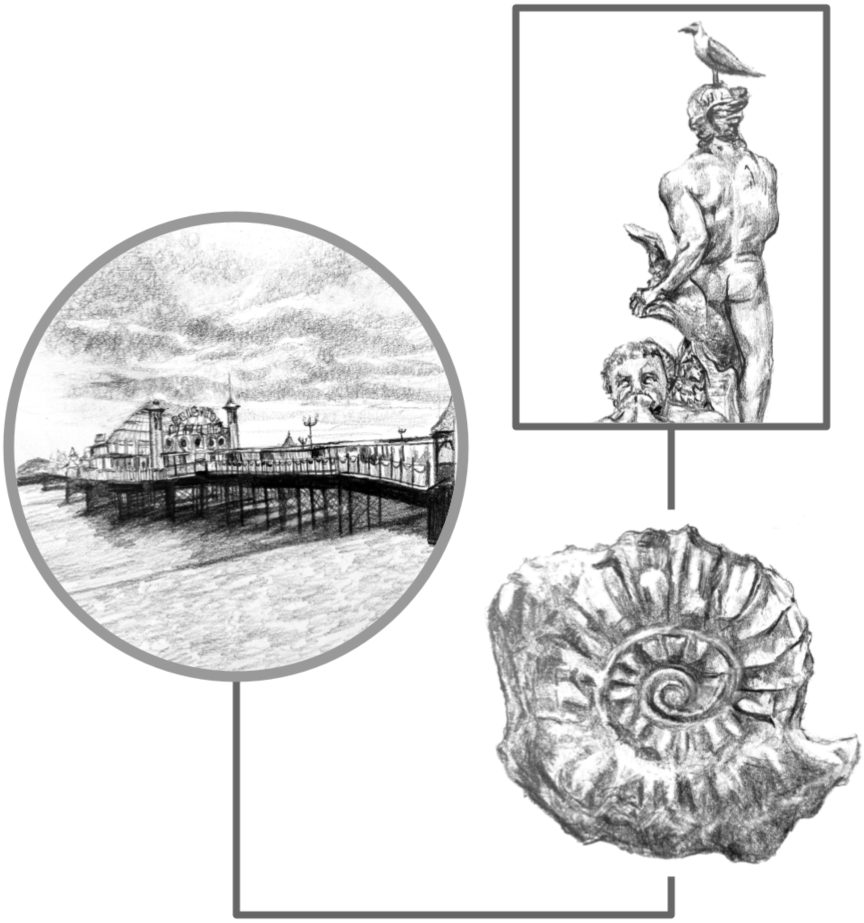

4

Empirical dynamical emulators

Concept. Leveraging the formalism that we developed in the previous chapter to be able to empirically emulate real-world data in a probabilistic way. This technique should enable a researcher to model complex trends in the data dynamically very well; at the cost of making the abstract interpretation of the model less immediately comprehensible in comparison to the statistical inference models of the next chapter. As our generalised framework applies to a wide variety stochastic phenomena, our emulator will be applicable to a great breadth of data modeling problems as well. In this chapter, we will also explore some examples which illustrate how our empirical emulator should be applied in practice and then follow this up with how the code is designed and implemented within the stochadex. For the mathematically-inclined, this chapter will take a detailed look at how our formalism can be extended to focus on probabilistic dynamical process emulation. For the programmers, we will be creating some new modules in the stochadex public Git repository: https://github.com/umbralcalc/stochadex.

4.1 Probabilistic formalism

The key distinction between the methods that we will develop in this chapter and the ones in the next chapter (on statistical inference tools) is in their utility when faced with the problem of attempting to model real-world data. In the proceeding chapter, we shall describe some powerful techniques that can be used most effectively when the researcher is aware of the family of models that generated the data. In the present chapter, we will go into the details of how a more 'empirical' approach can be derived for dynamical process modeling in a probabilistic framework which locally adapts the model to the data through time.

In this chapter, while we think that it's worth going into some mathematical detail to give a better sense of where our formalism comes from; we want to emphasise that the framework we discuss here is not especially new to the technical literature. We shall be using well-known techniques, such as Gaussian processes [1], and our overall framework is comparable to that of Empirical Dynamical Modeling (EDM) [2], or perhaps some classic nonparametric local regression techniques such as LOWESS/Savitzky-Golay filtering [3] as well. The novelties here, instead, lie more in the specifics of how we combine some of these ideas together when referencing the stochadex formalism, and how this manifests in designing more generally-applicable software for the user.

So, how do we begin? The general stochastic process, which we defined with the formula $X^i_{\mathsf{t}+1} = F^i_{\mathsf{t}+1}(X', z, \mathsf{t})$ in the first chapter of this book, also has an implicit *master equation* associated to it that fully describes the time evolution of the *probability density function* $P_{\mathsf{t}+1}(x)$ of the most recent matrix row $x = X_{\mathsf{t}+1}$ at time t. This can be written as

$$P_{\mathsf{t}+1}(x) = \frac{1}{\mathsf{t}}\sum_{\mathsf{t}'=0}^{\mathsf{t}}\int_{\omega_{\mathsf{t}'}} {\rm d}x' P_{\mathsf{t}'}(x')P_{(\mathsf{t}+1)\mathsf{t}'}(x\vert x')\,, \tag{4.1}$$

where at the moment we are assuming the state space is continuous in each dimension and $P_{(\mathsf{t}+1)\mathsf{t}'}(x\vert x')$ is the conditional probability that the matrix row at time $(\mathsf{t}+1)$ will be $x = X_{\mathsf{t}+1}$ given that the row at time t' was $x' = X_{\mathsf{t}'}$. This is a very general equation which should almost always apply to any continuous stochastic phenomenon we want to study in due course. To try and understand what this equation is 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. 4.1.

The factor of $1/\mathsf{t}$ in Eq. (4.1) is a normalisation factor — this just normalises the sum of all probabilities to 1 given that there is a sum over t'. Note that, if the process is defined over continuous time, we would need to replace

$$\frac{1}{\mathsf{t}}\sum_{\mathsf{t}'=0}^{\mathsf{t}} \rightarrow \frac{1}{t(\mathsf{t})}\sum_{\mathsf{t}'=0}^{\mathsf{t}}\delta t(\mathsf{t}')\,. \tag{4.2}$$

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

What if we wanted the joint distribution of both rows $P_{(\mathsf{t}+1)\mathsf{t}'}(x, x')$? One way to obtain this would be to extend Eq. (4.1) such that both matrix rows are marginalised over separately like so

$$\begin{aligned} &P_{(\mathsf{t}+1)\mathsf{t}'}(x, x') = \\ &\quad\frac{1}{(\mathsf{t}'-1)\mathsf{t}}\sum_{\mathsf{t}''=0}^{\mathsf{t}}\sum_{\mathsf{t}'''=0}^{\mathsf{t}'-1}\int_{\omega_{\mathsf{t}''}}{\rm d}x''\int_{\omega_{\mathsf{t}'''}}{\rm d}x''' P_{\mathsf{t}''\mathsf{t}'''}(x'', x''')P_{(\mathsf{t}+1)\mathsf{t}''}(x\vert x'')P_{\mathsf{t}'\mathsf{t}'''}(x'\vert x''')\,. \end{aligned} \tag{4.3}$$

Given Eqs. (4.1) and (4.3) it's also possible to work out what the conditional probabilities would look like using the simple relation

$$P_{(\mathsf{t}+1)\mathsf{t}'}(x\vert x') = \frac{P_{(\mathsf{t}+1)\mathsf{t}'}(x, x')}{P_{\mathsf{t}'}(x')}\,. \tag{4.4}$$

The implicit notation in Eq. (4.1) can hide some staggering complexity. To analyse the system in more detail, we can also do a kind of Kramers-Moyal expansion [4, 5] for each point in time to approximate the overall equation like this

$$\begin{aligned} P_{\mathsf{t}+1}(x) = \frac{1}{\mathsf{t}}\sum_{\mathsf{t}'=0}^{\mathsf{t}}P_{\mathsf{t}'}(x) - \frac{1}{\mathsf{t}}\sum_{\mathsf{t}'=0}^{\mathsf{t}}\sum_{i=1}^{d}\frac{\partial}{\partial x^i}\big[\alpha^i_{(\mathsf{t}+1)\mathsf{t}'}(x)P_{\mathsf{t}'}(x)\big] \\ + \frac{1}{2\mathsf{t}}\sum_{\mathsf{t}'=0}^{\mathsf{t}}\sum_{i=1}^{d}\sum_{j=1}^{d}\frac{\partial}{\partial x^i}\frac{\partial}{\partial x^j}\big[\beta^{ij}_{(\mathsf{t}+1)\mathsf{t}'}(x)P_{\mathsf{t}'}(x)\big] + \dots\,, \end{aligned} \tag{4.5}$$

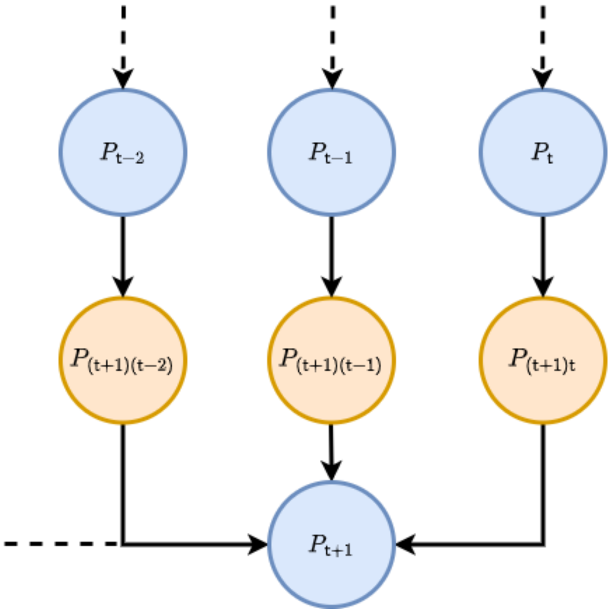
$P_{\mathsf{t}-2}$
$P_{\mathsf{t}-1}$
P_{t}
$P_{(\mathsf{t}+1)(\mathsf{t}-2)}$
$P_{(\mathsf{t}+1)(\mathsf{t}-1)}$
$P_{(\mathsf{t}+1)\mathsf{t}}$
$P_{\mathsf{t}+1}$

Figure 4.1: Graph representation of Eq. (4.1).

in which we have assumed that the state space is d-dimensional. In this expansion, we also needed to define these new integrals

$$\alpha^i_{(\mathsf{t}+1)\mathsf{t}'}(x) = \int_{\omega_{\mathsf{t}'}} \mathrm{d}x'(x'-x)^i P_{(\mathsf{t}+1)\mathsf{t}'}(x'|x) \tag{4.6}$$

$$\beta^{ij}_{(\mathsf{t}+1)\mathsf{t}'}(x) = \int_{\omega_{\mathsf{t}'}} \mathrm{d}x'(x'-x)^i(x'-x)^j P_{(\mathsf{t}+1)\mathsf{t}'}(x'|x)\,. \tag{4.7}$$

So the matrix notation of Eq. (4.1) can indeed hide a very complicated calculation. Truncating the expansion at second-order, Eq. (4.5) tells us that there can be first and second derivatives contributing to the flow of probability to each element of the row $x = X_{\mathsf{t}+1}$ which depend on every element of the matrix X'. The probability does indeed *flow*, in fact. We can define a quantity known as the 'probability current' $J_{(\mathsf{t}+1)\mathsf{t}'}(x)$ from t' to $(\mathsf{t}+1)$ which illustrates this through the following continuity relation

$$P_{\mathsf{t}+1}(x) - \frac{1}{\mathsf{t}}\sum_{\mathsf{t}'=0}^{\mathsf{t}} P_{\mathsf{t}'}(x) = \frac{1}{\mathsf{t}}\sum_{\mathsf{t}'=0}^{\mathsf{t}} \big[P_{\mathsf{t}+1}(x) - P_{\mathsf{t}'}(x)\big] = -\frac{1}{\mathsf{t}}\sum_{\mathsf{t}'=0}^{\mathsf{t}} J_{(\mathsf{t}+1)\mathsf{t}'}(x)\,. \tag{4.8}$$

By inspection of Eq. (4.5) we can therefore also deduce that

$$J^i_{(\mathsf{t}+1)\mathsf{t}'}(x) = \alpha^i_{(\mathsf{t}+1)\mathsf{t}'}(x)P_{\mathsf{t}'}(x) - \frac{1}{2}\sum_{j=1}^{d} \frac{\partial}{\partial x^j}\big[\beta^{ij}_{(\mathsf{t}+1)\mathsf{t}'}(x)P_{\mathsf{t}'}(x)\big] + \dots\,. \tag{4.9}$$

What would happen if we assumed that α and β were just arbitrary time-dependent functions? For example, let's make the following assumptions

$$\alpha^i_{(\mathsf{t}+1)\mathsf{t}'}(x) = \mu^i(\mathsf{t}') - x^i \tag{4.10}$$

$$\beta^{ij}_{(\mathsf{t}+1)\mathsf{t}'}(x) = 2\Sigma^{ij}(\theta, \mathsf{t}')\,, \tag{4.11}$$

where $\mu(\mathsf{t}')$ is an arbitrary vector-valued function of the timestep and $\Sigma(\theta, \mathsf{t}')$ is an arbitrary matrix (often known as the 'diffusion tensor') which depends on both the timestep and a set of hyperparameters θ. If we now also assume stationarity of $P_{\mathsf{t}'}(x) = P_{\mathsf{t}''}(x)$ for any t' and t'' such that

$$P_{\mathsf{t}+1}(x) = \frac{1}{\mathsf{t}}\sum_{\mathsf{t}'=0}^{\mathsf{t}} P_{\mathsf{t}'}(x)\,, \tag{4.12}$$

we can solve Eq. (4.5) to obtain the following stationary solution

$$P_{\mathsf{t}'}(x) = \mathsf{MultivariateNormalPDF}[x; \mu(\mathsf{t}'), \Sigma(\theta, \mathsf{t}')]\,. \tag{4.13}$$

Note in that last step the solution also required the identification that the flow of probability between timesteps vanishes uniquely for each and every t' such that $J_{(\mathsf{t}+1)\mathsf{t}'}(x) = 0$.

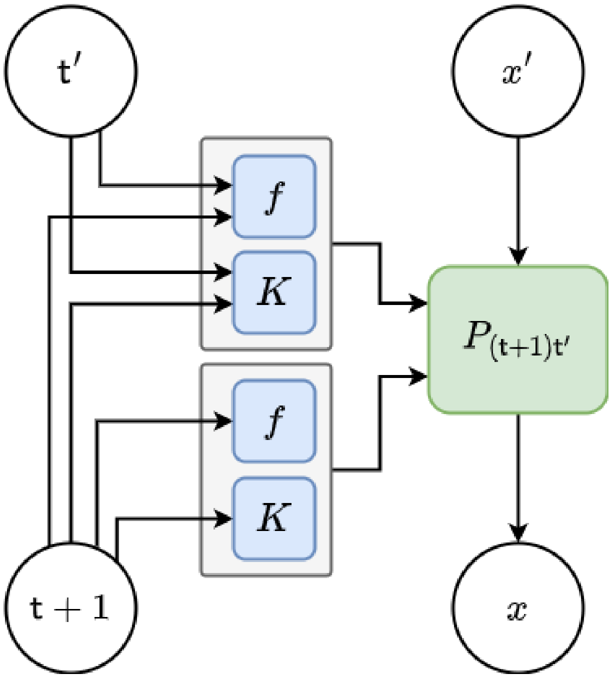

Figure 4.2: Graph representation of the generative process implied by Eq. (4.14).

It's possible to take this derivation a bit further by expanding Eq. (4.3) in a similar fashion, truncating it to second-order, assuming only time-dependent terms and then solving it in the stationary limit. By plugging this solution (and its corresponding marginal distribution equivalent) into Eq. (4.4), it's possible to get something that looks like this conditional distribution

$$P_{(\mathsf{t}+1)\mathsf{t}'}(x|x') \propto \exp\bigg\{ -\frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\big[x-f(\mathsf{t}+1)\big]^{i}[K^{-1}(\theta,\mathsf{t}+1,\mathsf{t}+1)]^{ij}\big[x-f(\mathsf{t}+1)\big]^{j} \\ +\sum_{i=1}^{d}\sum_{j=1}^{d}\big[x-f(\mathsf{t}+1)\big]^{i}[K^{-1}(\theta,\mathsf{t}+1,\mathsf{t}')]^{ij}\big[x'-f(\mathsf{t}')\big]^{j}\bigg\}\,, \tag{4.14}$$

where $K(\theta,\mathsf{t}+1,\mathsf{t}')$ is some arbitrary covariance matrix that encodes how the correlation structure varies with the between compared states at two different timesteps and $K^{-1}(\theta,\mathsf{t}+1,\mathsf{t}')$ denotes taking its inverse. Eq. (4.14) may look a bit familiar to some readers who like using Gaussian processes from the machine learning literature [1] — this version implies a *generative* model for a future x value (which we've illustrated in Fig 4.2), in contrast to the more standard equation used to *infer* values of f. These are two sides of the same coin though.

What other processes can be described by Eq. (4.1)? For Markovian phenomena, the equation no longer depends on timesteps older than the immediately previous one, hence the expression reduces to just

$$P_{\mathsf{t}+1}(x) = \int_{\omega_{\mathsf{t}}} \mathrm{d}x' P_{\mathsf{t}}(x')P_{(\mathsf{t}+1)\mathsf{t}}(x|x')\,. \tag{4.15}$$

It's also easy to show that Eq. (4.5) naturally simplifies into the more usually applied Kramers-Moyal expansion when considering a Markovian process — you just remove the sum over t' and the $1/\mathsf{t}$ normalisation factor.

Note that an analog of Eq. (4.1) 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

$$P_{\mathsf{t}+1}(x) = \frac{1}{\mathsf{t}}\sum_{\mathsf{t}'=0}^{\mathsf{t}}\sum_{\omega_{\mathsf{t}'}} P_{\mathsf{t}'}(x')P_{(\mathsf{t}+1)\mathsf{t}'}(x|x')\,, \tag{4.16}$$

where we note that the P's in the expression above all now refer to *probability mass functions*. Because the state space is now discrete, we cannot immediately intuit an approximative expansion from this expression.

Ok, as a brief mathematical aside; if one is really determined to use a similar approach to the one we derived above, it's quite straightforward to rewrite it in terms of continuous-valued characteristic functions like so

$$\varphi_{\mathsf{t}+1}(s) = \frac{1}{\mathsf{t}}\sum_{\mathsf{t}'=0}^{\mathsf{t}}\int_{\ell_{\mathsf{t}'}} \mathrm{d}s'\mathcal{C}(s')\varphi_{\mathsf{t}'}(s')\varphi_{(\mathsf{t}+1)\mathsf{t}'}(s|s') \tag{4.17}$$

$$\mathcal{C}(s') = \frac{1}{(2\pi)^d}\sum_{\omega_{\mathsf{t}'}} e^{-i(s'\cdot x')}\,, \tag{4.18}$$

where $\ell_{\mathsf{t}'}$ defines all the continuous values that the vector s' can possibly have at time t'. In the expression above, $\mathcal{C}(s')$ acts like is a kind of comb[1] to map the continuous frequency domain of s'

[1] This is very similar to how a 'Dirac comb' works in signal processing [6].

onto the discrete state space of x'. Note also that $\mathcal{C}(s')$ uses the imaginary number i and, to be visually tidier, the dot product notation $a \cdot b$ just means the sum of vector elements: $a \cdot b = \sum_{\forall k} a^k b^k$. In principle, one can perform an approximative expansion on Eq. (4.17) like we did for continuous state spaces. This isn't always the most practical way of analysing the system though.

We have one more important example to discuss and then we can cap off this more mathematical section. In the even-simpler case where x is just a vector of binary 'on' or 'off' states, Eq. (4.16) reduces to

$$P^i_{\mathsf{t}+1} = \frac{1}{\mathsf{t}} \sum_{\mathsf{t}'=0}^{\mathsf{t}} \sum_{j=1}^{d} P^j_{\mathsf{t}'} P^{ij}_{(\mathsf{t}+1)\mathsf{t}'} = \frac{1}{\mathsf{t}} \sum_{\mathsf{t}'=0}^{\mathsf{t}} \sum_{j=1}^{d} \left[P^j_{\mathsf{t}'} A^{ij}_{(\mathsf{t}+1)\mathsf{t}'} + (1 - P^j_{\mathsf{t}'}) B^{ij}_{(\mathsf{t}+1)\mathsf{t}'} \right] , \tag{4.19}$$

where $P^i_{\mathsf{t}'}$ now represents the probability that element $x^i = 1$ (is 'on') at time t'. The matrices A and B are defined as conditional probabilities where the previous state in time $P^j_{\mathsf{t}'}$ was either 'on' or 'off', respectively.

Bibliography

[1] K. P. Murphy, *Machine learning: a probabilistic perspective.* MIT press, 2012.

[2] G. Sugihara and R. M. May, *Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series*, *Nature* **344** (1990) 734–741.

[3] A. Savitzky and M. J. Golay, *Smoothing and differentiation of data by simplified least squares procedures.*, *Analytical chemistry* **36** (1964) 1627–1639.

[4] H. A. Kramers, *Brownian motion in a field of force and the diffusion model of chemical reactions*, *Physica* **7** (1940) 284–304.

[5] J. Moyal, *Stochastic processes and statistical physics*, *Journal of the Royal Statistical Society. Series B (Methodological)* **11** (1949) 150–210.

[6] D. Brandwood, *Fourier transforms in radar and signal processing.* Artech House, 2012.