2

Using the time evolution of probabilities

Concept. To extend the formalism that we developed in previous chapters to understand the time evolution of state probabilities. With this knowledge we shall derive popular machine learning methods like Gaussian processes, as well as creating numerical procedures to directly solve for discrete state spaces. In the latter case these can be very intensive computations to perform, so in this chapter we shall also discuss the practical limitations of a fully numerical approach. For the mathematically-inclined, this chapter will take a detailed look at how our formalism can be extended to focus on the time evolution of probabilities. For the programmers, the software described in this chapter lives in this public Git repository: https://github.com/umbralcalc/dennm-torch.

2.1 Probabilistic formalism

In this section we will return to the stochadex formalism that we introduced in the first chapter of this book. As we discussed at that point; this formalism is appropriate for sampling from nearly every stochastic phenomenon that one can think of. We are going to extend this description to consider what happens to the probability that the state history matrix takes a particular set of values over time.

So, how do we begin? In the first chapter, we defined the general stochastic process with the formula $X^i_{\mathtt{t}+1} = F^i_{\mathtt{t}+1}(X_{0:\mathtt{t}}, z, \mathtt{t})$. This equation also has an implicit *master equation* associated to it that fully describes the time evolution of the *probability density function* $P_{\mathtt{t}+1}(X|z)$ of $X_{0:\mathtt{t}+1} = X$ given that the parameters of the process are z. This can be written as

$$P_{\mathtt{t}+1}(X|z) = P_{\mathtt{t}}(X'|z)P_{(\mathtt{t}+1)\mathtt{t}}(x|X', z)\,, \tag{2.1}$$

where for the time being we are assuming the state space is continuous in each of the matrix elements and $P_{(\mathtt{t}+1)\mathtt{t}}(x|X', z)$ is the conditional probability that $X_{\mathtt{t}+1} = x$ given that $X_{0:\mathtt{t}} = X'$ at time $\mathtt{t}$ and the parameters of the process are z.

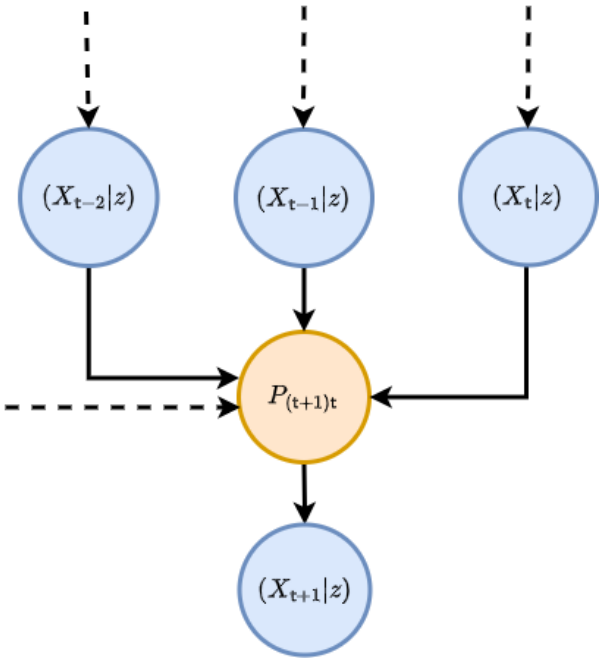

Figure 2.1: Graph representation of Eqs. (2.1) and (2.5).

Consider what happens when we extend the chain of conditional probabilities in Eq. (2.1) back in time by one step. In doing so, we retrieve a joint probability of rows $X_{t+1} = x$ and $X_t = x'$ on the right hand side of the expression

$$P_{t+1}(X|z) = P_{t-1}(X''|z)P_{(t+1)t(t-1)}(x, x'|X'', z)\,. \tag{2.2}$$

Since Eqs. (2.1) and (2.2) are both valid ways to obtain $P_{t+1}(X|z)$ we can average between them without loss of generality in the original expression, like this

$$P_{t+1}(X|z) = \frac{1}{2}\left[P_t(X'|z)P_{(t+1)t}(x|X', z) + P_{t-1}(X''|z)P_{(t+1)t(t-1)}(x, x'|X'', z)\right]. \tag{2.3}$$

Following this line of reasoning to its natural conclusion, Eq. (2.1) can hence be generalised to consider all possible joint distributions of rows at different timesteps like this

$$P_{t+1}(X|z) = \frac{1}{t}\sum_{t''=0}^{t} P_{t''}(X''|z)P_{(t+1)t\ldots t''}(x, x', \ldots|X'', z)\,. \tag{2.4}$$

If we wanted to just look at the distribution over the latest row $X_{t+1} = x$, we could achieve this through marginalisation over all of the previous matrix rows in Eq. (2.1) like this

$$P_{t+1}(x|z) = \int_{\Omega_t} \mathrm{d}X' P_{t+1}(X|z) = \int_{\Omega_t} \mathrm{d}X' P_t(X'|z)P_{(t+1)t}(x|X', z)\,. \tag{2.5}$$

But what is $\Omega_{\mathtt{t}}$? You can think of this as just the domain of possible matrix X' inputs into the integral which will depend on the specific stochastic process we are looking at.

The symbol $\mathrm{d}X'$ in Eq. (2.5) is our shorthand notation throughout the book for computing the sum of integrals over previous state history matrices which can further be reduced via Eq. (2.4) into a product of sub-domain integrals over each matrix row

$$P_{\mathtt{t}+1}(x|z) = \frac{1}{\mathtt{t}}\sum_{\mathtt{t}''=0}^{\mathtt{t}}\left\{\int_{\omega_{\mathtt{t}'}}\mathrm{d}^n x' ... \int_{\Omega_{\mathtt{t}''}}\mathrm{d}X''\right\} P_{\mathtt{t}''}(X''|z)P_{(\mathtt{t}+1)\mathtt{t}...\mathtt{t}''}(x,x',...|X'',z) \tag{2.6}$$

$$= \frac{1}{\mathtt{t}}\sum_{\mathtt{t}''=0}^{\mathtt{t}}\int_{\Omega_{\mathtt{t}''}}\mathrm{d}X'' P_{\mathtt{t}''}(X''|z)P_{(\mathtt{t}+1)\mathtt{t}''}(x|X'',z)\,, \tag{2.7}$$

where each row measure is a Cartesian product of n elements (a Lebesgue measure), i.e.,

$$\mathrm{d}^n x = \prod_{i=0}^{n}\mathrm{d}x^i\,, \tag{2.8}$$

and lowercase $x, x', \ldots$ values will always refer to individual rows within the state matrices. Note that $1/\mathtt{t}$ here is a normalisation factor — this just normalises the sum of all probabilities to 1 given that there is a sum over $\mathtt{t}'$. Note also that, if the process is defined over continuous time, we would need to replace

$$\frac{1}{\mathtt{t}}\sum_{\mathtt{t}'=0}^{\mathtt{t}} \to \frac{1}{t(\mathtt{t})}\sum_{\mathtt{t}'=0}^{\mathtt{t}}\delta t(\mathtt{t}')\,. \tag{2.9}$$

To try and understand what Eqs. (2.1) and (2.5) are saying, we find it's helpful to think of an iterative relationship between probabilities; each of which is connected by their relative conditional probabilities. This kind of thinking is also illustrated in Fig. 2.1.

We can approximate the probability $P_{\mathtt{t}+1}(X|z)$ with a logarithmic expansion like this

$$\ln P_{\mathtt{t}+1}(X|z) \simeq \ln P_{\mathtt{t}+1}(X_*|z) + \frac{1}{2}\sum_{\mathtt{t}'=0}^{\mathtt{t}+1}\sum_{i=0}^{n}\sum_{j=0}^{n}(x-x_*)^i\mathcal{I}^{ij}_{(\mathtt{t}+1)\mathtt{t}'}(x_*,x'_*)(x'-x'_*)^j \tag{2.10}$$

$$\mathcal{I}^{ij}_{(\mathtt{t}+1)\mathtt{t}'}(x_*,x'_*) = \frac{\partial}{\partial x^i}\frac{\partial}{\partial (x')^j}\ln P_{\mathtt{t}+1}(X|z)\Big|_{X=X_*}\,, \tag{2.11}$$

where the values for $X = X_*$ are defined by the vanishing of the first derivative, i.e., these are chosen such that

$$\frac{\partial}{\partial x^i}\ln P_{\mathtt{t}+1}(X|z)\Big|_{X=X_*} = 0\,. \tag{2.12}$$

If we keep the truncation up to second order in Eq. (2.10), note that this expression implies a pairwise correlation structure of the form

$$P_{\mathtt{t}+1}(X|z) \to \prod_{\mathtt{t}'=0}^{\mathtt{t}} P_{(\mathtt{t}+1)\mathtt{t}'}(x,x'|z) = \prod_{\mathtt{t}'=0}^{\mathtt{t}} P_{\mathtt{t}'}(x'|z)P_{(\mathtt{t}+1)\mathtt{t}'}(x|x',z)\,. \tag{2.13}$$

Given this pairwise temporal correlation structure, Eq. (2.7) reduces to this simpler sum of integrals

$$P_{\mathsf{t}+1}(x|z) = \frac{1}{\mathsf{t}}\sum_{\mathsf{t}'=0}^{\mathsf{t}}\int_{\omega_{\mathsf{t}'}} \mathrm{d}^n x' P_{\mathsf{t}'}(x'|z) P_{(\mathsf{t}+1)\mathsf{t}'}(x|x',z)\,. \tag{2.14}$$

In a similar fashion, we can increase the expansion order of Eq. (2.10) to include third-order correlations such that

$$P_{\mathsf{t}+1}(X|z) \to \prod_{\mathsf{t}'=0}^{\mathsf{t}}\prod_{\mathsf{t}''=0}^{\mathsf{t}'-1} P_{\mathsf{t}'\mathsf{t}''}(x',x''|z) P_{(\mathsf{t}+1)\mathsf{t}'\mathsf{t}''}(x|x',x'',z)\,, \tag{2.15}$$

and, in this instance, one can show that Eq. (2.7) reduces to

$$P_{\mathsf{t}+1}(x|z) = \frac{1}{\mathsf{t}}\sum_{\mathsf{t}'=0}^{\mathsf{t}}\frac{1}{\mathsf{t}'-1}\sum_{\mathsf{t}''=0}^{\mathsf{t}'-1}\int_{\omega_{\mathsf{t}'}} \mathrm{d}^n x' \int_{\omega_{\mathsf{t}''}} \mathrm{d}^n x'' P_{\mathsf{t}'\mathsf{t}''}(x',x''|z) P_{(\mathsf{t}+1)\mathsf{t}'\mathsf{t}''}(x|x',x'',z)\,. \tag{2.16}$$

Using $P_{\mathsf{t}'\mathsf{t}''}(x',x''|z) = P_{\mathsf{t}''}(x''|z)P_{\mathsf{t}'\mathsf{t}''}(x'|x'',z)$ one can also show that this integral is a marginalisation of this expression

$$P_{(\mathsf{t}+1)\mathsf{t}''}(x|x'',z) = \frac{1}{\mathsf{t}}\sum_{\mathsf{t}'=0}^{\mathsf{t}}\int_{\omega_{\mathsf{t}'}} \mathrm{d}^n x' P_{\mathsf{t}'\mathsf{t}''}(x'|x'',z) P_{(\mathsf{t}+1)\mathsf{t}'\mathsf{t}''}(x|x',x'',z)\,, \tag{2.17}$$

which describes the time evolution of the conditional probabilities.

Let's imagine that x is just a scalar (as opposed to a row vector) for simplicity in the expressions. We can then discretise the 1D space over x into separate i-labelled regions such that $[P]^i_{\mathsf{t}+1} - [P]^i_{\mathsf{t}} = \mathcal{J}^i_{\mathsf{t}+1}$, where the probability current $\mathcal{J}^i_{\mathsf{t}+1}$ for the factorised processes above would be defined as

$$\mathcal{J}^i_{\mathsf{t}+1} = -[P]^i_{\mathsf{t}} + \frac{1}{\mathsf{t}}\sum_{\mathsf{t}'=0}^{\mathsf{t}}\sum_{i'=0}^{N}\Delta x [P]^{i'}_{\mathsf{t}'}[P]^{ii'}_{(\mathsf{t}+1)\mathsf{t}'} \tag{2.18}$$

$$\mathcal{J}^i_{\mathsf{t}+1} = -[P]^i_{\mathsf{t}} + \frac{1}{\mathsf{t}}\sum_{\mathsf{t}'=1}^{\mathsf{t}}\frac{1}{\mathsf{t}'-1}\sum_{\mathsf{t}''=0}^{\mathsf{t}'-1}\sum_{i'=0}^{N}\sum_{i''=0}^{N}\Delta x^2 [P]^{i''}_{\mathsf{t}''}[P]^{i'i''}_{\mathsf{t}'\mathsf{t}''}[P]^{ii'i''}_{(\mathsf{t}+1)\mathsf{t}'\mathsf{t}''}\,. \tag{2.19}$$

The $[P]^{ii'i''}_{(\mathsf{t}+1)\mathsf{t}'\mathsf{t}''}$ tensor, in particular, will have $N^3\mathsf{t}(\mathsf{t}^2-1)$ elements. Note that the third-order temporal correlations can be evolved by identifying the pairwise conditional probabilities as time-dependent state variables and evolving them according to the following relation

$$[P]^{ii''}_{(\mathsf{t}+1)\mathsf{t}''} = \frac{1}{\mathsf{t}}\sum_{\mathsf{t}'=1}^{\mathsf{t}}\sum_{i'=0}^{N}\Delta x [P]^{i'i''}_{\mathsf{t}'\mathsf{t}''}[P]^{ii'i''}_{(\mathsf{t}+1)\mathsf{t}'\mathsf{t}''}\,. \tag{2.20}$$

What other classes of process can be described by Eqs. (2.1) and (2.5)? For Markovian phenomena, the equations no longer depend on timesteps older than the immediately previous one, hence Eq. (2.5) reduces to just

$$P_{\mathsf{t}+1}(x|z) = \int_{\omega_{\mathsf{t}}} \mathrm{d}^n x'\, P_{\mathsf{t}}(x'|z) P_{(\mathsf{t}+1)\mathsf{t}}(x|x',z)\,. \tag{2.21}$$

By performing a Kramers-Moyal expansion on the $P_{(\mathsf{t}+1)\mathsf{t}}(x|x',z)$ distribution up to second order,[1] we can approximate the right hand side of Eq. (2.21) like this

$$P_{(\mathsf{t}+1)}(x|z) \simeq P_{\mathsf{t}}(x|z) - \sum_{i=0}^{n} \frac{\partial}{\partial x^i}\left[f_{\mathsf{t}}(x,z)P_{\mathsf{t}}(x|z)\right] + \frac{1}{2}\sum_{i=0}^{n}\sum_{j=0}^{n}\frac{\partial}{\partial x^i}\frac{\partial}{\partial x^j}\left[K_{\mathsf{t}}(x,z)P_{\mathsf{t}}(x|z)\right], \quad (2.22)$$

where the components of $f_{\mathsf{t}}(x,z)$ and $K_{\mathsf{t}}(x,z)$ have been defined as

$$f^i_{\mathsf{t}}(x,z) = \int_{\omega_{\mathsf{t}+1}} \mathrm{d}^n x'(x'-x)^i P_{(\mathsf{t}+1)\mathsf{t}}(x'|x,z) \quad (2.23)$$

$$K^{ij}_{\mathsf{t}}(x,z) = \int_{\omega_{\mathsf{t}+1}} \mathrm{d}^n x'(x'-x)^i(x'-x)^j P_{(\mathsf{t}+1)\mathsf{t}}(x'|x,z)\,. \quad (2.24)$$

By inspection of Eq. (2.22) we can define the 'probability current'

$$J_{\mathsf{t}}(x,z) = f_{\mathsf{t}}(x,z)P_{\mathsf{t}}(x|z) - \frac{1}{2}\sum_{j=0}^{n}\frac{\partial}{\partial x^j}\left[K_{\mathsf{t}}(x,z)P_{\mathsf{t}}(x|z)\right]. \quad (2.25)$$

If the probability current vanishes $J_{\mathsf{t}}(x,z) = 0$ individually (this also implies that the distribution is stationary such that $P_{\mathsf{t}+1}(x|z) = P_{\mathsf{t}}(x|z)$), the implicit solution of Eq. (2.22) can be found to be

$$P_{\mathsf{t}}(x|z) \propto K^{-1}_{\mathsf{t}}(x,z)\exp\left[\int \mathrm{d}^n x\, K^{-1}_{\mathsf{t}}(x,z)f_{\mathsf{t}}(x,z)\right]. \quad (2.26)$$

An analog of Eq. (2.5) exists for discrete state spaces as well. We just need to replace the integral with a sum and the schematic would look something like this

$$P_{\mathsf{t}+1}(x|z) = \sum_{\Omega_{\mathsf{t}}} P_{\mathsf{t}}(X'|z)P_{(\mathsf{t}+1)\mathsf{t}}(x|X',z)\,, \quad (2.27)$$

where we note that the P's in the expression above all now refer to *probability mass functions*.

- Add in mean and covariance calculation of the process to generalise it continuous data and dev the Learnadex code using Gotch to support this with different data-linking distributions.
- Add some diagrams for the higher-order correlation expressions.
- Add a software design section and some examples.

[1] The Pawula theorem [1] states that this is, in fact, as far as we can truncate up to unless we include an infinite number of expansion terms.

Bibliography

[1] R. Pawula, *Generalizations and extensions of the fokker-planck-kolmogorov equations*, *IEEE Transactions on Information Theory* **13** (1967) 33–41.