}$$

while our analytic approximation technique leads to:

$$\overline{f}_{Anal_1} = \begin{bmatrix} 2.67 & 3.48 \\ 3.48 & 15.4 \end{bmatrix}.$$

If we adopt a right-hand-side of the flux constraint equations that is consistent with the terminal state being `state 2`, then two columns of $E(A^{-1})$ enter into our approximation formulae. (The approximate expected increment in (x, τ) coordinates is given by our previous estimate plus $2\times$ first column of $E(A^{-1})$.) This leads to an estimate equal to

$$\overline{f}_{Anal_2} = \begin{bmatrix} 3.19 & 4.23 \\ 3.23 & 14.3 \end{bmatrix}.$$

A reasonable approach would be to average the estimates in some way. If we form the certainty-equivalent estimate for the transition probability matrix from its prior matrix-beta parameters, $\overline{P}_{CE} \equiv \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}$,[3] then solve for the associated stationary distribution, π, via $\pi = \pi P$, $\sum \pi_i = 1$, we may construct a $\pi-$weighted combination of the $\overline{f}_{Anal_i}$ (for our example, $\pi = \begin{bmatrix} \frac{3}{11} & \frac{8}{11} \end{bmatrix}$):

$$\overline{f}_{Anal_\pi} = \begin{bmatrix} 3.05 & 4.02 \\ 3.30 & 14.6 \end{bmatrix}.$$

[3]For reference, we note the certainty-equivalence estimate: $\overline{f}_{CE} = \begin{bmatrix} 2.54 & 5.07 \\ 4.35 & 13.0 \end{bmatrix}$.

The diffusion model for the basic Bayes-Bernoulli process exactly matches the first moment of the true process, but systematically overestimates the second moment (tails of the propagating normal density exceed the natural boundaries of propagation). To derive a more accurate model of the second moment, one may naturally consider Fokker-Planck equations for information state densities with suitable (reflecting) boundary conditions, though the resulting forward equations are not time-homogeneous, which discourages solution via separation-of-variables. It is possible to adjust the predictions suggested by the simple Gaussian process model[4] by condensing the portions of the propagating distribution tails that exceed the natural boundaries onto the boundaries, and then calculating the resulting decrement in the second moment. Considerations of this kind lead to a model with reduced slope-variance estimates and improved estimates for mean transition frequencies:

$$\overline{f}_{ScaledVar} = \begin{bmatrix} 3.13 & 4.39 \\ 3.66 & 13.8 \end{bmatrix}.$$

The computational complexity of our analytic approach is independent of the time horizon; it requires coordinate transformations and evaluations of the A^{-1} column formulae, which entails total work of order $O(N^2)$,

4. Summary

Our goal has been the synthesis of an analytic procedure for estimating the mean state-transition frequencies of a Markov chain whose transition probabilities are unknown but modeled in a Bayesian way. We have noted that information state components (the parameters specifying distributions that model uncertainty in transition probabilities leading from each state) may be viewed as embedded Markov chains that are amenable to diffusion modeling. Analysis yields simple formulae for means and variances of Gaussian processes that model the flow of information state density.

Conceptually we may imagine generating sample paths of joint information state by (see Figure 4): (1) Generating a set of terminal information states consistent with the Gaussian process models, (2) Constructing a corresponding set of linearly-interpolated information-state trajectories, and (3) Retracting these trajectories so that they obey flux constraints, so that together they form a set of trajectories whose joint extent is consistent with a feasible physical-state trajectory. Our

[4]We note that there exist values of $\sigma^2_{s_1}$ and $\sigma^2_{s_2}$ that make the diffusion approximation prediction essentially exact.

analysis gauges the mean result of performing these sample path constructions many times and computing the mean terminal joint information state. Our approach relies upon analytic methods, rather than simulation or brute-force computation, and its complexity scales polynomially with the number of physical states ($O(N^2)$), rather than exponentially with the time horizon. In (Duff, 2002), it is shown how this approach can provide estimates for the value of a stationary, stochastic policy. One next step would be to combine the policy evaluation scheme with a procedure for policy improvement. These ideas are rather preliminary in nature, admittedly suggestive rather than definitive. Some open issues include: (1) We seek a more-refined diffusion model. The variance of the simple Gaussian model overestimates that of true information state components. (2) We seek a more thorough accounting of terminal physical state; our method of combining estimates by weighting them by the certainty-equivalent stationary distribution is provisional, and could almost surely be improved upon.

Acknowledgements

The author wishes to thank Andy Barto and Peter Dayan for helpful comments. This work was supported in part by a grant from the Gatsby Charitable Foundation.

References

Arnold, L. (1974). *Stochastic Differential Equations.* Wiley, New York.

DeGroot, M. H. (1970) *Optimal Statistical Decisions*, McGraw-Hill, New York.

Duff, M. O. (2002) *Optimal learning: Computational procedures for Bayes-adaptive Markov decision processes.* Ph.D. Thesis. Department of Computer Science, University of Massachusetts, Amherst.

Kloeden, P.E. and Platen, E. (1992). *Numerical solution of stochastic differential equations*, Springer, Berlin.

Martin, J. J. (1967). *Bayesian Decision Problems and Markov Chains.* John Wiley & Sons, New York.

Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill, New York.

Silver, E. A. (1963). *Markovian decision processes wih uncertain transition probabilities or rewards,* (Technical Report No. 1). Research in the Control of Complex Systems, Operations Research Center, MIT.

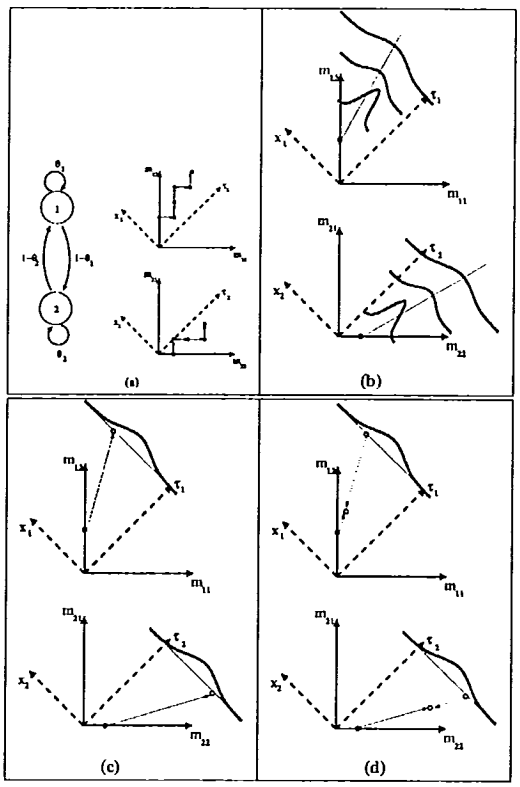

Figure 4. A pictorial summary of our approach: (a) Figure 2 repeated: information components as embedded Markov chains, (b) Information density modeled as a Gaussian process, (c) Sampling from the Gaussian process and construction of a linearly-interpolated trajectories, (d) Retraction to extents that satisfy flux constraints. Our analysis gauges the mean result of performing these sample path constructions many times and computing the mean terminal joint information state.

Using the Triangle Inequality to Accelerate k-Means

Charles Elkan ELKAN@CS.UCSD.EDU
Department of Computer Science and Engineering
University of California, San Diego
La Jolla, California 92093-0114

Abstract

The k-means algorithm is by far the most widely used method for discovering clusters in data. We show how to accelerate it dramatically, while still always computing exactly the same result as the standard algorithm. The accelerated algorithm avoids unnecessary distance calculations by applying the triangle inequality in two different ways, and by keeping track of lower and upper bounds for distances between points and centers. Experiments show that the new algorithm is effective for datasets with up to 1000 dimensions, and becomes more and more effective as the number k of clusters increases. For $k \geq 20$ it is many times faster than the best previously known accelerated k-means method.

1. Introduction

The most common method for finding clusters in data used in applications is the algorithm known as k-means. k-means is considered a fast method because it is not based on computing the distances between all pairs of data points. However, the algorithm is still slow in practice for large datasets. The number of distance computations is nke where n is the number of data points, k is the number of clusters to be found, and e is the number of iterations required. Empirically, e grows sublinearly with k, n, and the dimensionality d of the data.

The main contribution of this paper is an optimized version of the standard k-means method, with which the number of distance computations is in practice closer to n than to nke.

The optimized algorithm is based on the fact that most distance calculations in standard k-means are redundant. If a point is far away from a center, it is not necessary to calculate the exact distance between the point and the center in order to know that the point should not be assigned to this center. Conversely, if a point is much closer to one center than to any other, calculating exact distances is not necessary to know that the point should be assigned to the first center. We show below how to make these intuitions concrete.

We want the accelerated k-means algorithm to be usable wherever the standard algorithm is used. Therefore, we need the accelerated algorithm to satisfy three properties. First, it should be able to start with any initial centers, so that all existing initialization methods can continue to be used. Second, given the same initial centers, it should always produce exactly the same final centers as the standard algorithm. Third, it should be able to use any black-box distance metric, so it should not rely for example on optimizations specific to Euclidean distance.

Our algorithm in fact satisfies a condition stronger than the second one above: after each iteration, it produces the same set of center locations as the standard k-means method. This stronger property means that heuristics for merging or splitting centers (and for dealing with empty clusters) can be used together with the new algorithm. The third condition is important because many applications use a domain-specific distance metric. For example, clustering to identify duplicate alphanumeric records is sometimes based on alphanumeric edit distance (Monge & Elkan, 1996), while clustering of protein structures is often based on an expensive distance function that first rotates and translates structures to superimpose them. Even without a domain-specific metric, recent work shows that using a non-traditional L_p norm with $0 < p < 1$ is beneficial when clustering in a high-dimensional space (Aggarwal et al., 2001).

This paper is organized as follows. Section 2 explains how to use the triangle inequality to avoid redundant distance calculations. Then Section 3 presents the new algorithm, and Section 4 discusses experimental results on six datasets of dimensionality 2 to 1000. Section 5 outlines possible improvements to the method, while Section 6 reviews related work, and Section 7 explains three open research issues.

Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), Washington DC, 2003.