Worlds Of Observation

Building more realistic environments for machine learning

Robert J. Hardwick and C. M. Gomez-Perales

November 10, 2023

Introduction

In many systems of practical importance to the real world it's common to find that observing the state of the system itself is only partially possible. One needs to only think of the measurement uncertainties in any scientific experiment, the latent demand behind orders in a financial market, the unknown reservoirs of infection for a disease pathogen or even the limits to complete supply chain component observability in recognising just how ubiquitous this situation is. This obscurity can make the learning of algorithms to control these systems using observations alone an extreme — if not sometimes impossible — challenge without further insight provided by a model.

Worlds Of Observation is a book about building more realistic training environments for machine learning algorithms to control these 'noisy' systems in the real world. While model-free reinforcement learning is a popular and very powerful approach to generating such algorithms [1] (especially when there is plenty of data and the system is fully observable), this book is better suited to creating algorithms using a more model-based approach. Those readers who are data scientists, research engineers, statistical programmers or computational scientists may find our mathematically descriptive, yet practically-minded, approach in this book quite interesting and a little different to the usual perspectives. Hopefully, however, the significant overlaps that our approach has with the machine learning literature will still mean that we are in familiar territory for many readers.

As can be expected from a book about algorithms, this text accompanies a lot of new open-source scientific software written in various combinations of Go [2], C++ [3], Python [4] and TypeScript [5]. A major motivation for creating these new tools is to prepare a foundation of code from which to develop new and more complex applications. We also hope that the resulting framework will enable anyone to study new phenomena and explore complex control problems effectively, regardless of their scientific background.

The need to properly test all this software has also provided a wonderful excuse to study and play with an extensive range of models which simulate real-world systems. We've chosen these models based on a fairly broad background of interests, but also to illustrate the cross-disciplinary applicability of our algorithmic framework. To achieve this generalisation, an essential part of this book is the mathematical framework that it introduces and uses throughout.

It seems silly to us that mathematical formalities can obscure the practical computations that a programmer is being asked to implement when reading an equation. So, while we've tried to be as ambitious as possible with the level of technical detail in this book, we've also attempted to write many of the mathematical expressions in a more computer-friendly way where feasible,[1] in contrast with the more conventional formal descriptions. To help with this goal of explainability, we also make use of quite a lot of illustrations and diagrams.

[1] For example, we'll typically be thinking more in terms of 'matrices' and less about 'operators'.

A quick note on the code: any software that we describe in this book (including the software which compiles the book itself [6]) will always be shared under a MIT License [7] in a public Git repository.[2] Forking these repositories and submitting pull requests for new features or applications is strongly encouraged too, though we apologise in advance if we don't follow these up very quickly as all of this work has to be conducted independently in free time, outside of work hours.

The book is split into two main parts: **Part 1**, which details the theoretical background and design of code; and **Part 2**, which explores all of the applications to realistic examples that we wanted to initially try. We hope you, the reader, really enjoy reading through this book and using the all of the code that was built while writing it. We're very grateful to have been able to make use of all the amazing open source software which would otherwise have made this project impossible to achieve.

[2]The repositories will always be somewhere on these lists: https://github.com/umbralcalc?tab=repositories, https://github.com/orgs/WorldsOfObservation/repositories.

Table of contents

Part 1

1

Building a generalised simulator

Concept. To design and build a generalised simulation engine that is able to generate samples from practically any real-world stochastic processes that a researcher could encounter. With such a thing pre-built and self-contained, it can become the basis upon which to build generalised software solutions for a lot of different problems. For the mathematically-inclined, this chapter will require the introduction of a new formalism which we shall refer back to throughout the book. For the programmers, the public Git repository for the code that is described in this chapter can be found here: https://github.com/umbralcalc/stochadex.

1.1 Computational formalism

Before we dive into software design of the stochadex, we need to mathematically define the general computational approach that we're going to take. Because the language of stochastic processes is primarily mathematics, we'd argue this step is essential in enabling a really general description. From experience, it seems reasonable to start by writing down the following formula which describes iterating some arbitrary process forward in time (by one finite step) and adding a new row each to some matrix $X_{0:\mathtt{t}} \to X_{0:\mathtt{t}+1}$

$$X^i_{\mathtt{t}+1} = F^i_{\mathtt{t}+1}(X_{0:\mathtt{t}}, z, \mathtt{t})\,, \tag{1.1}$$

where: i is an index for the dimensions of the 'state' space; $\mathtt{t}$ is the current time index for either a discrete-time process or some discrete approximation to a continuous-time process; $X_{0:\mathtt{t}+1}$ is the next version of $X_{0:\mathtt{t}}$ after one timestep (and hence one new row has been added); z is a vector of arbitrary size which contains the 'hidden' other parameters that are necessary to iterate the process; and $F^i_{\mathtt{t}+1}(X_{0:\mathtt{t}}, z, \mathtt{t})$ as the latest element of an arbitrary matrix-valued function.

Throughout the book, the notation $A_{\mathtt{b}:\mathtt{c}}$ will always refer to a slice of rows from index $\mathtt{b}$ to $\mathtt{c}$ in a matrix (or row vector) A. As we shall discuss shortly, $F^i_{\mathtt{t}+1}(X_{0:\mathtt{t}}, z, \mathtt{t})$ may represent not just operations on deterministic variables, but also on stochastic ones. There is also no requirement for the function to be continuous.

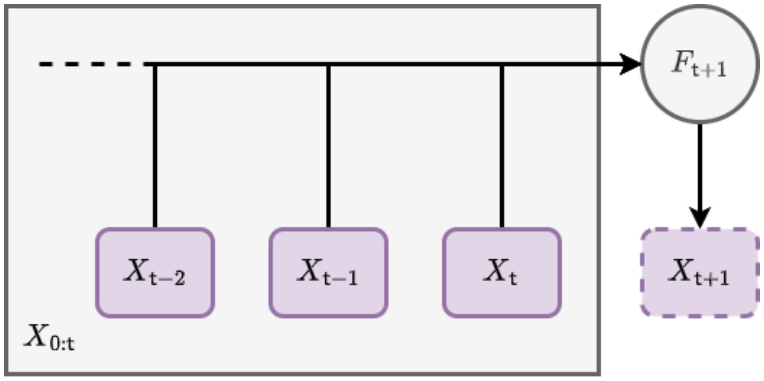

Figure 1.1: Graph representation of Eq. (1.1).

The basic computational idea here is illustrated in Fig. 1.1; we iterate the matrix X forward in time by a row, and use its previous version $X_{0:t}$ as an entire matrix input into a function which populates the elements of its latest rows. In pseudocode you could easily write something with the same idea in it, and it would probably look something like the method diagram in Fig. 1.2.

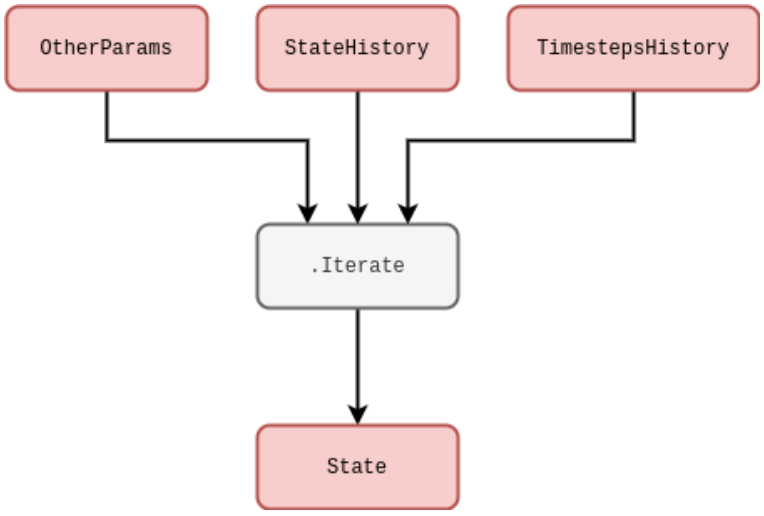

Figure 1.2: Pseudocode representation of Eq. (1.1).

Pretty simple! But why go to all this trouble of storing matrix inputs for previous values of the same process? It's true that this is mostly redundant for *Markovian* phenomena, i.e., processes where their only memory of their history is the most recent value they took. However, for a large class of stochastic processes a full memory[1] of past values is essential to consistently construct the sample paths moving forward. This is true in particular for *non-Markovian* phenomena, where the

[1] Or memory at least within some window.

latest values don't just depend on the immediately previous ones but can depend on values which occured much earlier in the process as well.

For more complex physical models and integrators, the distinct notions of 'numerical timestep' and 'total elapsed continuous time' will crop up quite frequently. Hence, before moving on further details, it will be important to define the total elapsed time variable $t(\mathtt{t})$ for processes which are defined in continuous time. Assuming that we have already defined some function $\delta t(\mathtt{t})$ which returns the specific change in continuous time that corresponds to the step $\mathtt{t}-1 \to \mathtt{t}$, we will always be able to compute the total elapsed time through the relation

$$t(\mathtt{t}) = \sum_{\mathtt{t}'=0}^{\mathtt{t}} \delta t(\mathtt{t}') \,. \tag{1.2}$$

This seems a lot of effort, no? Well it's important to remember that our steps in continuous time may not be constant, so by defining the $\delta t(\mathtt{t})$ function and summing over it we can enable this flexibility in the computation. In case the summation notation is no fun for programmers; we're simply adding up all of the differences in time to get a total. We've illustrated this in Fig. 1.3 for more clarity.

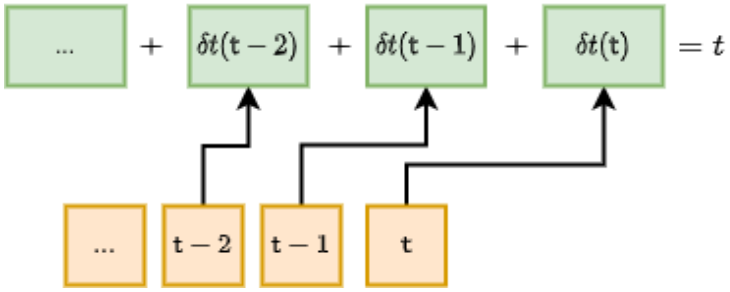

Figure 1.3: Illustration of Eq. (1.2).

So, now that we've mathematically defined a really general notion of iterating the stochastic process forward in time, it makes sense to discuss some simple examples. For instance, it is frequently possible to split F up into deteministic (denoted D) and stochastic (denoted S) matrix-valued functions like so

$$F^i_{\mathtt{t}+1}(X_{0:\mathtt{t}}, z, \mathtt{t}) = D^i_{\mathtt{t}+1}(X_{0:\mathtt{t}}, z, \mathtt{t}) + S^i_{\mathtt{t}+1}(X_{0:\mathtt{t}}, z, \mathtt{t}) \,. \tag{1.3}$$

In the case of stochastic processes with continuous sample paths, it's also nearly always the case with mathematical models of real-world systems that the deterministic part will at least contain the term $D^i_{\mathtt{t}+1}(X_{0:\mathtt{t}}, z, \mathtt{t}) = X^i_{\mathtt{t}}$ because the overall system is described by some stochastic differential equation. This is not a really requirement in our general formalism, however.

What about the stochastic term? For example, if we wanted to consider a *Wiener process noise*, we can define $W^i_{\mathtt{t}}$ is a sample from a Wiener process for each of the state dimensions indexed by i and our formalism becomes

$$S^i_{\mathtt{t}+1}(X_{0:\mathtt{t}}, z, \mathtt{t}) = W^i_{\mathtt{t}+1} - W^i_{\mathtt{t}} \,. \tag{1.4}$$

One draws the increments $W^i_{\mathsf{t}+1} - W^i_{\mathsf{t}}$ from a normal distribution with a mean of 0 and a variance equal to the length of continuous time that the step corresponded to $\delta t(\mathsf{t}+1)$, i.e., the probability density $P_{\mathsf{t}+1}(x^i)$ of the increments $x^i = W^i_{\mathsf{t}+1} - W^i_{\mathsf{t}}$ is

$$P_{\mathsf{t}+1}(x^i) = \mathsf{NormalPDF}[x^i; 0, \delta t(\mathsf{t}+1)]\,. \tag{1.5}$$

Note that for state spaces with dimensions > 1, we could also allow for non-trivial cross-correlations between the noises in each dimension. In pseudocode, the Wiener process is schematically represented by Fig. 1.4.

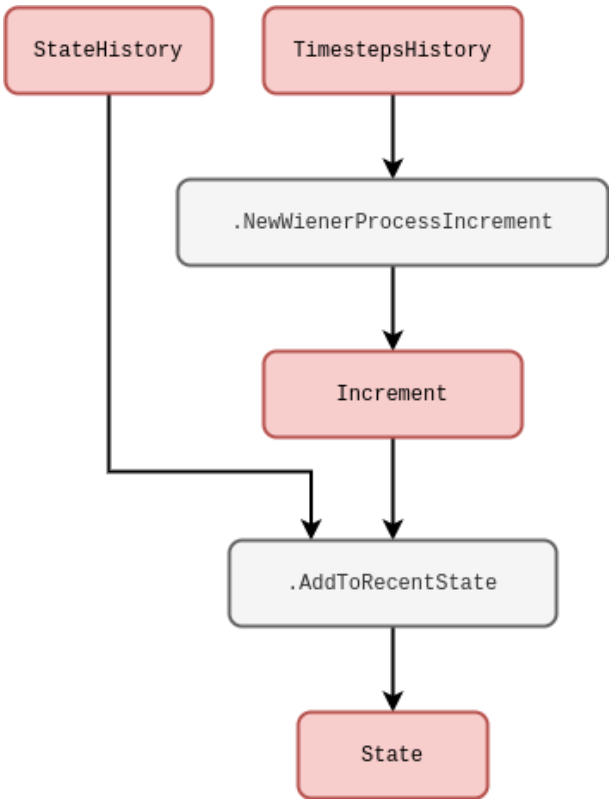

Figure 1.4: Schematic of code for a Wiener process.

In another example, to model *geometric Brownian motion noise* we would simply have to multiply X^i_{t} to the Wiener process like so

$$S^i_{\mathsf{t}+1}(X_{0:\mathsf{t}}, z, \mathsf{t}) = X^i_{\mathsf{t}}(W^i_{\mathsf{t}+1} - W^i_{\mathsf{t}})\,. \tag{1.6}$$

Here we have implicitly adopted the Itô interpretation to describe this stochastic integration. Given a carefully-defined integration scheme other interpretations of the noise would also be possible with our formalism too, e.g., Stratonovich[2] or others within the more general 'α-family' [8, 9, 10]. The

[2]Which would implictly give $S^i_{\mathsf{t}+1}(X_{0:\mathsf{t}}, z, \mathsf{t}) = (X^i_{\mathsf{t}+1} + X^i_{\mathsf{t}})(W^i_{\mathsf{t}+1} - W^i_{\mathsf{t}})/2$ for Eq. (1.6).

pseudocode for any of these should hoepfully be fairly straightforward to deduce based on the lines we've already written above.

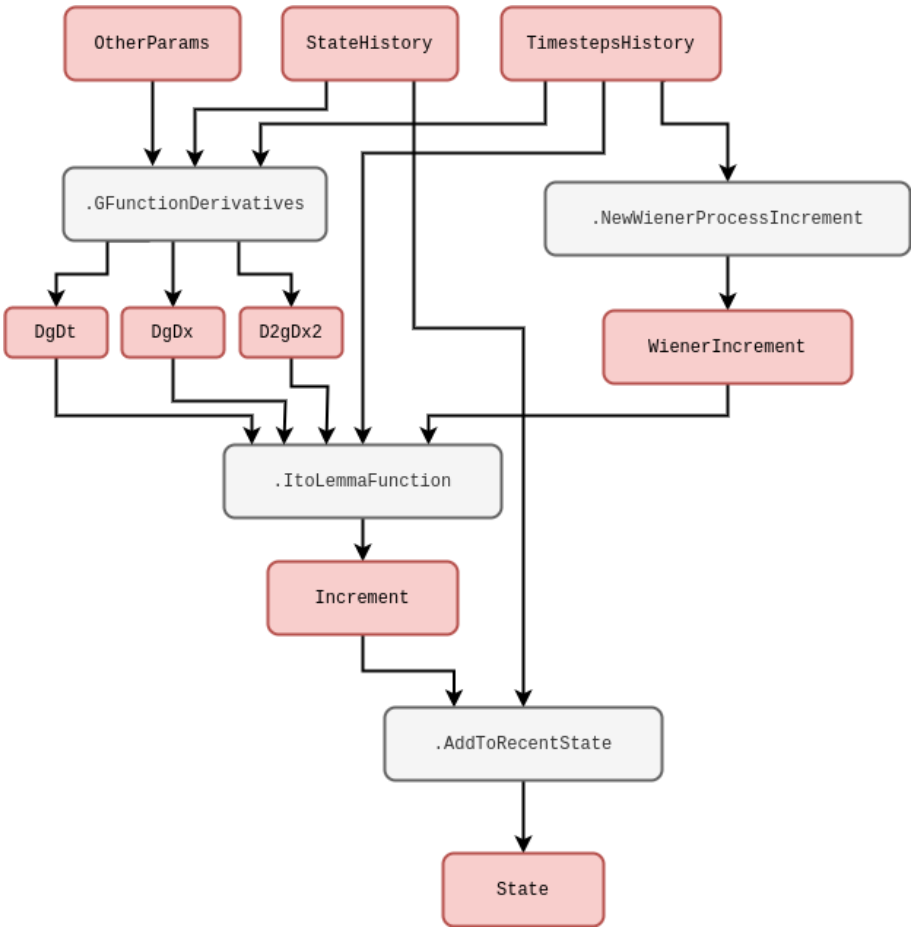

Figure 1.5: Schematic of code for Eq. (1.8).

We can imagine even more general processes that are still Markovian. One example of these in a single-dimension state space would be to define the noise through some general function of the Wiener process like so

$$S^0_{\mathtt{t}+1}(X_{0:\mathtt{t}}, z, \mathtt{t}) = g[W^0_{\mathtt{t}+1}, t(\mathtt{t}+1)] - g[W^0_{\mathtt{t}}, t(\mathtt{t})] \tag{1.7}$$

$$= \left[\frac{\partial g}{\partial t} + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}\right]\delta t(\mathtt{t}+1) + \frac{\partial g}{\partial x}(W^0_{\mathtt{t}+1} - W^0_{\mathtt{t}})\,, \tag{1.8}$$

where $g(x,t)$ is some continuous function of its arguments which has been expanded out with Itô's Lemma on the second line. Note also that the computations in Eq. (1.8) could be performed with numerical derivatives in principle, even if the function were extremely complicated. This is unlikely to be the best way to describe the process of interest, however, the mathematical expressions above

can still be made a bit more meaningful to the programmer in this way. The pseudocode in general would look something like Fig. 1.5.

Let's now look at a more complicated type of noise. For example, we might consider sampling from a *fractional Brownian motion* process $[B_H]_{\mathtt{t}}$, where H is known as the 'Hurst exponent'. Following Ref. [11], we can simulate this process in one of our state space dimensions by modifying the standard Wiener process by a fairly complicated integral factor which looks like this

$$S^0_{\mathtt{t}+1}(X_{0:\mathtt{t}}, z, \mathtt{t}) = \frac{(W^0_{\mathtt{t}+1} - W^0_{\mathtt{t}})}{\delta t(\mathtt{t})} \int_{t(\mathtt{t})}^{t(\mathtt{t}+1)} \mathrm{d}t' \frac{(t'-t)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} {}_2F_1\left(H-\frac{1}{2}; \frac{1}{2}-H; H+\frac{1}{2}; 1-\frac{t'}{t}\right), \quad (1.9)$$

where $S^0_{\mathtt{t}+1}(X_{0:\mathtt{t}}, z, \mathtt{t}) = [B_H]_{\mathtt{t}+1} - [B_H]_{\mathtt{t}}$. The integral in Eq. (1.9) can be approximated using an appropriate numerical procedure (like the trapezium rule, for instance). In the expression above, we have used the symbols ${}_2F_1$ and Γ to denote the ordinary hypergeometric and gamma functions, respectively. A computational form of this integral is illustrated in Fig. 1.6 to try and disentangle some of the mathematics as a program.

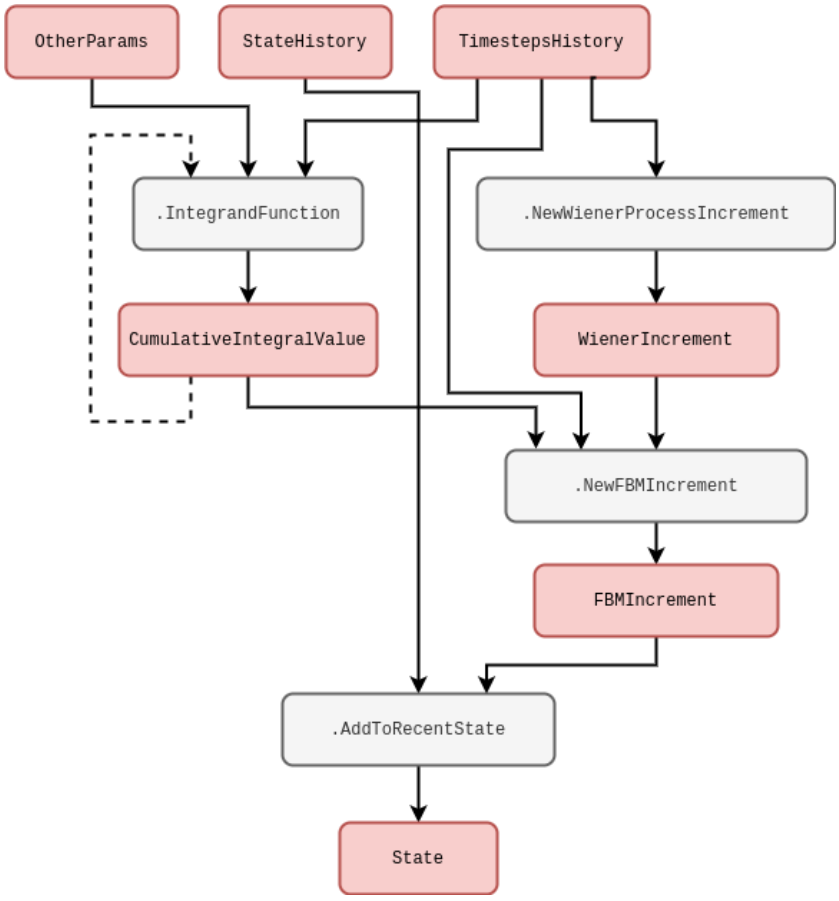

Figure 1.6: Schematic of code for Eq. (1.9).

So far we have mostly been discussing noises with continuous sample paths, but we can easily adapt our computation to discontinuous sample paths as well. For instance, *Poisson process noises* would generally take the form

$$S^i_{\mathsf{t}+1}(X_{0:\mathsf{t}}, z, \mathsf{t}) = [N_\lambda]^i_{\mathsf{t}+1} - [N_\lambda]^i_{\mathsf{t}}\,, \tag{1.10}$$

where $[N_\lambda]^i_{\mathsf{t}}$ is a sample from a Poisson process with rate λ. One can think of this process as counting the number of events which have occured up to the given interval of time, where the intervals between each succesive event are exponentially distributed with mean $1/\lambda$. Such a simple counting process could be simulated exactly by explicitly setting a newly-drawn exponential variate to the next continuous time jump $\delta t(\mathsf{t}+1)$ and iterating the counter. Other exact methods exist to handle more complicated processes involving more than one type of 'event', such as the Gillespie algorithm [12] — though these techniques are not always be applicable in every situation.

Is using step size variation always possible? If we consider a *time-inhomogeneous Poisson process noise*, which would generally take the form

$$S^i_{\mathsf{t}+1}(X_{0:\mathsf{t}}, z, \mathsf{t}) = [N_{\lambda(\mathsf{t}+1)}]^i_{\mathsf{t}+1} - [N_{\lambda(\mathsf{t})}]^i_{\mathsf{t}}\,, \tag{1.11}$$

the rate $\lambda(\mathsf{t})$ has become a deterministically-varying function in time. In this instance, it likely not be accurate to simulate this process by drawing exponential intervals with a mean of $1/\lambda(\mathsf{t})$ because this mean could have changed by the end of the interval which was drawn. An alternative approach (which is more generally capable of simulating jump processes but is an approximation) first uses a small time interval τ such that the most likely thing to happen in this period is nothing, and then the probability of the event occuring is simply given by

$$p(\mathsf{event}) = \frac{\lambda(\mathsf{t})}{\lambda(\mathsf{t}) + \frac{1}{\tau}}\,. \tag{1.12}$$

This idea can be applied to phenomena with an arbitrary number of events and works well as a generalised approach to event-based simulation, though its main limitation is worth remembering; in order to make the approximation good, τ often must be quite small and hence our simulator must churn through a lot of steps. From now on we'll refer to this well-known technique as the *rejection method*. Fig. 1.7 may also help to understand this concept from the programmer's perspective.

There are a few extensions to the simple Poisson process that introduce additional stochastic processes. *Cox (doubly-stochastic) processes*, for instance, are basically where we replace the time-dependent rate $\lambda(\mathsf{t})$ with independent samples from some other stochastic process $\Lambda(\mathsf{t})$. For example, a Neyman-Scott process [13] can be mapped as a special case of this because it uses a Poisson process on top of another Poisson process to create maps of spatially-distributed points. In our formalism, a two-state implementation of the Cox process noise would look like

$$S^0_{\mathsf{t}+1}(X_{0:\mathsf{t}}, z, \mathsf{t}) = \Lambda(\mathsf{t}+1) \tag{1.13}$$

$$S^1_{\mathsf{t}+1}(X_{0:\mathsf{t}}, z, \mathsf{t}) = [N_{S^0_{\mathsf{t}+1}}]^i_{\mathsf{t}+1} - [N_{S^0_{\mathsf{t}}}]^i_{\mathsf{t}}\,. \tag{1.14}$$

This process could be simulated using the pseudocode we wrote for the time-inhomogeneous Poisson process previously — where we would just replace `EventRateLambdaFunction` with a function that generates the stochastic rate $\Lambda(\mathsf{t})$.

Another extension is *compound Poisson process noise*, where it's the count values $[N_\lambda]^i_{\mathsf{t}}$ which are replaced by independent samples $[J_\lambda]^i_{\mathsf{t}}$ from another probability distribution, i.e.,

$$S^i_{\mathsf{t}+1}(X_{0:\mathsf{t}}, z, \mathsf{t}) = [J_\lambda]^i_{\mathsf{t}+1} - [J_\lambda]^i_{\mathsf{t}}\,. \tag{1.15}$$

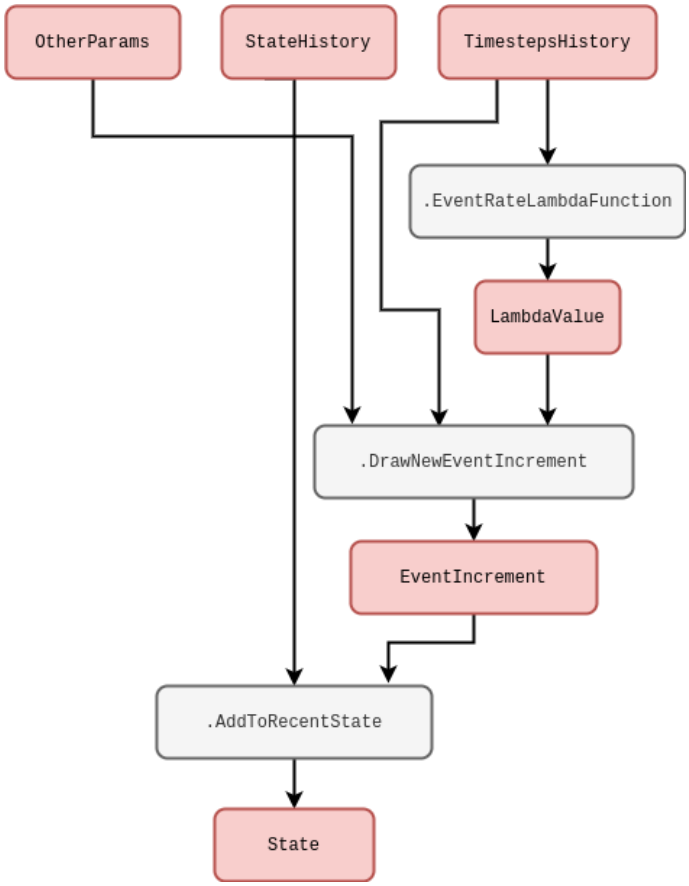

Figure 1.7: Schematic of code for an inhomogeneous Poisson process.

Note that the rejection method of Eq. (1.12) can be employed effectively to simulate any of these extensions as long as a sufficiently small τ is chosen. Once again, the pseudocode we wrote previously would be sufficient to simulate this process with one tweak: add into the `DrawNewEventIncrement` function the calling of a function which generates the $[J_\lambda]^i_{\mathtt{t}}$ samples and output these if the event occurs.

All of the examples we have discussed so far are Markovian. Given that we have explicitly constructed the formalism to handle non-Markovian phenomena as well, it would be worthwhile going some examples of this kind of process too. *Self-exciting process noises* would generally take the form

$$S^0_{\mathtt{t}+1}(X_{0:\mathtt{t}}, z, \mathtt{t}) = \mathcal{I}_{\mathtt{t}+1}(X_{0:\mathtt{t}}, z, \mathtt{t}) \tag{1.16}$$

$$S^1_{\mathtt{t}+1}(X_{0:\mathtt{t}}, z, \mathtt{t}) = [N_{S^0_{\mathtt{t}+1}}]^i_{\mathtt{t}+1} - [N_{S^0_{\mathtt{t}}}]^i_{\mathtt{t}}\,, \tag{1.17}$$

where the stochastic rate $\mathcal{I}_{\mathtt{t}+1}(X_{0:\mathtt{t}}, z, \mathtt{t})$ now depends on the history explicitly. Amongst other

potential inputs we can see, e.g., Hawkes processes [14] as an example of above by substituting

$$\mathcal{I}_{\mathsf{t}+1}(X_{0:\mathsf{t}}, z, \mathsf{t}) = \mu + \sum_{\mathsf{t}'=0}^{\mathsf{t}} \gamma[t(\mathsf{t}) - t(\mathsf{t}')] S^1_{\mathsf{t}'} \,, \tag{1.18}$$

where γ is the 'exciting kernel' and μ is some constant background rate. In order to simulate a Hawkes process using our formalism, the pseudocode would be something like Fig. 1.8.

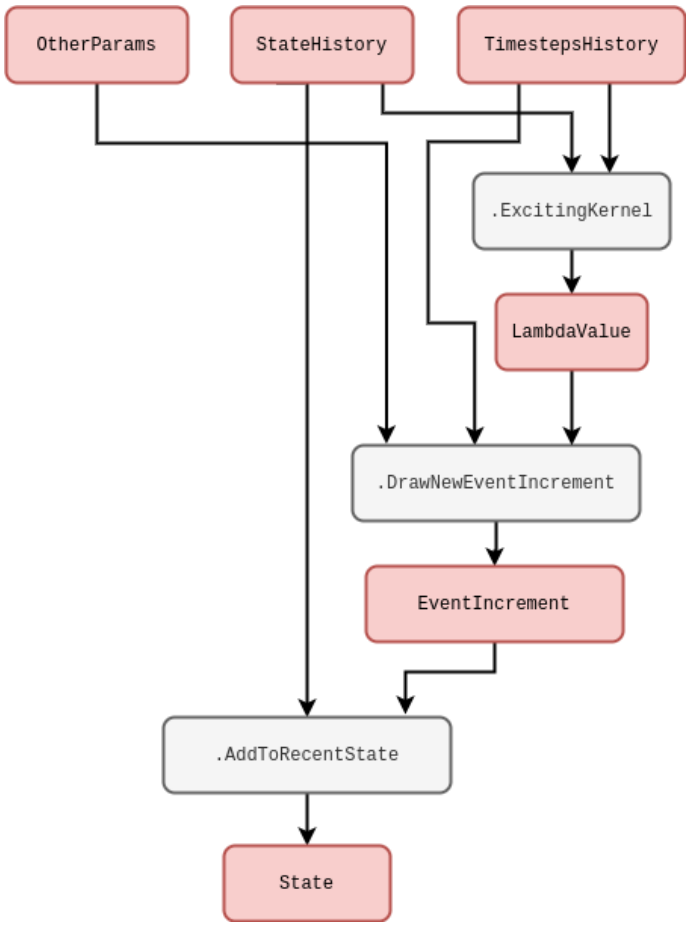

Figure 1.8: Schematic of code for a Hawkes process.

Note that this idea of integration kernels could also be applied back to our Wiener process. For example, another type of non-Markovian phenomenon that frequently arises across physical and life systems integrates the Wiener process history like so

$$S^0_{\mathsf{t}+1}(X_{0:\mathsf{t}}, z, \mathsf{t}) = W^0_{\mathsf{t}+1} - W^0_{\mathsf{t}} \tag{1.19}$$

$$S^1_{\mathsf{t}+1}(X_{0:\mathsf{t}}, z, \mathsf{t}) = u \sum_{\mathsf{t}'=0}^{\mathsf{t}} e^{-u[t(\mathsf{t})-t(\mathsf{t}')]} S^0_{\mathsf{t}'} \,, \tag{1.20}$$

where u is inversely proportional to the length of memory in continuous time.

1.2 Software design

So we've proposed a computational formalism and then studied it in more detail to demonstrate that it can cope with a variety of different stochastic phenomena. Now we're ready to summarise what we want the stochadex software package to be able to do. But what's so complicated about Eq. (1.1)? Can't we just implement an iterative algorithm with a single function? It's true that the fundamental concept is very straightforward, but as we'll discuss in due course; the stochadex needs to have a lot of configurable features so that it's applicable in different situations. Ideally, the stochadex sampler should be designed to try and maintain a balance between performance and flexibility of utilisation.

If we begin with the obvious first set of criteria; we want to be able to freely configure the iteration function F of Eq. (1.1) and the timestep function t of Eq. (1.2) so that any process we want can be described. The point at which a simulation stops can also depend on some algorithm termination condition which the user should be able to specify up-front.

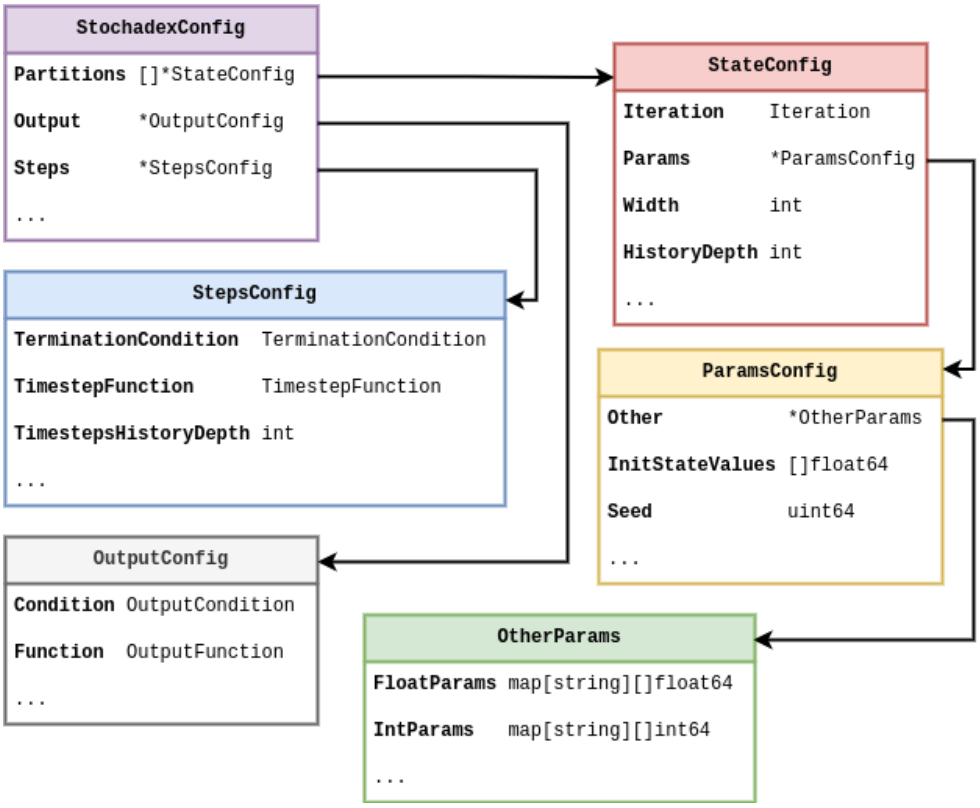

Figure 1.9: A relational summary of the configuration data types in the stochadex.

Once the user has written the code to create these functions for the stochadex, we want to

then be able to recall them in future only with configuration files while maintaining the possibility of changing their simulation run parameters. This flexibility should facilitate our uses for the simulation later in the book, and from this perspective it also makes sense that the parameters should include the random seed and initial state value.

The state history matrix X should be configurable in terms of its number of rows — what we'll call the 'state width' — and its number of columns — what we'll call the 'state history depth'. If we were to keep increasing the state width up to millions of elements or more, it's likely that on most machines the algorithm performance would grind to a halt when trying to iterate over the resulting X within a single thread. Hence, before the algorithm or its performance in any more detail, we can pre-empt the requirement that X should represented in computer memory by a set of partitioned matrices which are all capable of communicating to one-another downstream. In this paradigm, we'd like the user to be able to configure which state partitions are able to communicate with each other without having to write any new code.

For convenience, it seems sensible to also make the outputs from stochadex runs configurable. A user should be able to change the form of output that they want through, e.g., some specified function of X at the time of outputting data. The times that the stochadex should output this data can also be decided by some user-specified condition so that the frequency of output is fully configurable as well.

In summary, we've put together a schematic of configuration data types and their relationships in Fig. 1.9. In this diagram there is some indication of the data type that we propose to store each piece information in (in Go syntax), and the diagram as a whole should serve as a useful guide to the basic structure of configuration files for the stochadex.

It's clear that in order to simulate Eq. (1.1), we need an interative algorithm which reapplies a user-specified function to the continually-updated history. But let's now return to the point we made earlier about how the performance of such an algorithm will depend on the size of the state history matrix X. The key bit of the algorithm design that isn't so straightforward is: how do we sucessfully split this state history up into separate partitions in memory while still enabling them to communicate effectively with each other? Other generalised simulation frameworks — such as SimPy [15], StoSpa [16] and FLAME GPU [17] — have all approached this problem in different ways, and with different software architectures.

In Fig. 1.10 we've illustrated what a loop involving separate state partitions looks like in the stochadex simulator. Each partition is handled by concurrently running execution threads of the same process, while a separate process may be used to handle the outputs from the algorithm. As the diagram shows, the main sequence of each loop iteration follows the pattern:

1. The `PartitionCoordinator` requests more iterations from each state partition by sending an `IteratorInputMessage` to a concurrently running goroutine.

2. The `StateIterator` in each goroutine executes the iteration and stores the resulting state update in a variable.

3. Once all of the iterations have been completed, the `PartitionCoordinator` then requests each goroutine to update its relevant partition of the state history by sending another `IteratorInputMessage` to each.

This pattern ensures that no partition has access to values in the state history which are out of sync with its current state in time, and hence prevents anachronisms from occuring in the overall state iteration.

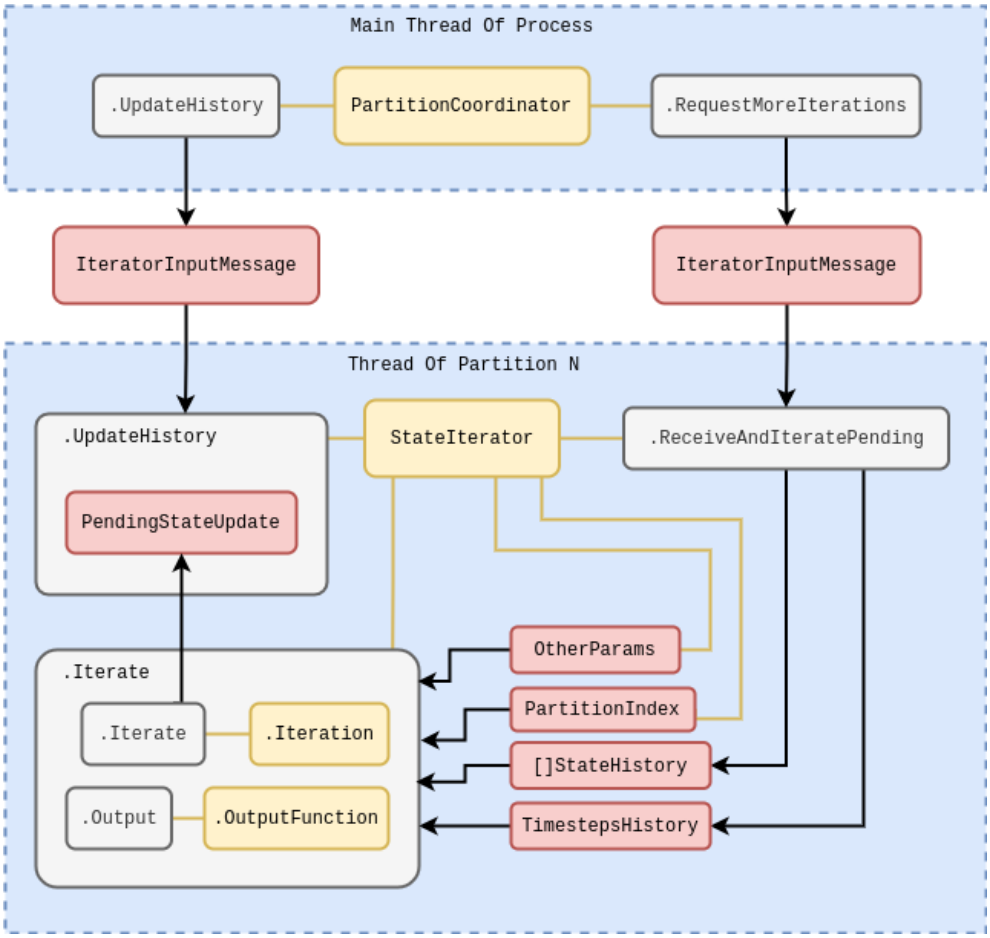

Figure 1.10: Schematic for a step of the stochadex simulation algorithm.

It's also worth noting that while Fig. 1.10 illustrates only a single process; it's obviously true that we may run many of these whole diagrams at once to parallelise generating independent realisations of the simulation, if necessary.

As we stated at the beginning of this chapter: the full implementation of the stochadex can be found on GitHub by following this link: https://github.com/umbralcalc/stochadex. Users can build the main binary executable of this repository and determine what configuration of the stochadex they would like to have through config at runtime (one can infer these configurations from Fig. 1.9). As Go is a statically typed language, this level of flexibility has been achieved using code templating proceeding runtime build and execution via `go run` 'under-the-hood'. Users who find this particular execution pattern undesirable can also use all of the stochadex types, tools and methods as part of a standard library import.

In order to debug the simulation code and gain a more intuitive understanding of the outputs from a model as it is being developed, we have also written a lightweight frontend dashboard

React [18] app in TypeScript to visualise any stochadex simulation as it is running. This dashboard can be launched by passing config at runtime to the main stochadex executable, and we have illustrated how all this fits together in a flowchart shown in Fig. 1.11.

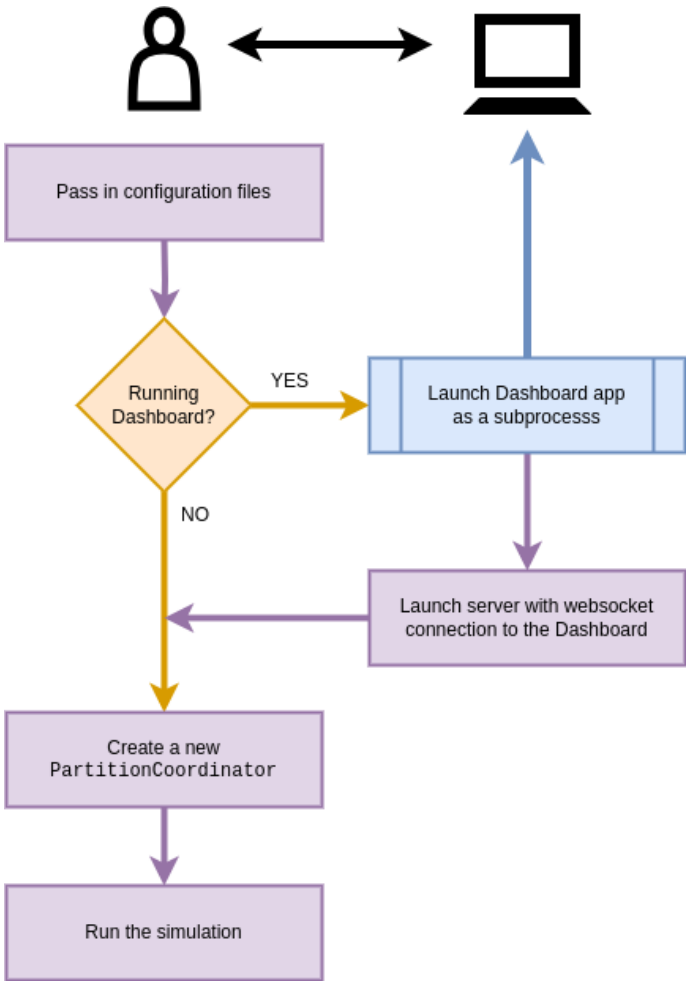

Figure 1.11: A diagram of the main stochadex binary executable.

2

Numerical time evolution of probabilities

Concept. To extend the formalism that we developed in previous chapters to enable the numerical simulation of state probabilities. In the general case this is a very intensive computation to perform, so in this chapter we shall also discuss the practical limitations of this 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.

If we wanted to just look at the distribution over the latest row $X_{\mathtt{t}+1} = x$, we could achieve this

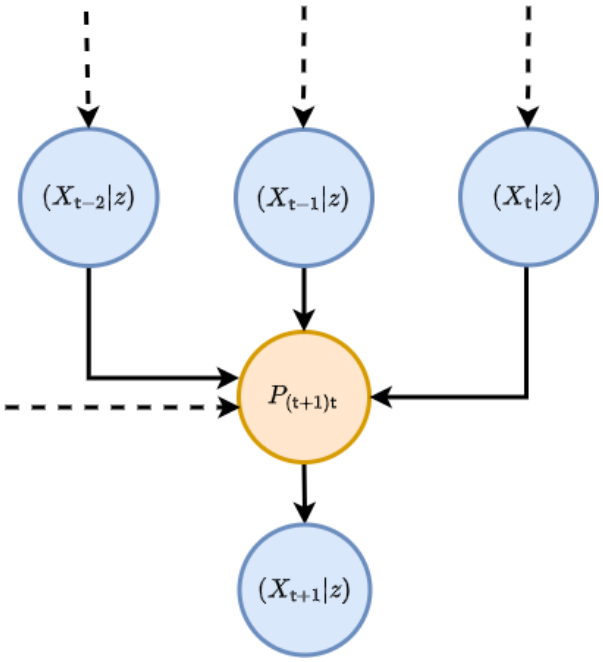

Figure 2.1: Graph representation of Eqs. (2.1) and (2.2).

through marginalisation over all of the previous matrix rows in Eq. (2.1) like this

$$P_{t+1}(x|z) = \int_{\Omega_t} \mathrm{d}X' P_{t+1}(X|z) = \int_{\Omega_t} \mathrm{d}X' P_t(X'|z) P_{(t+1)t}(x|X', z)\,. \tag{2.2}$$

But what is Ω_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. (2.2) is our shorthand notation throughout the book for taking a sum of sub-domain integrals over each matrix row; where each row measure is a Cartesian product of n elements (a Lebesgue measure), i.e.,

$$\int_{\Omega_t} \mathrm{d}X' = \frac{1}{t} \sum_{t'=0}^{t} \int_{\omega_{t'}} \mathrm{d}^n x' = \frac{1}{t} \sum_{t'=0}^{t} \int_{\omega_{t'}} \prod_{i=0}^{n} \mathrm{d}(x')^i\,, \tag{2.3}$$

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

$$\frac{1}{t} \sum_{t'=0}^{t} \to \frac{1}{t(t)} \sum_{t'=0}^{t} \delta t(t')\,. \tag{2.4}$$

To try and understand what Eqs. (2.1) and (2.2) 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.

Without loss of generality, we can relate the latest probabilities to those from deeper into the past by chaining conditional probabilities together in a non-Markovian equivalent of the Chapman-Kolmogorov equation

$$\begin{aligned}
P_{\mathsf{t}+1}(x|z) &= \int_{\Omega_{\mathsf{t}-1}} \mathrm{d}X'' P_{\mathsf{t}-1}(X''|z) \int_{\omega_{\mathsf{t}}} \mathrm{d}^n x' P_{\mathsf{t}(\mathsf{t}-1)}(x'|X'',z) P_{(\mathsf{t}+1)\mathsf{t}}(x|X',z) \\
&= \int_{\Omega_{\mathsf{t}-2}} \mathrm{d}X''' P_{\mathsf{t}-2}(X'''|z) \int_{\omega_{\mathsf{t}-1}} \mathrm{d}^n x'' P_{(\mathsf{t}-1)(\mathsf{t}-2)}(x''|X''',z) \\
&\qquad\qquad \times \int_{\omega_{\mathsf{t}}} \mathrm{d}^n x' P_{\mathsf{t}(\mathsf{t}-1)}(x'|X'',z) P_{(\mathsf{t}+1)\mathsf{t}}(x|X',z) \\
&= \dots \\
&= \int_{\Omega_{\mathsf{t}-\mathsf{s}}} \mathrm{d}X''' P_{\mathsf{t}-\mathsf{s}}(X'''|z) \prod_{\mathsf{s}'=0}^{\mathsf{s}-1} \left\{ \int_{\omega_{\mathsf{t}-\mathsf{s}'}} \mathrm{d}^n x' P_{(\mathsf{t}-\mathsf{s}')(\mathsf{t}-\mathsf{s}'-1)}(x'|X'',z) \right\} P_{(\mathsf{t}+1)\mathsf{t}}(x|X',z)\,.
\end{aligned} \tag{2.5}$$

Depending on the temporal correlation structure of the process, the conditional probabilities can be factorised. For example, processes with second or third-order temporal correlations would be described by the following expressions

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

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

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

$$\mathcal{J}^i_{\mathsf{t}+1} = -[P]^i_{\mathsf{t}} + \frac{1}{\mathsf{t}} \sum_{\mathsf{t}'=0}^{\mathsf{t}} \sum_{i'=0}^{N} \Delta x [P]^{i'}_{\mathsf{t}'} [P]^{ii'}_{(\mathsf{t}+1)\mathsf{t}'} \tag{2.8}$$

$$\mathcal{J}^i_{\mathsf{t}+1} = -[P]^i_{\mathsf{t}} + \frac{1}{\mathsf{t}} \sum_{\mathsf{t}'=1}^{\mathsf{t}} \frac{1}{\mathsf{t}'-1} \sum_{\mathsf{t}''=0}^{\mathsf{t}'-1} \sum_{i'=0}^{N} \sum_{i''=0}^{N} \Delta x^2 [P]^{i''}_{\mathsf{t}''} [P]^{i'i''}_{\mathsf{t}'\mathsf{t}''} [P]^{ii'i''}_{(\mathsf{t}+1)\mathsf{t}'\mathsf{t}''}\,. \tag{2.9}$$

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

$$[P]^{ii''}_{(\mathsf{t}+1)\mathsf{t}''} = \frac{1}{\mathsf{t}} \sum_{\mathsf{t}'=1}^{\mathsf{t}} \sum_{i'=0}^{N} \Delta x [P]^{i'i''}_{\mathsf{t}'\mathsf{t}''} [P]^{ii'i''}_{(\mathsf{t}+1)\mathsf{t}'\mathsf{t}''}\,. \tag{2.10}$$

What other classes of process can be described by Eqs. (2.1) and (2.2)? For Markovian phenomena, the equations no longer depend on timesteps older than the immediately previous one, hence Eq. (2.2) reduces to just

$$P_{t+1}(x|z) = \int_{\omega_t} \mathrm{d}^n x' \, P_t(x'|z) P_{(t+1)t}(x|x', z) \,. \tag{2.11}$$

An analog of Eq. (2.2) 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

$$P_{t+1}(x|z) = \sum_{\Omega_t} P_t(X'|z) P_{(t+1)t}(x|X', z) \,, \tag{2.12}$$

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

- Add some diagrams for the higher-order correlation expressions.
- Add a software design section and some examples.

3

Empirical probabilistic reweighting algorithms

Concept. To extend the formalism that we developed in previous chapters to enable the empirical emulation of real-world data via a probabilistic reweighting. This technique should enable a researcher to model complex dynamical trends in the data very well; at the cost of making the abstract interpretation of the model less immediately comprehensible than the statistical inference models in some proceeding chapters. For the mathematically-inclined, this chapter will take a detailed look at how our formalism can be extended to focus on probabilistic reweightings and their optimisation using real-world data. For the programmers, the software described in this chapter lives in the public Git repository: https://github.com/umbralcalc/learnadex.

3.1 Probabilistic formalism

The key distinction between the methods that we will develop in this chapter and the ones in the proceeding chapters 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.

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 new to the technical literature at all. Our overall framework draws on influences from Empirical Dynamical Modeling (EDM) [19], some classic nonparametric local regression techniques — such as LOWESS/Savitzky-Golay filtering [20] — and also Gaussian processes [21] 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. In addition, readers well-versed in machine learning will note that our software design and mathematical formalism reflects our preference for 'online' learning[1] in the context of time series prediction, in contrast to some of the more standard frameworks.

When trying robustly assess how far a model is from accurately describing a set of real-world data, trying to use only generated samples of the model process can be diffcult. Instead, in this section, we are going to extend this formalism to look at how probability theory can help with this data comparison problem in a systematic way. In order to do this, we need to return to the probabilistic formalism which we discussed in the previous chapter.

Given the general master equation $P_{\mathsf{t}+1}(X|z) = P_{\mathsf{t}}(X'|z)P_{(\mathsf{t}+1)\mathsf{t}}(x|X', z)$, if we wanted to compute the mean $M_{\mathsf{t}+1}(z)$ of the distribution over the matrix row corresponding to time $(\mathsf{t}+1)$, it would be straightforward to just multiply both sides by x and integrate over it in its $\omega_{\mathsf{t}+1}$ sub-domain. However, there is another similar expression for the mean that we can derive under certain conditions which will be valuable to us when developing the empirical reweighting. If the probability distribution over each row of the state history matrix is *stationary* — meaning that $P_{\mathsf{t}+1}(x|z) = P_{\mathsf{t}'}(x|z)$ — it's possible to derive

$$M_{\mathsf{t}+1}(z) = \int_{\omega_{\mathsf{t}+1}} \mathrm{d}^n x\, x\, P_{\mathsf{t}+1}(x|z) = \frac{1}{\mathsf{t}} \sum_{\mathsf{t}'=0}^{\mathsf{t}} \int_{\omega_{\mathsf{t}'}} \mathrm{d}^n x'\, x'\, P_{\mathsf{t}'}(x'|z) \int_{\omega_{\mathsf{t}+1}} \mathrm{d}^n x\, P_{(\mathsf{t}+1)\mathsf{t}'}(x|x', z)\,. \tag{3.1}$$

To see that Eq. (3.1) is true, first note that a joint distribution over both x and x' can be derived like this $P_{(\mathsf{t}+1)\mathsf{t}'}(x, x'|z) = P_{(\mathsf{t}+1)\mathsf{t}'}(x|x', z)P_{\mathsf{t}'}(x'|z)$. Secondly, note that this joint distribution will always allow variable swaps trivially like this $P_{(\mathsf{t}+1)\mathsf{t}'}(x, x'|z) = P_{\mathsf{t}'(\mathsf{t}+1)}(x', x|z)$. Then, lastly, note that stationarity of $P_{\mathsf{t}+1}(x|z) = P_{\mathsf{t}'}(x|z)$ means

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

where we've used the trivial variable swap and integration variable relabelling to arrive at the second equality in the expressions above.

The standard covariance matrix elements can also be computed in a similar fashion

$$\begin{aligned}
C^{ij}_{\mathsf{t}+1}(z) &= \int_{\omega_{\mathsf{t}+1}} \mathrm{d}^n x\, [x - M_{\mathsf{t}+1}(z)]^i [x - M_{\mathsf{t}+1}(z)]^j P_{\mathsf{t}+1}(x|z) \\
&= \frac{1}{\mathsf{t}} \sum_{\mathsf{t}'=0}^{\mathsf{t}} \int_{\omega_{\mathsf{t}'}} \mathrm{d}^n x'\, [x' - M_{\mathsf{t}+1}(z)]^i [x' - M_{\mathsf{t}+1}(z)]^j\, P_{\mathsf{t}'}(x'|z) \int_{\omega_{\mathsf{t}+1}} \mathrm{d}^n x\, P_{(\mathsf{t}+1)\mathsf{t}'}(x|x', z)\,.
\end{aligned} \tag{3.2}$$

While they look quite abstract, Eqs. (3.1) and (3.2) express the core idea behind how our probabilistic reweighting will function. By assuming a stationary distribution, we gain the ability to directly estimate the statistics of the probability distribution of the next sample from the stochastic

[1] We'll explain this later in the chapter.

process $P_{\mathsf{t}+1}(x|z)$ from past samples it may have in empirical data; which are represented here by $P_{\mathsf{t}'}(x'|z)$. More on this later.

It may seem needlessly more complex to deal with probability distributions over matrices instead of marginal distributions over separate vectors which represent the rows of these matrices. However, the matrix description is more general, and it turns out to be quite neat to describe correlations across time (which would be lost by marginal distributions over row vectors). In order to study these out-of-time-order correlations, we need only consider using the statistical moments computed across the columns of X. For example, pairwise correlations can be analysed through the out-of-time-order covariance matrix elements

$$C^{ij}_{(\mathsf{t}+1)\mathsf{t}'}(z) = \int_{\omega_{\mathsf{t}+1}} \mathrm{d}^n x \int_{\omega_{\mathsf{t}'}} \mathrm{d}^n x' [x - M_{\mathsf{t}+1}(z)]^i [x' - M_{\mathsf{t}'}(z)]^j P_{\mathsf{t}+1}(X|z)\,. \tag{3.3}$$

- Need to finish the descriptions in this section to move onto the next...

3.2 Online learning the optimal reweighting

Probabilistic reweighting depends on the stationarity of $P_{\mathsf{t}+1}(x|z) = P_{\mathsf{t}'}(x|z)$ such that, e.g., Eq. (3.1) is applicable. The core idea behind it is to represent the past distribution of state values $P_{\mathsf{t}'}(x'|z)$ with the samples from a real time series dataset. If the user then specifies a good model for the relationships in this data by providing a weighting function which returns the conditional probability mass

$$\mathsf{w}_{\mathsf{t}'}(y, z) = \int_{\omega_{\mathsf{t}+1}} \mathrm{d}^n x\, P_{(\mathsf{t}+1)\mathsf{t}'}(x|x'{=}y, z)\,, \tag{3.4}$$

we can apply this as a *reweighting* of the historical time series samples to estimate any statistics of interest. Taking Eqs. (3.1) and (3.2) as the examples; we are essentially approximating these integrals through weighted sample estimations like this

$$M_{\mathsf{t}+1}(z) \simeq \frac{1}{\mathsf{t}} \sum_{\mathsf{t}'=0}^{\mathsf{t}} Y_{\mathsf{t}'} \mathsf{w}_{\mathsf{t}'}(Y_{\mathsf{t}'}, z) \tag{3.5}$$

$$C^{ij}_{\mathsf{t}+1}(z) \simeq \frac{1}{\mathsf{t}} \sum_{\mathsf{t}'=0}^{\mathsf{t}} [Y_{\mathsf{t}'} - M_{\mathsf{t}+1}(z)]^i [Y_{\mathsf{t}'} - M_{\mathsf{t}+1}(z)]^j \mathsf{w}_{\mathsf{t}'}(Y_{\mathsf{t}'}, z)\,, \tag{3.6}$$

where we have defined the data matrix Y with rows $Y_{\mathsf{t}+1}, Y_{\mathsf{t}}, \ldots$, each of which representing specific observations of the rows in X at each point in time from a real dataset.

Our goal in this section will be to learn the optimal reweighting function $\mathsf{w}_{\mathsf{t}'}(Y_{\mathsf{t}'}, z)$ with respect to z, i.e., the ones which most accurately represent a provided dataset. But before we think about the various kinds of conditional probability we could use, we need to think about how to connect the post-reweighting statistics to the data by defining an objective function.

If the mean is a sufficient statistic for the distribution which describes the data, a choice of, e.g., Exponential, Poisson or Binomial distribution could be used where the mean is estimated directly from the time series using Eq. (3.1), given a conditional probability $P_{(\mathsf{t}+1)\mathsf{t}'}(x|x', z)$. Extending this

idea further to include distributions which also require a variance to be known, e.g., the Normal, Gamma or Negative Binomial distributions could be used where the variance (and/or covariance) could be estimated using Eq. (3.2). These are just a few simple examples of distributions that can link the estimated statistics from Eqs. (3.1) and (3.2) to a time series dataset. However, the algorithmic framework is very general to whatever choice of 'data linking' distribution that a researcher might need.

We should probably make what we've just said a little more mathematically concrete. We can define $P_{\mathsf{t}+1}[y; M_{\mathsf{t}+1}(z), C_{\mathsf{t}+1}(z), \dots]$ as representing the likelihood of $y = Y_{\mathsf{t}+1}$ given the estimated statistics from Eqs. (3.1) and (3.2) (and maybe higher-orders). Note that in order to do this, we need to identify the x' and t' values that are used to estimate, e.g., $M_{\mathsf{t}+1}(z)$ with the past data values which are observed in the dataset time series itself. Now that we have this likelihood, we can immediately evaluate an objective function (a cumulative log-likelihood) that we might seek to optimise over for a given dataset

$$\ln \mathcal{L}_{\mathsf{t}+1}(Y|z) = \sum_{\mathsf{t}'=0}^{\mathsf{t}+1} \ln P_{\mathsf{t}'}[y; M_{\mathsf{t}'}(z), C_{\mathsf{t}'}(z), \dots] \,, \tag{3.7}$$

where the summation continues until all of the past measurements $Y_{\mathsf{t}+1}, Y_{\mathsf{t}}, \dots$ which exist as rows in the data matrix Y have been taken into account. The code to compute this objective function follows the schematic we have provided in Fig. 3.1.

In order to specify what $P_{(\mathsf{t}+1)\mathsf{t}'}(x|x', z)$ is, it's quite natural to define a set of hyperparameters for the elements of z. To get a sense of how the data-linking function relates to these hyperparameters, it's instructive to consider an example. One generally-applicable option for the conditional probability could be a purely time-dependent kernel

$$P_{(\mathsf{t}+1)\mathsf{t}'}(x|x', z) \propto \mathcal{K}(z, \mathsf{t}+1, \mathsf{t}') \,, \tag{3.8}$$

and the data-linking distribution, e.g., could be a Gaussian

$$P_{\mathsf{t}+1}[y; M_{\mathsf{t}+1}(z), C_{\mathsf{t}+1}(z), \dots] = \mathsf{MultivariateNormalPDF}[y; M_{\mathsf{t}+1}(z), C_{\mathsf{t}+1}(z)] \,. \tag{3.9}$$

It's worth pointing out that other machine learning frameworks could easily be used to model these conditional probabilities. For example, neural networks could be used to infer the optimal reweighting scheme and this would still allow us to use the data-linking distribution.[2] It would still be desirable to keep the data-linking distribution as it can usually be sampled from very easily — something that can be quite difficult to achieve with a purely machine learning-based representation of the distribution. Sampling itself could even be made more flexible by leveraging a Variational Autoencoder (VAE) [23]; these use neural networks not just on the compression (or 'encode') step to estimate the statistics but also use them as a layer between the sample from the data distribution model and the output (the 'decode' step).

In the case of Eqs. (3.8) and (3.9) above, the hyperparameters that would be optimised could relate to the kernel in a wide variety of ways. Optimising them would make our optimised reweighting very similar to (but not quite the same as) evaluating maximum a posteriori (MAP) of a Gaussian process regression. The main differences here are that the mean of a Gaussian process as a function of time is typically included within z, and hence must be obtained through optimisation. In

[2] One can think of using this neural network-based reweighting scheme as similar to constructing a normalising flow model [22] with an autoregressive layer. Invertibility and further network structural constraints mean that these are not exactly equivalent, however.

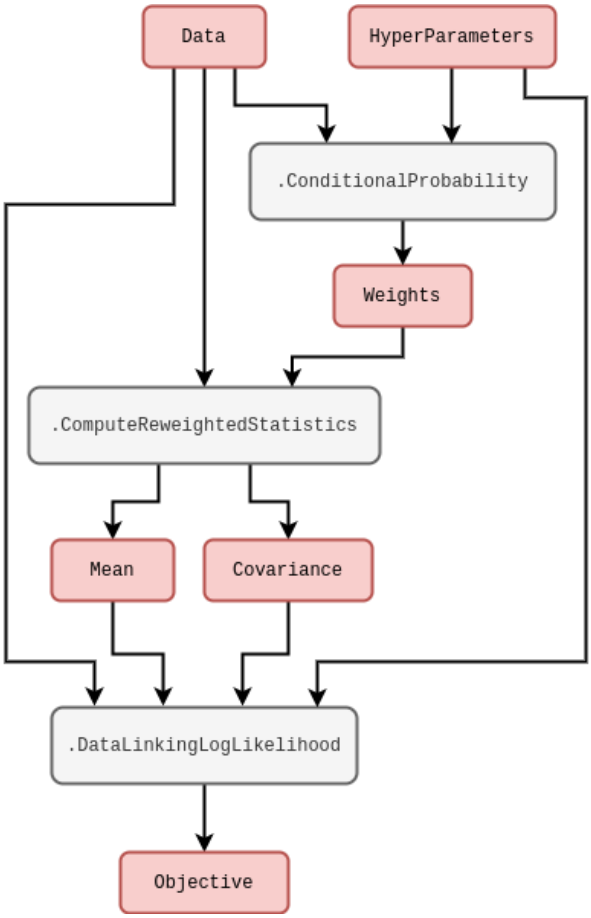

Figure 3.1: Code schematic of the probability reweighting optimisation.

contrast, our methodology relies on the fact that the mean estimator can be computed directly by weighted sample estimation and can then be fed to some data-linking distribution. By doing it this way, we enable many different kinds of data to be described by the same underlying conditional probability-based reweighting and also make incorporating future data into the current model much easier then is the case for a standard Gaussian process (which would require re-optimising with respect to the next data point every time). The time-dependent kernel we have chosen in Eq. (3.8) also only represents one particular choice, but we could consider a wide range of state-dependent conditional probability weightings for the algorithm as well.

As another form of flexibility; we could also try adapting the data-linking distributions to include an intercept term and linear coefficients for the statistics which are passed to it. These could then be treated as additional hyperparameters and optimised jointly with the others if there is sufficient constraining power in the data.

The optimisation approach that we choose to use for obtaining the best hyperparameters in the conditional probability of Eq. (3.7) will depend on a few factors. For example, if the number of hy-

perparameters is relatively low, but their gradients are difficult to calculate exactly; then a gradient-free optimiser (such as the Nelder-Mead [24] method or something like a particle swarm [25, 26]) would likely be the most effective choice. On the other hand, when the number of hyperparameters ends up being relatively large, it's usually quite desriable to utilise the gradients in algorithms like vanilla Stochastic Gradient Descent [27] (SGD) or Adam [28].

If the gradients of Eq. (3.7) are needed, we can always factorise each derivative with respect to hyperparameter z^i in the following way through the chain rule

$$\frac{\partial}{\partial z^i} \ln \mathcal{L}_{\mathfrak{t}+1}(Y|z) = \sum_{\mathfrak{t}'=0}^{\mathfrak{t}+1} \frac{\partial M_{\mathfrak{t}'}}{\partial z^i} \frac{\partial}{\partial M_{\mathfrak{t}'}} \ln P_{\mathfrak{t}'}[y; M_{\mathfrak{t}'}(z), C_{\mathfrak{t}'}(z), \dots] + \sum_{\mathfrak{t}'=0}^{\mathfrak{t}+1} \frac{\partial C_{\mathfrak{t}'}}{\partial z^i} \frac{\partial}{\partial C_{\mathfrak{t}'}} \ln P_{\mathfrak{t}'}[y; M_{\mathfrak{t}'}(z), C_{\mathfrak{t}'}(z), \dots] . \tag{3.10}$$

By factoring derivatives in this manner, the computation can be separated into two parts: the derivatives with respect to $M_{\mathfrak{t}'}$ and $C_{\mathfrak{t}'}$, which are typically quite straightforward; and the derivatives with respect to z elements, which typically need a more involved calculation depending on the model. Incidentally, this separation also neatly lends itself to abstracting gradient calculations as having a simpler, general purpose component that can be built directly into a library of data models and a more complex, model-specific component that the user must specify.

Before moving on to the software design aspects, we need to consider how we might structure learning by optimisation of Eq. (3.7) for a sequence of observations in time. One of the issues that can arise when learning streams of data is 'concept drift'. In our context, this would be when the optimal value for z does not match the optimal value at some later point in time. In order to solve this issue, our learning algorithm should track an up-to-date optimal value for z as data is continually passed into it. Iteratively updating the optimal parameters as new data is ingested into the objective function is typically called 'online learning' [29, 1], in contrast to 'offline learning' which would correspond to learning an optimal z only once with the entire dataset provided upfront.[3]

3.3 Software design

Let's now take a step back from the specifics of the probabilistic reweighting algorithm to introduce our new sofware package for this part of the book: the 'learnadex'. At its core, the learnadex algorithm adapts the stochadex iteration engine to iterate through streams of data in order to accumulate a global objective function value with respect to that data. The user may then choose which optimisation algorithm (or write their own) to use in order to leverage this objective for learning a better representation of the data.

As we discussed at the end of the last section, the algorithms in the learnadex are all applied in an 'online' fashion — refitting for the optimal hyperparameters z as new data is streamed into them. A challenging aspect of online learning is in managing the computational expense of recomputing the optimal value for z after each new datapoint is sent. To help with this; the user may configure the algorithm recompute the optimum value after larger batches of data have been ingested. The last value of optimum z will also frequently be close to the next optimum in the sequence, so using

[3] This book will mostly be focussed on using online learning techniques due to the inherently sequential nature of stochastic processes.

the former as the initial input into the optimisation routine for the latter is typically very valuable for aiding efficiency.

Reusing the `PartitionCoordinator` code of the stochadex to facilitate online learning makes neat use of software which has already been designed and tested in earlier chapters of this book. However, in order to fully achieve this, a few minor extensions to the typing structures and code abstractions are necessary; as we show in Fig. 3.2. To start with, we separate out 'learning' from the kind of optimiser in the overall config so as to enable multiple optimisation algorithms to be used for the same learning problem. The hyperparameters that define that optimisation problem domain can be determined by the user with an extension to the `OtherParams` object so that it includes some optional Boolean masks over the parameters, i.e., `OtherParams.FloatParamMask` and `OtherParams.IntParamMask`. These masks are used to extract the parameters of interest, which can then be flattened and formatted to fit into any generic optimisation algorithm.

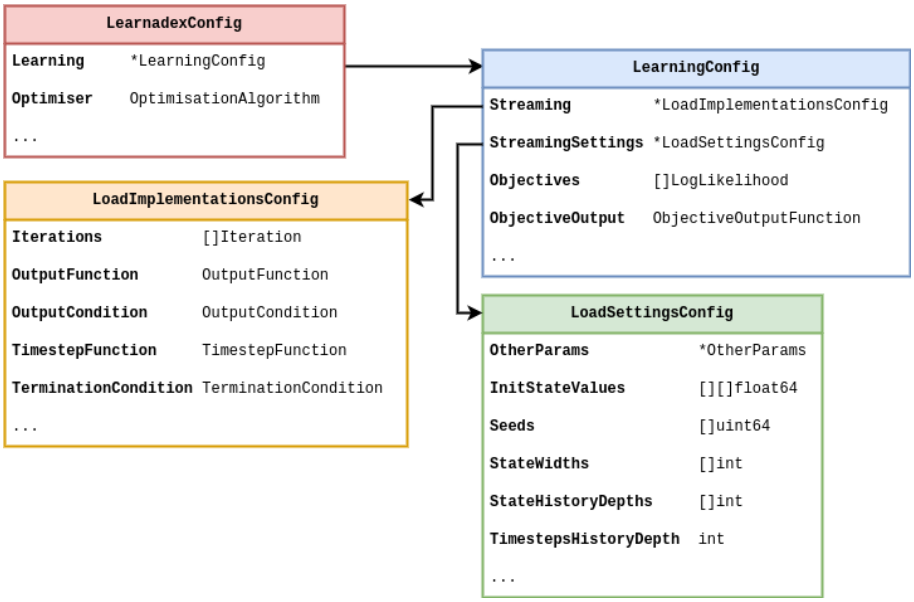

Figure 3.2: A relational summary of the core data types in the learnadex.

On the learning side; in order to define a specific objective for each data iterator to compute while the data streams through it, we have abstracted a 'log-likelihood' type. Similarly, each iterator also gets a data streamer configuration which defines where the data is streaming from — e.g., from a file on disk, from a local database instance or maybe via a network socket — and also some inherited abstractions from the stochadex which define the time stepping function and when the data stream ends. In Fig. 3.3 below, we provide a schematic of the method calls of (and within) each data iterator.

- introduce the β past discounting factor in this section and explain what it's for
- refactor the code so that it's always doing online learning under the hood — this can either be rolling refits in blocks on a refitting schedule with any optimisation algorithm of choice or

full online learning Adam optimisation

$$z_*(\mathrm{t}+1) = -\alpha[\mathrm{t}+1, \text{stats of gradient history like Adam}]\frac{\partial}{\partial z}\ln\mathcal{L}_{\mathrm{t}+1} + z_*(\mathrm{t})$$

- refactor the code and integrate the reweighting algorithm with Libtorch models for the conditional probabilities — describe how this is supported
- describe the method calls diagram in more detail — in particular, point out how it can replace the `Iteration.Iterate` method which is called when the `StateIterator` is asked for another iteration from the `PartitionCoordinator` of the stochadex
- then talk about the optimiser! starting with non-gradient-based: the two packages that are supported out of the box are gonum and eaopt (still need to do gago — see here: github.com/maxhalford/eaopt)
- also need to then support gradient-based algorithms (like vanilla SGD) by implementing Eq. (3.10) for the current basic implementations in the learnadex — shouldn't be too difficult!
- then talk about the output - talk about the possibilities for output and what the default setting to json logs is for
- could also be written to, e.g., a locally-hosted database server and the best-suited would be a NoSQL document database, e.g., MongoDB [30], but building something bespoke and simpler is more aligned with the use-case here and with the principles of this book
- describe the need for log exploration and visualisation and then introduce logsplorer - a REST API for querying the json logs (with basic filtering and selection capabilities but could be extended to more advanced options) and optionally also launches a visualisation React app written in Typescript
- note how this could be scaled to cloud services easily and remotely queried through the logsplorer API and visualised

As with the software we wrote for the stochadex, the learnadex main binary executable leverages templating to enable full configurability of all the implementations and settings of Fig. 3.2 through passing configs at runtime. Users can alternatively use the learnadex as a library for import, if they desire more control over the code execution.

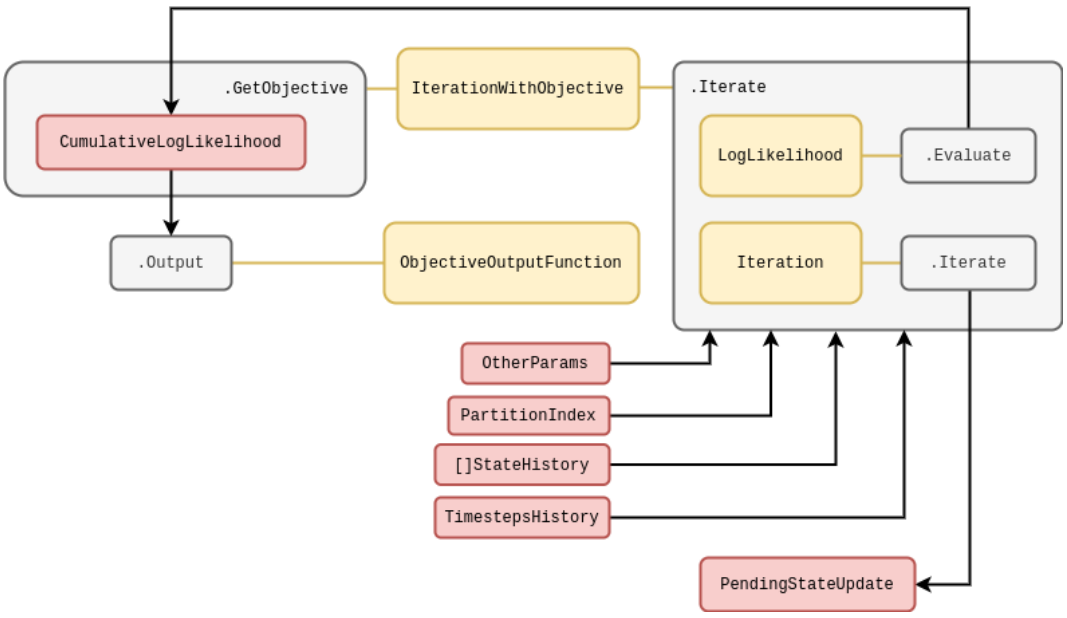

Figure 3.3: A schematic of an iteration with an objective function evaluation.

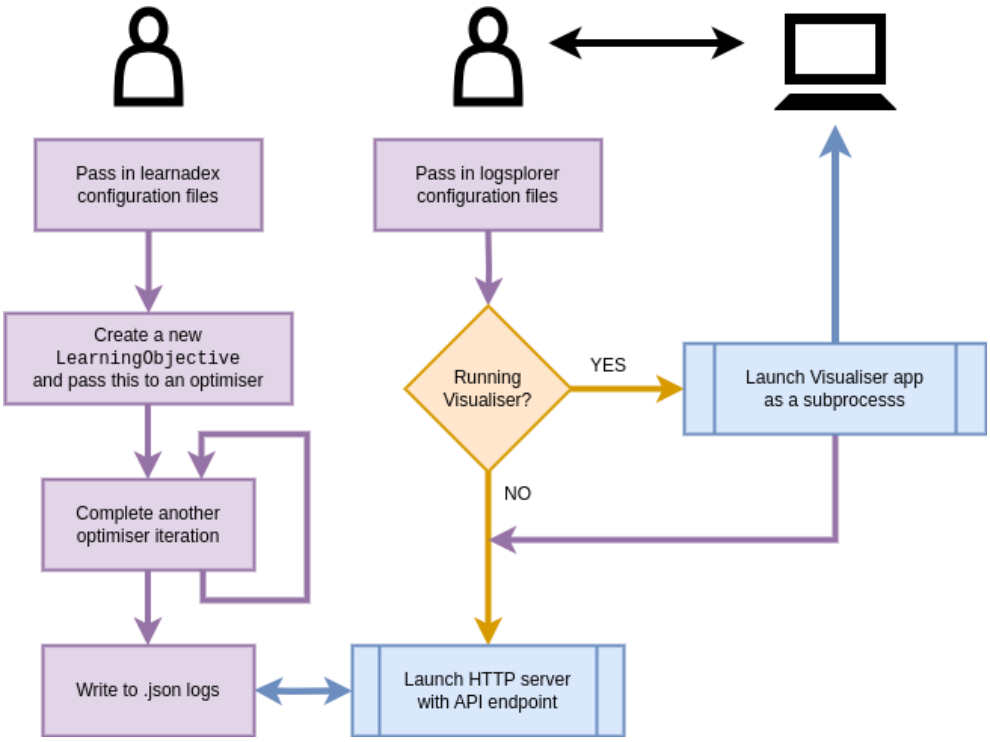

Figure 3.4: A diagram of the main learnadex and logsplorer executables.

4

Generalised simulation inference

Concept. To generalise the procedure of statistical inference for any simulation model using an algorithm which builds from techniques we developed in the previous chapter. When we say 'statistical inference' here; we specifically mean computing the maximum a posteriori (MAP) estimate for any arbitrary stochastic model which has been defined in the stochadex simulator. For the mathematically-inclined, this chapter will give a very brief exposition for Bayesian statistical inference metholodology — in particular, how it relates to the evaluation of MAP estimates. For the programmers, the software described in this chapter lives in the public Git repository: https://github.com/umbralcalc/learnadex.

4.1 Inference formalism

In Bayesian inference, one applies Bayes' rule to the problem of statistically inferring a model from some dataset. This typically involves the following formula for a posterior distribution

$$\mathcal{P}_{\mathsf{t}+1}(z|Y) \propto \mathcal{L}_{\mathsf{t}+1}(Y|z)\mathcal{P}(z)\,. \tag{4.1}$$

In the formula above, one relates the prior probability distribution over a parameter set $\mathcal{P}(z)$ and the likelihood $\mathcal{L}_{\mathsf{t}+1}(Y|z)$ of some data matrix Y up to timestep $\mathsf{t}+1$ given the parameters z of a model to the posterior probability distribution of parameters given the data $\mathcal{P}_{\mathsf{t}+1}(z|Y)$ up to some proportionality constant. All this may sound a bit technical in statistical language, so it can also be helpful to summarise what the formula above states verbally as follows: the initial (prior) state of knowledge about the parameters z we want to learn can be updated by some likelihood function of the data to give a new state of knowledge about the values for z (the 'posterior' probability).

From the point of view of statistical inference, if we seek to maximise $\mathcal{P}_{\mathsf{t}+1}(z|Y)$ — or its logarithm — in Eq. (4.1) with respect to z, we will obtain what is known as a maximum posteriori (MAP) estimate of the parameters. In fact, we have already encountered this metholodology in the previous chapter when discussing the algorithm which obtains the best fit parameters for the empirical probability reweighting. In this case; while it appears that we optimised the log-likelihood

directly as our objective function, one can easily show that this is also technically equivalent obtaining a MAP estimate where one chooses a specfic prior $\mathcal{P}(z) \propto 1$ (typically known as a 'flat prior').

How might we calulate the posterior in practice with some arbitrary stochastic process model that has been defined in the stochadex? In order to make the comparison to a real dataset, any stochadex model of interest will always need to be able to generate observations which can be directly compared to the data. To formalise this a little; a stochadex model could be represented as a map from z to a set of stochastic measurements $\mathsf{Y}_{\mathsf{t}+1}(z), \mathsf{Y}_{\mathsf{t}}(z), \dots$ that are directly comparable to the rows in the real data matrix Y. The values in Y may only represent a noisy or partial measurement of the latent states of the simulation X, so a more complete picture can be provided by the following probabilistic relation

$$P_{\mathsf{t}+1}(\mathsf{y}|z) = \int_{\omega_{\mathsf{t}+1}} \mathrm{d}^n x \, P_{\mathsf{t}+1}(\mathsf{y}|x) P_{\mathsf{t}+1}(x|z) \,, \tag{4.2}$$

where, in practical terms, the measurement probability $P_{\mathsf{t}+1}(\mathsf{y}|x)$ of $\mathsf{Y}_{\mathsf{t}+1} = \mathsf{y}$ given $X_{\mathsf{t}+1} = x$ can be represented by sampling from another stochastic process which takes the state of the stochadex simulation as input. Given that we have this capability to compare like-for-like between the data and the simulation; the next problem is to figure out how this comparison between two sequences of vectors can be done in a way which ensures the the statistics of the posterior are ultimately respected.

For an arbitrary simulation model which is defined by the stochadex, the likelihood in Eq. (4.1) is typically not describable as a simple function or distribution. While we could train the probability reweighting we derived in the previous chapter to match the simulation; to do this well would require having an exact formula for the conditional probability, and this is not always easy to derive in the general case. Instead, there is a class of Bayesian inference methods which we shall lean on to help us compute the posterior distribution (and hence the MAP), which are known as 'Likelihood-Free' methods [31, 32, 33, 34].

'Likelihood-Free' methods work by separating out the components of the posterior which relate to the closeness of rows in Y to the rows in Y from the components which relate the states X and parameters z of the simulation stochastically to Y. To achieve this separation, we can make use of chaining conditional probability like this

$$\mathcal{P}_{\mathsf{t}+1}(X, z|Y) = \int_{\Upsilon_{\mathsf{t}+1}} \mathrm{d}\mathsf{Y} \, \mathcal{P}_{\mathsf{t}+1}(\mathsf{Y}|Y) P_{\mathsf{t}+1}(X, z|\mathsf{Y}) \,, \tag{4.3}$$

where $\Upsilon_{\mathsf{t}+1}$ here corresponds to the domain of the simulated measurements matrix Y at time $\mathsf{t}+1$.

As we demonstrated in the previous chapter, it's possible for us to also optimise a probability distribution $\mathcal{P}_{\mathsf{t}'}(\mathsf{y}|Y) = P_{\mathsf{t}'}(\mathsf{y}; \mathcal{M}_{\mathsf{t}'}, \mathcal{C}_{\mathsf{t}'}, \dots)$ for each step in time to match the statistics of the measurements in Y as well as possible, given some statistics $\mathcal{M}_{\mathsf{t}'} = \mathcal{M}_{\mathsf{t}'}(Y)$ and $\mathcal{C}_{\mathsf{t}'} = \mathcal{C}_{\mathsf{t}'}(Y)$. Assuming the independence of samples (rows) in Y, this distribution can be used to construct the distribution over all of Y through the following product

$$\mathcal{P}_{\mathsf{t}+1}(\mathsf{Y}|Y) = \prod_{\mathsf{t}'=0}^{\mathsf{t}+1} P_{\mathsf{t}'}(\mathsf{y}; \mathcal{M}_{\mathsf{t}'}, \mathcal{C}_{\mathsf{t}'}, \dots) \,. \tag{4.4}$$

We do not necessarily need to obtain these statistics from the probability reweighting method, but could instead try to fit them via some other objective function. Either way, this represents a lossy

compression of the data we want to fit the simulation to, and so the best possible fit is desirable; regardless of overfitting. This choice to summarise the data with statistics means we are using what is known as a Bayesian Synthetic Likelihood (BSL) method [32, 33] instead of another class of methods which approximate an objective function directly using a proximity kernel — known as Approximate Bayesian Computation (ABC) methods [31].

Let's consider a few concrete examples of $P_{\mathsf{t}'}(\mathsf{y}; \mathcal{M}_{\mathsf{t}'}, \mathcal{C}_{\mathsf{t}'}, \dots)$. If the data measurements were well-described by a multivariate normal distribution, then

$$P_{\mathsf{t}'}(\mathsf{y}; \mathcal{M}_{\mathsf{t}'}, \mathcal{C}_{\mathsf{t}'}, \dots) = \mathsf{MultivariateNormalPDF}(\mathsf{y}; \mathcal{M}_{\mathsf{t}'}, \mathcal{C}_{\mathsf{t}'})\,, \tag{4.5}$$

Similarly, if the data measurements were instead better described by a Poisson distribution, we might disregard the need for a covariance matrix statistic $\mathcal{C}_{\mathsf{t}'}$ and instead use

$$P_{\mathsf{t}'}(\mathsf{y}; \mathcal{M}_{\mathsf{t}'}, \mathcal{C}_{\mathsf{t}'}, \dots) = \mathsf{PoissonPMF}(\mathsf{y}; \mathcal{M}_{\mathsf{t}'})\,. \tag{4.6}$$

The more statistically-inclined readers may notice that the probability mass function here would require the integrals in Eq. (4.3) to be replaced with summations over the relevant domains.

Eq. (4.3) demonstrates how one can construct a statistically meaningful way to compare the sequence of real data measurements $Y_{\mathsf{t}+1}, Y_{\mathsf{t}}, \dots$ to their modelled equivalents $\mathsf{Y}_{\mathsf{t}+1}(z), \mathsf{Y}_{\mathsf{t}}(z), \dots$. But we still haven't shown how to compute $P_{\mathsf{t}+1}(X, z|\mathsf{Y})$ for a given simulation, and this can be the most challenging part. To begin with, we can reapply Bayes' rule and the chaining of conditional probability to find

$$P_{\mathsf{t}+1}(x, z|\mathsf{Y}) \propto P_{\mathsf{t}+1}(\mathsf{y}|z) P_{\mathsf{t}}(z|\mathsf{Y}') = P_{\mathsf{t}+1}(\mathsf{y}|x) P_{\mathsf{t}+1}(x|z) P_{\mathsf{t}}(z|\mathsf{Y}')\,, \tag{4.7}$$

where here $P_{\mathsf{t}}(z|\mathsf{Y}')$ is the probability of $\mathsf{Y}_{\mathsf{t}} = \mathsf{Y}'$.

The relationship between $P_{\mathsf{t}+1}(X|z)$ and previous timesteps can be directly inferred from the probabilistic iteration formula that we introduced in the previous chapter. So we can map probabilities of $X_{0:\mathsf{t}+1} = X$ throughout time and learned information about the state of the system can be applied from previous values, given z. But is there a similar relationship we might consider for $P_{\mathsf{t}+1}(z|\mathsf{Y})$? Yes there is! The marginalisation

$$P_{\mathsf{t}+1}(z|\mathsf{Y}) \propto \left[\int_{\Omega_{\mathsf{t}+1}} \mathrm{d}^n x\, P_{\mathsf{t}+1}(\mathsf{y}|x) P_{\mathsf{t}+1}(x|z) \right] P_{\mathsf{t}}(z|\mathsf{Y}')\,, \tag{4.8}$$

shows how the z updates can occur in an iterative fashion. The reader may also recognize the factor above in brackets as Eq. (4.2). To complete the picture, one can combine the X and z updates into a joint distribution update which takes the following form

$$P_{\mathsf{t}+1}(X, z|\mathsf{Y}) \propto P_{\mathsf{t}+1}(\mathsf{y}|x) P_{(\mathsf{t}+1)\mathsf{t}}(x|X', z) P_{\mathsf{t}}(X', z|\mathsf{Y}')\,. \tag{4.9}$$

We can also marginalise this distribution over the past state history rows to get a distribution over the latest state row $X_{\mathsf{t}+1} = x$ like this

$$P_{\mathsf{t}+1}(x, z|\mathsf{Y}) = \int_{\Omega_{\mathsf{t}}} \mathrm{d}X' P_{\mathsf{t}+1}(X, z|\mathsf{Y}) \propto P_{\mathsf{t}+1}(\mathsf{y}|x) \int_{\Omega_{\mathsf{t}}} \mathrm{d}X' P_{(\mathsf{t}+1)\mathsf{t}}(x|X', z) P_{\mathsf{t}}(X', z|\mathsf{Y}')\,. \tag{4.10}$$

In the next section, we're going to discuss how to translate all of this probabilistic language into some MAP inference algorithms. Before we do this, however, it will be instructive (particularly for

'online' learning algorithms) to consider what happens if the model changes over time and z needs to change in order to better represent the real data. In such situations, we propose to apply the same formula as Eq. (4.10) but instead replace the distribution over (X', z) on the right hand side with its 'past discounted' version[1]

$$\int_{\Omega_{\mathsf{t}}} \mathrm{d}X' P_{\mathsf{t}}(X', z|\mathsf{Y}') \;\longrightarrow\; \frac{1}{\mathsf{t}} \sum_{\mathsf{t}'=0}^{\mathsf{t}} \int_{\omega_{\mathsf{t}'}} \mathrm{d}^n x' \beta^{\mathsf{t}-\mathsf{t}'} P_{\mathsf{t}'}(x', z|\mathsf{Y}')\,, \tag{4.11}$$

where $0 < \beta < 1$ and we recall the notation which considers distributions over the individual rows x' within the matrix X' in this new version. This time-dependent discount factor could be used to reduce the dependence of the update on data which is much further in the past, and hence will ultimately lead to a more responsive algorithm. This responsiveness would have to be balanced with the tradeoffs associated with discounting potentially valuable data that may offer greater long-term stability. Readers who are familiar with reinforcement learning may be starting to feel in familiar territory here — they will have to wait for the latter parts of the book to see more on discounting though!

4.2 Online learning the MAP

Eq. (4.9) tells us how to probabilistically translate the current state of knowledge about (x, z) forward through time in response to the arrival of new data. We also know how to connect the simulated measurements to the real data because Eq. (4.3) essentially gives us an objective function to maximise for each step in time. This is all great in theory; but in practice, this optimisation problem typically has several layers of difficulty to it. Since the model has been defined by its stochastically generated samples of measurements $\mathsf{Y}_{\mathsf{t}+1}(z), \mathsf{Y}_{\mathsf{t}}(z), \ldots$, the objective function will manifestly be stochastic too. Another layer of difficulty is that gradients of the objective function are not immediately computable and so navigation around the optimisation domain could be difficult, especially in high-dimensional problems. Lastly, given that the simulation model in the stochadex needs to be running multiple times for each timestep, we need a way of mitigating computational expense.

So how should we proceed? To solve this problem in the general case, Eqs. (4.3) and (4.9) tell us we need to synthesize the following components into a single algorithm:

1. A process $P_{(\mathsf{t}+1)\mathsf{t}}(x|X', z)$ which iterates the state matrix of the simulation X forward in time.
2. A process $P_{\mathsf{t}+1}(\mathsf{y}|x)$ which generates a simulated measurement from the simulated state x.
3. A probability distribution $P_{\mathsf{t}'}(\mathsf{y}; \mathcal{M}_{\mathsf{t}'}, \mathcal{C}_{\mathsf{t}'}, \ldots)$ which represents the posterior distribution of the simulated measurement vector y given an optimised compression of the real data into summary statistics.
4. Some way of representing samples from the distribution $P_{\mathsf{t}+1}(X, z|\mathsf{Y})$ so that their distribution can be updated and will coverge towards the posterior over (X, z).

[1] In the continuous-time version, this past-discounting factor can depend on the stepsize such that we replace

$$\beta^{\mathsf{t}-\mathsf{t}'} \longrightarrow \frac{1}{\beta[\delta t(\mathsf{t})]} \prod_{\mathsf{t}''=\mathsf{t}'}^{\mathsf{t}} \beta[\delta t(\mathsf{t}'')]\,.$$

Before writing this up, should read this paper on efficient amortized inference using neural networks with `BayesFlow` here in particular: [35]. But also, should cite other works to make amortized inference more efficient by using neural networks to learn convenient functions of the Bayes factor in Evidence networks [36].

- amortized online inference of the posterior update over just z can be achieved by running lots of simulations and solving the inverse problem with the `y` outputs i.e., neural network modelling of the update in Eq. (4.8)

The algorithm is specifically: 1. if this is a refit step, sample new values for (X, z) for all members of the ensemble from the current (X, z) distribution points and run the iterations for all of these ensemble members from the back of the window all the way up to the current point in time (hence the full matrix X is sampled) 2. take all of the ensemble members a step forward in time 3. approximate the mode by computing the average values of z within the q-th percentile of the sampled probability mass (where q is set by the user and is ideally $< 68\%$) — this idea comes from nested sampling 4. stream in the data for the next point in time and go to 1.

As such an algorithm converges, we can recompute (and hence iteratively improve) the MAP estimate with respect to each iteration of the posterior.

Readers with some machine learning experience may be familiar with the classic exploration vs exploitation tradeoffs. It's clear that these tradeoffs will manifest in our case here when trying to strike a balance between iterating the posterior distribution and optimizing the current posterior with respect to (X, z) to compute the MAP.

Readers of the previous section may also have recognized that Eq. (4.9) contains the same conditional probability $P_{(\mathsf{t}+1)\mathsf{t}}(x|X', z)$ as the reweighting algorithm. This structure enables us to reuse all of the exposition we provided for the probabilistic reweighting and highlights how the reweighting itself can be used in the algorithm to optimise the posterior.

If we now synthesize both of these observations together, we can see how a stochastic variant of the well-known Expectation-Maximisation Algorithm [37, 38, 21] naturally emerges.

5

Optimising interactions with any system

Concept. To design and build software which enables the optimisation of automated control objectives over stochastic phenomena of any kind. The theory in this chapter will overlap significantly with that of Reinforcement Learning (RL), however, in contrast to more standard RL approaches, we shall be relying on all of the work from previous parts of this book to help agents characterise, measure and learn from their environment. In particular, the online learning of stochastic simulation state and parameters will be crucial to this model-based approach. The software which implements our generalised control optimisation algorithm will be implemented as an extension to the learnadex. For the mathematically-inclined, this chapter will cover how we formalise model-based automated control optimisation within the frameworks that we have already introduced in this book. For the programmers, the public Git repository for the code described in this chapter can be found here: https://github.com/umbralcalc/stochadex and here: https://github.com/umbralcalc/learnadex.

5.1 Formalising general interactions

Let's start by considering how we might adapt the mathematical formalism we have been using so far to be able to take actions which can manipulate the state at each timestep. Using the mathematical notation that we inherited from the stochadex, we may extend the formula for updating the state history matrix $X_{0:\mathtt{t}} \to X_{0:\mathtt{t}+1}$ to include a new layer of possible interactions which is facilitated by a new vector-valued 'take action' function $G_{\mathtt{t}}$. In doing so we shall be defining the domain of an acting entity in the stochastic process environment — which we shall hereafter refer to as simply the 'agent'.

During a timestep over which actions are performed by the agent, the stochadex state update formula can be extended to include interactions by composition with the original state update function like so

$$X^i_{\mathtt{t}+1} = G^i_{\mathtt{t}+1}[F_{\mathtt{t}+1}(X_{0:\mathtt{t}}, z, \mathtt{t}), A_{\mathtt{t}+1}] = \mathcal{F}^i_{\mathtt{t}+1}(X_{0:\mathtt{t}}, z, A_{\mathtt{t}+1}, \mathtt{t})\,, \tag{5.1}$$

where we have also introduced the concept of the 'actions' performed A_{t+1} on the system; some vector of parameters which define what actions are taken at timestep $t+1$. The code for the new iteration formula would look something like Fig. 5.1.

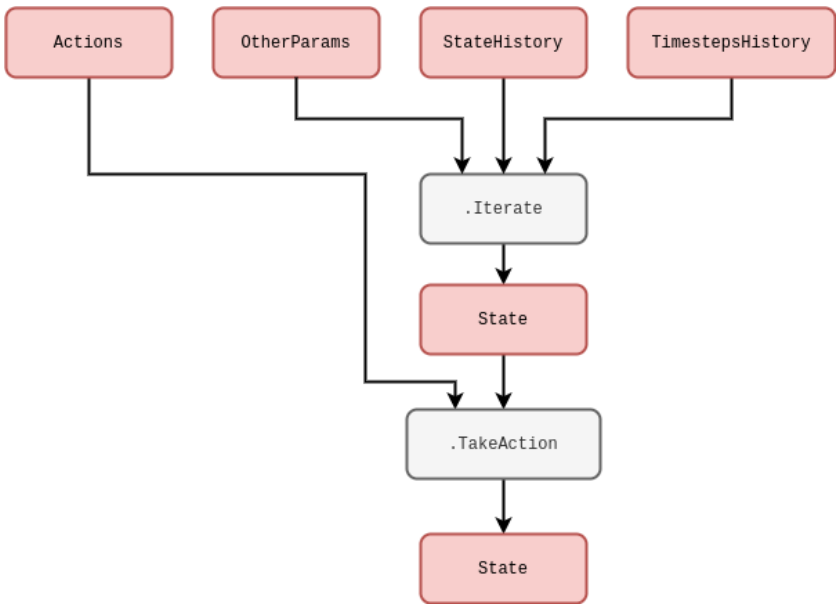

Figure 5.1: Code schematic of Eq. (5.1).

So far, Eq. (5.1) on its own will allow the agent to take actions that are scheduled up front through some fixed process or perhaps through user interaction via a game interface. So what's next? In order to start creating algorithms which will act on the system state for us, we need to develop a formalism which 'closes the loop' by feeding information back from the stochastic process to the agent's decision-making algorithm.

If we use $A_{0:t+1}$ a referring to the matrix of historically-taken actions which up to time $t+1$, we can build up a more generalised, non-Markovian picture of automated interactions with the system which matches the notation we are already using for $X_{0:t+1}$. Let us now define a Non-Markovian Decision Process (NMDP) as a probabilistic model which draws an actions matrix $A_{0:t+1} = A$ from a 'policy' distribution $\Pi_{(t+1)t}(A|X,\theta)$ given $X_{0:t} = X$ and a new vector of parameters which fully specify the automated interations. Using the probabilistic notation from the previous part of the book, the joint probability that $X_{0:t+1} = X$ and $A_{0:t+1} = A$ at time $t+1$ is

$$P_{t+1}(X, A|z, \theta) = P_t(X'|z, \theta)\, \Pi_{(t+1)t}(A|X', \theta) P_{(t+1)t}(x|X', z, A)\,, \tag{5.2}$$

where we recall that $P_{(t+1)t}(x|X', z, A)$ is the conditional probability of $X_{t+1} = x$ given $X_{0:t} = X'$ and z that we have encountered before, but it now requires $A_{0:t+1} = A$ as another given input. We have illustrated Eq. (5.1) and how it relates to the policy distribution of Eq. (5.2) with a new graph representation in Fig. 5.2.

For additional clarity, let's take a moment to think about what $\Pi_{(t+1)t}(A|X,\theta)$ represents and how generally descriptive it can be. If an agent is acting under and entirely deterministic policy,

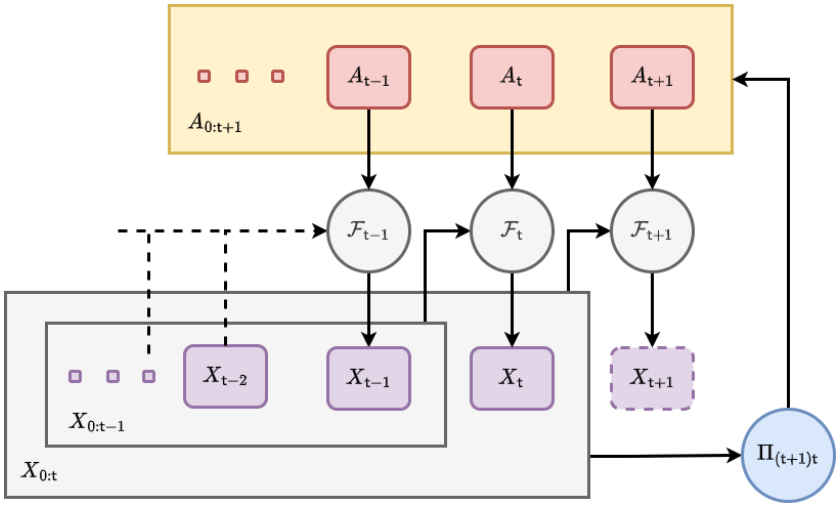

Figure 5.2: Graph representation of Eq. (5.1) with the policy distribution of Eq. (5.2).

then the policy distribution may be simplified to a direct function mapping which is parameterised by θ. At the other extreme, the distribution may also describe a fully stochastic policy where actions are drawn randomly in time. If we combine this consideration of noise with the observation that policies described by a distribution $\Pi_{(t+1)t}(A|X,\theta)$ permit a memory of past actions and states, it's easy to see that this structure can be used in a wide variety of different use cases.

By marginalising over Eq. (5.2) we find an updated probabilistic iteration formula for the stochastic process state which now takes the influence of agent actions into account

$$P_{t+1}(X|z,\theta) = \int_{\Xi_{t+1}} \mathrm{d}A\, P_t(X'|z,\theta)\, \Pi_{(t+1)t}(A|X',\theta) P_{(t+1)t}(x|X',z,A)\,. \tag{5.3}$$

This relationship will be very useful in the last part of this book when we begin to look at optimising control algorithms.

What are the main categories of action which are possible in the rows of A? Since the NMDP described by $\Pi_{(t+1)t}(A|X',\theta)$ is just another form of stochastic process, the main categories of action will fall into the same as those we covered in defining the stochadex formalism. The first, and perhaps most obvious, category would probably where the actions are defined in a continuous space and are continuously applied on every timestep. Some examples of these 'continuously-acting' decision processes include controlling the temperature of chemical reactions [39] (such as those in a brewery), spacecraft control [40] and guidance systems, as well as the driving of autonomous vehicles [41]. Within a kind of subset of the continuously-acting category; we can also find the event-based acting decision processes (where actions are not necessarily taken every timestep), e.g. controlling traffic through signal timings [42], managing disease spread through treatment intervals [43] and automated trading on stock markets [44].

Many of the examples we have given above have continuous action spaces, but we might also consider classes of decision processes where actions are defined discretely. Examples of these include the famous multi-armed bandit problem [45] (like choosing between website layouts for E-

commerce [46]), managing a sports team through player substitutions, sensor measurement scheduling [47] and the sequential design prioritisation of large-scale scientific experiments [48].

5.2 States, actions and attributing rewards

In the previous parts of this book we laid out the concept for a generalised framework to simulate and learn stochastic phenomena continually as data is received. Given that we have also introduced a framework for the automated control of these phenomena, we have all the ingredients we need to create optimal decision-making algorithms. The key question to answer then, is: *optimal with respect to what objective?*

The objective of an automated control algorithm could take many forms depending on the specific context. Since there is no loss in generality in doing so, it seems natural to follow the naming convention used by Markov Decision Processes (MDP) [49, 1] by referring to the objective outcome of an action at a particular point in time as having a 'reward' value r. Since the relationship between reward, actions and states may be stochastic, it makes sense to relate the reward outcome r given a state history X and action history A at timestep $\mathsf{t}+1$ through the probability distribution $P_{\mathsf{t}+1}(r|X,A)$. Hence, generally, this reward signal is non-Markovian — as is the case in many real-world problems [50].

We can use the reward probability distribution to derive a joint distribution over both state history X' and reward r at timestep $\mathsf{t}+1$ like so

$$P_{(\mathsf{t}+1)\mathsf{t}}(r,x'|X,z,\theta) = P_{\mathsf{t}+1}(r|X',A)\Pi_{(\mathsf{t}+1)\mathsf{t}}(A|X,\theta)P_{(\mathsf{t}+1)\mathsf{t}}(x'|X,z,A)\,. \tag{5.4}$$

In this expression, let's recall that we are using the policy distribution $\Pi_{(\mathsf{t}+1)\mathsf{t}}(A|X,\theta)$ for agent interactions and the fundamental state update conditional probability for the underlying stochastic process $P_{(\mathsf{t}+1)\mathsf{t}}(x'|X,z,A)$.

Note that in most use cases, the state of real-world phenomena cannot be measured perfectly. So to enable any agent trained on simulated phenomena to potentially act in the real world, we will need to include a measurement process as part of the information retrieval step. This is the part where we can leverage our work in a previous chapter which develops an online learning system for stochastic process models. But we're jumping ahead with this thinking and will return to this point later on.

Using Eq. (5.4), we can now define a 'state value function' V_{t} at timestep t which is the expected γ-discounted future reward given the current state history X and the other parameters like this[1]

$$\begin{aligned} V_{\mathsf{t}}(X,z,\theta) &= \mathrm{E}_{\mathsf{t}}(\mathsf{Discounted\ Return}|X,z,\theta) \\ &= \sum_{\mathsf{t}'=\mathsf{t}}^{\infty}\int_{\omega_{\mathsf{t}'+1}} \mathrm{d}^n x' \int_{\rho_{\mathsf{t}'+1}} \mathrm{d}r\, r\, \gamma^{\mathsf{t}'-\mathsf{t}} \prod_{\mathsf{t}''=\mathsf{t}}^{\mathsf{t}'} P_{(\mathsf{t}''+1)\mathsf{t}''}(r,x'|X,z,\theta)\,, \end{aligned} \tag{5.5}$$

[1] The discount factor in continuous time could also be explicitly dependent on the stepsize such that we would replace the discount factor in Eq. (5.5) with

$$\gamma^{\mathsf{t}'-\mathsf{t}} \longrightarrow \frac{1}{\gamma[\delta t(\mathsf{t}+1)]}\prod_{\mathsf{t}''=\mathsf{t}}^{\mathsf{t}'}\gamma[\delta t(\mathsf{t}''+1)]\,.$$

where $0 < \gamma < 1$. The idea behind this discount factor γ is to decrease the contribution of rewards to the optimisation objective (often called the 'expected discounted return' in RL) more and more as the prediction increases into the future. Note also that the state value function is inherently recursively defined, such that

$$V_{\mathsf{t}}(X, z, \theta) = \int_{\omega_{\mathsf{t}+1}} \mathrm{d}^n x \int_{\rho_{\mathsf{t}+1}} \mathrm{d}r \, P_{(\mathsf{t}+1)\mathsf{t}}(r, x'|X, z, \theta) \Big\{ r + \gamma V_{\mathsf{t}+1}(X', z, \theta) \Big\}, \tag{5.6}$$

and the optimal θ can hence be derived from

$$\theta^*_{\mathsf{t}}(X, z) = \underset{\theta}{\operatorname{argmax}} \big[V_{\mathsf{t}}(X, z, \theta) \big] . \tag{5.7}$$

By deriving the optimal policy in terms of the parameters $\theta^*_{\mathsf{t}}(X, z)$, the optimal state value function and policy distribution can therefore be derived from

$$V^*_{\mathsf{t}}(X, z) = V_{\mathsf{t}}[X, z, \theta^*_{\mathsf{t}}(X, z)] \tag{5.8}$$

$$\Pi^*_{(\mathsf{t}+1)\mathsf{t}}(A|X, z) = \Pi_{(\mathsf{t}+1)\mathsf{t}}[A|X, \theta^*_{\mathsf{t}}(X, z)] . \tag{5.9}$$

Note that the type of decision process optimisation which we have introduced above differs from standard RL methodology. In the more conventional 'model-free' RL approaches, the state-action value function

$$Q_{\mathsf{t}}(X, A, z) = \mathrm{E}_{\mathsf{t}}(\mathsf{Discounted\ Return}|X, A, z) , \tag{5.10}$$

would be used to evaluate the optimal policy instead of the state value function $V_{\mathsf{t}}(X, z, \theta)$ that we are using above. We are able to use the latter here because the simulation model gives us explicit knowledge of the $P_{(\mathsf{t}+1)\mathsf{t}}(x'|X, z, A)$ distribution which is utilised by Eq. (5.4). When this model is not known, the state-action value function $Q_{\mathsf{t}}(X, A, z)$ must be learned explicitly through sample estimation from the measured state and experienced outcomes of actions taken by the agent.

When an agent takes an action to measure the state of the system (or when it is given measurements without needing to take action) there will typically be some uncertainty in how the history of measured real-world data Y maps to the latent states of the system X and its parameters z at time $\mathsf{t}+1$. It is natural, then, to represent this uncertainty with a posterior probability distribution $\mathcal{P}_{\mathsf{t}+1}(X, z|Y)$ as we did in the previous chapters of this book.

Follow-up this bit with the model-based approach that we're going to take in this book.

- Introduce broad concept of dynamic programming — partitioning a optimal global control into smaller optimal control segments/iterations.
- Talk about the utility of the model-based online learning approach in the case of partially observed systems [51].
- Look into the overlaps with this approach and Thompson sampling for exploration — discuss here.
- Looking at a stochastic policy iteration algorithm here combined with Monte Carlo rollouts.
- The value learning can be facilitated in software using a predictive model which is able to roll forecast rewards forward in time in a Monte Carlo fashion up to a window from a certain point given an input prior distribution of policies.
- This input prior distribution of policies can itself be optimised by maximising expected discounted utility in a Bayesian design framework. Draw parallels.

Part 2

6

Controlling parasitic infections

Concept. The idea here is to limit the spread of some abstract spatial parasitic infections through the correct time-dependent resource allocation.

6.1 Adapting the probabilistic formalism

Let's by returning to the probabilistic formalism that we introduced earlier and noting that the covariance matrix estimate with elements $C^{ij}_{\mathsf{t}+1}(z)$ represents a matrix that could get very large, depending on the problem. For example; if we encoded the state of a 2-dimensional spatial field of values into the elements X^i_{t}, the number of elements in the covariance matrix $C^{ij}_{\mathsf{t}+1}(z)$ would scale as $4N^2$ — where N here is the number of spatial points we wanted to encode.

One solution to this scaling problem is to exploit the fact that, in many spatial processes, the proximity of points can strongly determine how correlated they are. Hence, for pairwise distances further than some threshold, the covariance matrix elements should tend towards 0. If we were to place points along the diagonal of $C^{ij}_{\mathsf{t}+1}(z)$ in order of how close they are to each other, this threshold would then be represented as a *banded matrix*. We have illustrated such a matrix in Fig. 6.1 in which the 'bandwidth' is defined as the number of diagonals one needs to traverse from the main diagonal before encountering a diagonal of 0s.

- At some point it might be sensible to move into the Fourier domain here — at least for derivations and calculations. Probably more intuitive for the reader to keep it mostly in real space though if possible.
- The extra detail that's also needed here is to consider how we encode a 2-dimensional spatial process into our state vector, and how the elements of the resulting state vector might be correlated to one another depending on their spatial proximity. If we start with a Markovian Gaussian random field, we can derive the Matérn kernel over these spatial coordinates in order to correlate the state vectors in such a way.

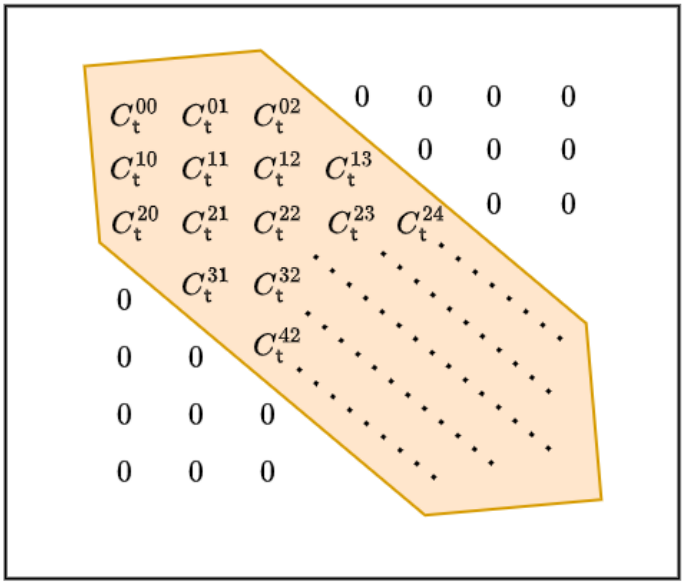

Figure 6.1: An illustration of a banded covariance matrix with a bandwidth of 2.

- Also look into the Radial Basis Function (RBF) and higher-order derived kernels based on DALI expansion [52] in order to try and capture non-Gaussianity.
- Spatial malaria models here: [53, 54]
- Spatial helminth-like infections too would be a good template...

7

Algo-trading on financial markets

Concept. The idea here is to use the Q-Hawkes processes and the Bouchaud work to come up with some interesting simulations of financial markets.

- Fundamental simulation should be that of a proper limit order book microsim [55]
- Algo trades using online learning of the market dynamics through Q-Hawkes processes

8

Sustainable angling for fish

Concept. The idea here is

8.1 A large-scale Lotka-Volterra model

Inspired by the empirical dynamical modeling approach to sockeye salmon in Ref. [56], but also desiring a generative model which has some link to the classic causal models promoted by mathematical ecology; the goal here is to create and calibrate a stochastic model which predicts the fish counts, weights, lengths and ages for each species in each area based on the past system states. To do this, we will combine some well-known models from mathematical ecology with supervised learning.

The one-step master equation for the proposed stochastic simulation is given implictly by

$$\frac{\mathrm{d}}{\mathrm{d}t}P(\ldots,n_i,\ldots,t) = \sum_{\forall i} \mathcal{T}_i^{+}(\ldots,n_i-1,\ldots,\mathsf{f},t)P(\ldots,n_i-1,\ldots,t) \tag{8.1}$$

$$+ \sum_{\forall i} \mathcal{T}_i^{-}(\ldots,n_i+1,\ldots,\mathsf{f},t)P(\ldots,n_i+1,\ldots,t) \tag{8.2}$$

$$- \sum_{\forall i} \Big[\mathcal{T}_i^{+}(\ldots,n_i,\ldots,\mathsf{f},t) + \mathcal{T}_i^{-}(\ldots,n_i,\ldots,\mathsf{f},t)\Big]P(\ldots,n_i,\ldots,t)\,, \tag{8.3}$$

where the time t is defined in units of years and $\mathcal{T}_i^{+}$ and $\mathcal{T}_i^{-}$ are the transition coefficients for the i-th species, which depend not only on the counts for all species $n_1, n_2, \ldots$, but also (in principle) on a larger feature space f generated by the available data up to time t.

The famous Lotka-Volterra system, with some modficiations for fishing and a larger set of species, would suggest transition coefficients of the form

$$\mathcal{T}_i^+(\dots, n_i, \dots, \mathsf{f}, t) = \mathcal{T}_i^+(\dots, n_i, \dots) = \Lambda_i(n_i) + n_i\alpha_i \sum_{\forall i' \text{ prey}} n_{i'} \tag{8.4}$$

$$\mathcal{T}_i^-(\dots, n_i, \dots, \mathsf{f}, t) = \mathcal{T}_i^-(\dots, n_i, \dots) = n_i\mu_i + n_i\gamma_i + n_i\beta_i \sum_{\forall i' \text{ pred}} n_{i'} \,, \tag{8.5}$$

where: $\Lambda_i(n_i) = \tilde{\Lambda}_i n_i e^{-\lambda_i(n_i-1)}$ is the density-dependent birth rate; μ_i is the species death rate; α_i is the increase in the baseline birth rate per fish caused by the increase in prey population; β_i is the rate per fish of predation of the species; and γ_i accounts for the rate of recreational fishing per fish of the species. To approach the present data-driven simulation problem, we're going to generalise this model by training $\mathcal{T}_i^+(\dots, n_i, \dots, \mathsf{f}, t)$ and $\mathcal{T}_i^-(\dots, n_i, \dots, \mathsf{f}, t)$ directly from the data and generated features.

Look into the likelihood from, e.g., an electrofishing survey such as in Ref. [57]...

$$\mathsf{Likelihood} = \sum_{\mathsf{data}} \mathrm{NB}\big[\mathsf{data}; w_{i,\mathsf{survey}}\langle n_i(t_{\mathsf{data}})\rangle, k_{i,\mathsf{survey}}\big] \,, \tag{8.6}$$

9

Managing a rugby match

Concept. Building a toy model simulation of a rugby match whose outcome can be manipulated through correctly-timed player substitutions and game management decisions. The state manipulation framework we have built around the stochadex can meet these requirements, and a dashboard can be created for user interaction. All this combines together to make a simple dashboard game, which we call: 'trywizard'. For the mathematically-inclined, this chapter will motivate the construction of a specific modeling framework for rugby match simulation. For the programmers, the public Git repository for the code described in this chapter can be found here: https://github.com/umbralcalc/trywizard.

9.1 Designing the event simulation engine

Since the basic state manipulation framework and simulation engine will run using the stochadex, the mathematical novelties in this project are all in the design of the rugby match model itself. And, as ever, we're not especially keen on spending a lot of time doing detailed data analysis to come up with the most realistic values for the parameters that are dreamed up here. Even though this would also be interesting.[1]

Let's begin by specifying an appropriate state space to live in when simulating a rugby match. It is important at this level that events are defined in quite broadly applicable terms, as it will define the state space available to our stochastic sampler and hence the simulated game will never be allowed to exist outside of it. It seems reasonable to characterise a rugby union match by the following set of states: Penalty, Free Kick (the punitive states); Penalty Goal, Drop Goal, Try (the scoring states); Kickoff, Kick Phase, Run Phase, Knock-on, Scrum, Lineout, Maul and Ruck (the general play states). Using this set of states, in Fig. 9.1 we have summarised our approach to match state transitions into a single event graph. In order to capture the fully detailed range of events that are possible in a real-world match, we've needed to be a little imaginative in how we define the

[1] One could do this data analysis, for instance, by scraping player-level performance data from one of the excellent websites that collect live commentary data such as rugbypass.com or espn.co.uk/rugby.

kinds of state transitions which occur.[2] For example, kickoffs, 22m dropouts and goal line dropouts are all modelled here as a Kickoff state but with different initial ball locations on the pitch (we'll get to how ball location changes later on).

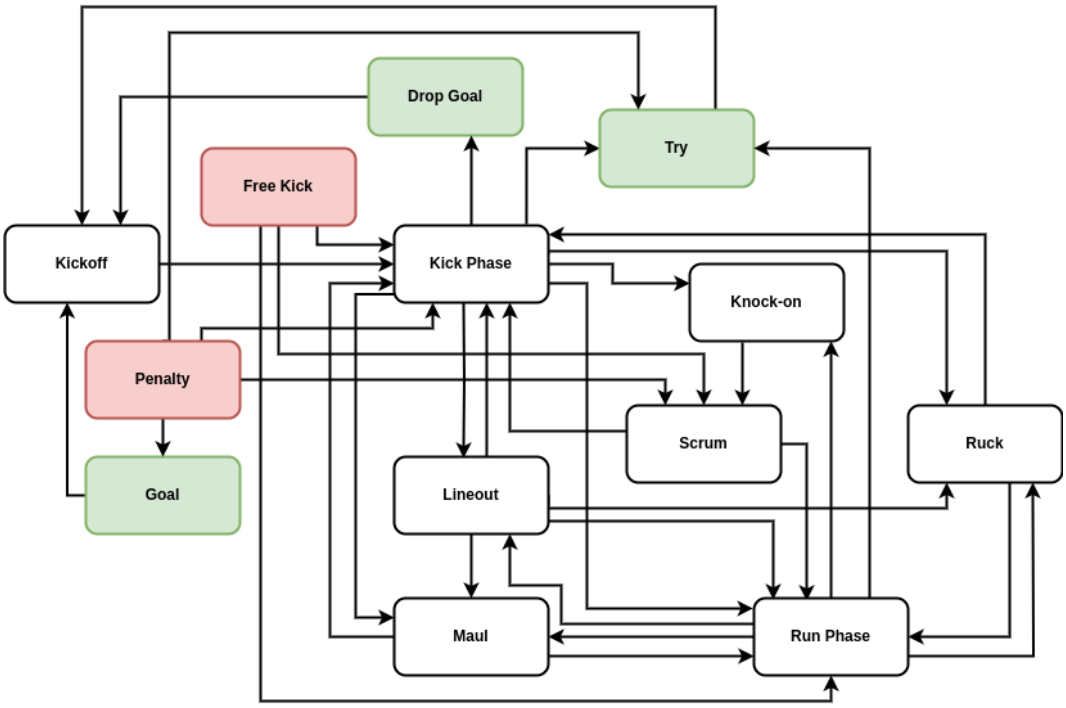

Figure 9.1: Simplified event graph of a rugby union match.

In addition to occupying some state in the event graph, the state of a rugby match must also include a binary 'possession' element which encodes which team has the ball at any moment. We should also include the 2-dimensional pitch location of the ball as an element of the match state in order to get a better sense of how likely some state transitions are, e.g., when playing on the edge of the pitch near the touchline it's clearly more likely that a Run Phase is going to result in a Lineout than if the state is currently in the centre of the pitch. To add even more detail; in the next section we will introduce states for each player.

Since a rugby match exists in continuous time, it is natural to choose a continuous-time event-based simulation model for our game engine. As we have discussed in previous chapters already, this means we will be characterising transition probabilities of the event graph in Fig. 9.1 by ratios of event rates in time. Recalling our notation in previous chapters, if we consider the current state vector of the match X_t, we can start by assigning each transition $\mathsf{a} \to \mathsf{b}$ on the event graph an associated expected rate of occurance $\lambda_{\mathsf{a}\to\mathsf{b}}$ which is defined in units of continuous time, e.g., seconds. In addition to the transitions displayed on the graph, we can add a 'possession change transition'; where the possession of the ball in play moves to the opposing team. This transition

[2] It's also fair to say that our simplified model here represents just a subset of states that a real rugby match could exist in.

may occur while the match is also in most of the white-coloured states on the graph apart from Knock-on (which determines a possession change immediately through a Scrum) or a Kickoff (which directly proceeds a Kick Phase from which the ball may change possession). Let's assign possession changes with a state and timestep-dependent expected rate of occurance $\lambda_{\text{pos}}(X_{\mathsf{t}}, \mathsf{t})$.

Based on our dicussion above, an appropriate encoding for the overall game state at timestep index t could be a state vector X_{t} whose elements are

$$X^0_{\mathsf{t}} = \begin{cases} 0 & \text{Match State} = \textsf{Penalty} \\ 1 & \text{Match State} = \textsf{Free Kick} \\ \dots & \end{cases} \tag{9.1}$$

$$X^1_{\mathsf{t}} = \begin{cases} 0 & \text{Possession} = \textsf{Home Team} \\ 1 & \text{Possession} = \textsf{Away Team} \end{cases} . \tag{9.2}$$

But how does this overall game state connect to the event rates? The probabilistic answer is quite straightforward. If the probability of the match state being $X^0_{\mathsf{t}} = \mathsf{a}$ at timestep t is written as $P^0_{\mathsf{t}}(\mathsf{a})$, then the probability of $X^0_{\mathsf{t}+1} = \mathsf{b}$ in the following timestep is

$$P^0_{\mathsf{t}+1}(\mathsf{b}) = \frac{\frac{1}{\tau} P^0_{\mathsf{t}}(\mathsf{b}) + \sum_{\forall \mathsf{a} \neq \mathsf{b}} \lambda_{\mathsf{a}\to\mathsf{b}} \mathcal{T}_{\mathsf{a}\to\mathsf{b}}(X_{\mathsf{t}}, \mathsf{t}) P^0_{\mathsf{t}}(\mathsf{a})}{\left[\frac{1}{\tau} + \sum_{\forall \mathsf{a} \neq \mathsf{b}} \lambda_{\mathsf{a}\to\mathsf{b}} \mathcal{T}_{\mathsf{a}\to\mathsf{b}}(X_{\mathsf{t}}, \mathsf{t})\right]} , \tag{9.3}$$

where $\forall \mathsf{a} \neq \mathsf{b}$ in the summation indicates that all the available transitions from a to b, where $\mathsf{a} \neq \mathsf{b}$, should be summed over and $\mathcal{T}_{\mathsf{a}\to\mathsf{b}}(X_{\mathsf{t}}, \mathsf{t})$ is a time and state-dependent transition probability that is determined by the playing tactics of each team as well as the general likelihoods of gameplay which are expected from a real rugby match. Note that in the expression above, we have also defined τ as a timescale short enough such that no transition is likely to occur during that interval. An equivalent to Eq. (9.3) should also apply to the possession change transition rate, i.e., the probability that the Home Team has possession $P^1_{\mathsf{t}}(\mathsf{H})$ at time t evolves according to

$$P^1_{\mathsf{t}+1}(\mathsf{H}) = \frac{\frac{1}{\tau} P^1_{\mathsf{t}}(\mathsf{H}) + \lambda_{\text{pos}}(X_{\mathsf{t}}, \mathsf{t})[1 - P^1_{\mathsf{t}}(\mathsf{H})]}{\left[\frac{1}{\tau} + \lambda_{\text{pos}}(X_{\mathsf{t}}, \mathsf{t})\right]} . \tag{9.4}$$

Before we move on to other details, it's quite important to recognise that because our process is defined in continuous time, the possession change rate may well vary continuously (this will be especially true when we talk about, e.g., player fatigue). Hence, Eq. (9.3) is only an *approximation* of the true underlying dynamics that we are trying to simulate — and this approximation will only be accurate if τ is small. The reader may recall that we discussed this same issue from the point of view of simulating time-inhomogeneous Poisson processes with the rejection method when we were building the stochadex in an earlier chapter.

While these match state transitions and possession changes are taking place, we also need to come up with a model for how the ball location L_{t} changes during the course of a game, and as a function of the current game state. Note that, because the ball location is a part of the overall game state, it will be included as information contained within some elements of X_{t} as well. To make this explicit, we can simply set $X^2_{\mathsf{t}} = L^{\text{lon}}_{\mathsf{t}}$ and $X^3_{\mathsf{t}} = L^{\text{lat}}_{\mathsf{t}}$ — where $L^{\text{lon}}_{\mathsf{t}}$ denotes the longitudinal component (lengthwise along the pitch) and $L^{\text{lat}}_{\mathsf{t}}$ denotes the lateral component (widthwise across the pitch). If we associate every state on the event graph with a single change in spatial location of the ball on the pitch, we then need to construct a process which makes 'jumps' in 2-dimensional

space each time a state transition occurs. To keep things simple and intuitive, we will say that movements of the ball are only allowed to occur during either a Run Phase or a Kick Phase. In most cases this restriction makes sense given the real-world game patterns, but perhaps the only clear exception is the Penalty $\longrightarrow$ Goal transition; which is easier to think of as a kind of 'Kick Phase transition' anyway.

In the case of a Run Phase, let's choose the longitudinal component of the ball location $L_{\mathsf{t}}^{\text{lon}}$ to be updated by the difference between samples drawn from two exponential distributions (one associated to each team). Hence, the probability density $P_{\mathsf{t}+1}(\ell)$ of $L_{\mathsf{t}+1}^{\text{lon}} - L_{\mathsf{t}}^{\text{lon}} = \ell$, evolves according to

$$P_{\mathsf{t}+1}(\ell) = \int_0^\infty \mathrm{d}\ell' \, \mathsf{ExponentialPDF}(\ell + \ell'; a_{\text{run}}) \mathsf{ExponentialPDF}(\ell'; d_{\text{run}}) \,, \tag{9.5}$$

where a_{run} and d_{run} are the exponential distribution scale parameters for an attacking and defending player, respectively, and we have chosen positive values of ℓ to be aligned with the forward direction for the attacking team. We shall elaborate on where a_{run} and d_{run} come from when we discuss associating events for player abilities in due course. If we now consider lateral component of the ball location $L_{\mathsf{t}}^{\text{lat}}$ during a Run Phase; it makes sense that this wouldn't be affected much by either team within the scope of detail in this first version of our model. Hence, the probability density $P_{\mathsf{t}+1}(w)$ of $L_{\mathsf{t}+1}^{\text{lat}} - L_{\mathsf{t}}^{\text{lat}} = w$ can just be updated like so

$$P_{\mathsf{t}+1}(w) = \mathsf{NormalPDF}(w; 0, \sigma_{\text{run}}^2) \,, \tag{9.6}$$

where σ_{run} is the typical jump in lateral motion (the standard deviation parameter of the normal distribution).

Turning our attention to the Kick Phase; the longitudinal and lateral components are only realistically controlled by the attacking team — specifically, by the player who is currently the kicker. Referring back to the state transitions which precede a Kick Phase in Fig. 9.1, we note that there are several types of kick which can dictate what the mechanics of the process should look like. To keep things simple, we can cluster these types of event into the following categories

1. Kicks at the goalposts for points:
   - Penalty $\longrightarrow$ Goal
   - Kick Phase $\longrightarrow$ Drop Goal
2. Kicks in the general field of play where the ball does not leave the field:
   - Kick Phase $\longrightarrow$ Try
   - Kick Phase $\longrightarrow$ Run Phase
   - Kick Phase $\longrightarrow$ Knock-on
   - Kick Phase $\longrightarrow$ Ruck
   - Kick Phase $\longrightarrow$ Maul
3. Kicks to the touchline where the ball leaves the field:
   - Kick Phase $\longrightarrow$ Lineout

To model case 1. above, the simplest option would be to associate the attempt at goal with a goal success probability for the kicker p_{goal} which, in the simple first version of our model, will not depend on the location on the pitch from which the kick is taken. We will, however, restrict kickers to only be allowed an attempt at goal if they are within their opposing team's half — this is not strictly a rule in rugby, but it simplifies the automation of the decision logic quite nicely for now.

In case 2 above, we can think of two main tactical options that a team might be employing. The first of these is kicking a further longitudinal distance in the field of play in order to gain territory, but lose possession, and the second is to kick to regain possession but with a shorter longitudinal distance. When kicking to gain territory, we'll just assign a uniform probability to the lateral position update of the ball location for simplicity and for the probability density $P_{\mathsf{t}+1}(\ell)$ of $L^{\text{lon}}_{\mathsf{t}+1} - L^{\text{lon}}_{\mathsf{t}} = \ell$, we'll use

$$P_{\mathsf{t}+1}(\ell) = \mathsf{ExponentialPDF}(\ell; a_{\text{kick}})\,, \tag{9.7}$$

where a_{kick} is the kicking player's longitudinal scale parameter. When kicking to regain possession, we will use another exponential distribution with another constant scale parameter ℓ_{typ} for the typical distance gained by this type of kick (unassociated to either team's abilities) and also assign a 'regain possession' probability p_{reg} which is associated to the abilities of the players chasing the kick (on the kicker's team).

Lastly, to model case 3. above, the event has determined that the ball will leave the field of play and so the remaining unknowns that need to be determined are: which side of the field this occurred (we'll just choose the side closest to touch when the ball was last in play), the longitudinal distance of the kick and whether or not the ball bounced before leaving the field. For simplicity, let's determine the last of these through another kind of kick accuracy probability p_{kick} associated to the kicker. This just leaves the longitudinal distance that the kick achieved along the touchline; in this case we'll just assign the probability density $P_{\mathsf{t}+1}(\ell)$ of $L^{\text{lon}}_{\mathsf{t}+1} - L^{\text{lon}}_{\mathsf{t}} = \ell$ to that of Eq. (9.7).

Generally, the proceeding Lineout will be taken with the opposition team (to the kicker) in possession at the point where the ball left the field. However, there are two notable exceptions to this rule. The first is if the ball does not bounce before leaving the field of play and the player is outside of their team's 22m line — the Lineout then must occur where the ball was last kicked from with the opposition team in possession. The second exception is if the ball bounces into touch when the player who kicked it was inside their team's half of the pitch — the Lineout then occurs where the ball left the field but is taken with the kicker's team in possession.

Before moving on to player states and abilities it's important to note that, in addition to the other state variables we have discussed above, the score of the match obviously needs to be recorded in the overall match state X_{t} as well. To be explicit, we'll say that $X^4_{\mathsf{t}} = \mathsf{Home\ Team\ Score}$ and $X^5_{\mathsf{t}} = \mathsf{Away\ Team\ Score}$, recalling that kicks at goal from a Penalty Goal or Drop Goal are worth 3 points, a Try is worth 5 points and if this is proceeded by a conversion (another kick at goal taken laterally inline with where the Try was scored across the pitch) then this is worth an additional 2 points.

9.2 Associating events to player states and abilities

In the last section we introduced a continuous-time event-based simulation model for a rugby union match. In this section we are going to add more detail into this model by inventing how to associate specific player states and abilities to the event rates of the simulation. Before continuing, we want

to reiterate that this model is entirely made up and, while we hope it illustrates some interesting mathematical modeling ideas in the context of rugby, there's no particular reason why a statistical inference with a reliable dataset should prefer our model to others which may exist.

In Fig. 9.2 we have begun by separating playing positions on the rugby field into their usual descriptions and then associating each player type with a short list of simplified attributes. Note that our model associates a player with an 'possession attacking' and 'possession defending' ability which corresponds to each of their possession attributes that are indicated by the diagram. For example, a Front Row Forward will have 10 possession abilities associated to them: 2 for each of their Scrum, Lineout, Ruck, Maul and Run attributes.

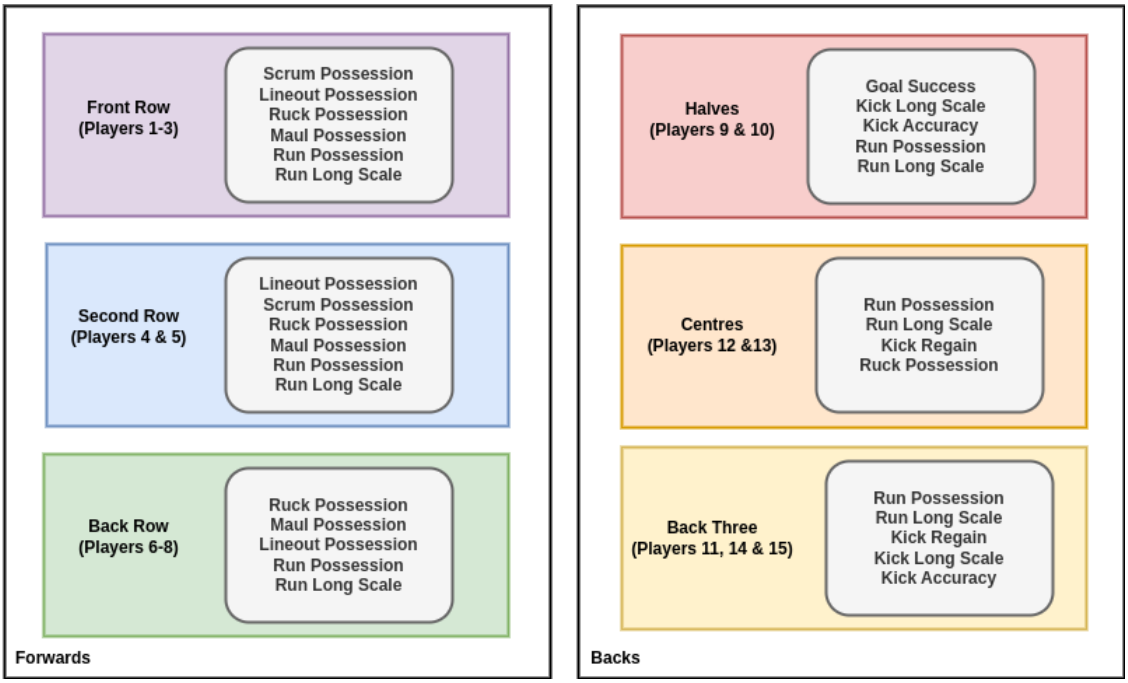

Figure 9.2: Associated playing abilities for each position type.

Let's now say that z contains all of these parameters for all of the players on both sides, including those on the bench. With these parameters, the knowledge of which team is in possession from X_{t}^1, and the identifiers of those players who are actively playing on the field, it should be simple to create a vector-valued function $a_{\mathrm{pos}}(X_{\mathrm{t}})$ which returns all of the possession attacking attributes that are associated to match state X_{t}^0 and an analogous one $d_{\mathrm{pos}}(X_{\mathrm{t}})$ for the possession defending attributes. The dependencies of these functions on the ball possession state X_{t}^1 comes from the fact that when, e.g., the Home Team has possession of the ball it will be their possession attacking attributes that are returned by $a_{\mathrm{pos}}(X_{\mathrm{t}})$ and the Away Team's possession defending attributes that are returned by $d_{\mathrm{pos}}(X_{\mathrm{t}})$.

In order to model the effect of player fatigue over the course of a match, we can add some vectors of player fatigue values f into the collection of parameters that are contained within z. These new parameters can then be used to define a formula for the decline of each attribute over the course of

a match. Let's redefine these declining values as

$$a^i_{\text{pos}}(X_{\mathsf{t}}, \mathsf{t}) = a^i_{\text{pos}}(X_{\mathsf{t}}) e^{-f^i[t(\mathsf{t}) - t^i_{\text{start}}]} \tag{9.8}$$

$$d^i_{\text{pos}}(X_{\mathsf{t}}, \mathsf{t}) = d^i_{\text{pos}}(X_{\mathsf{t}}) e^{-f^i[t(\mathsf{t}) - t^i_{\text{start}}]} \,. \tag{9.9}$$

So how does each player affect the events of a match? In our model, we would argue that players should be able to directly influence the possession change rate $\lambda_{\text{pos}}(X_{\mathsf{t}}, \mathsf{t})$ through a balance of attacking and defensive attributes in the following relation

$$\lambda_{\text{pos}}(X_{\mathsf{t}}, \mathsf{t}) = \frac{\lambda^*_{\text{pos}} \sum_{\forall i} d^i_{\text{pos}}(X_{\mathsf{t}}, \mathsf{t})}{\sum_{\forall i} a^i_{\text{pos}}(X_{\mathsf{t}}, \mathsf{t}) + \sum_{\forall i} d^i_{\text{pos}}(X_{\mathsf{t}}, \mathsf{t})} \,, \tag{9.10}$$

where λ^*_{pos} is the maximum rate that is physically possible and the $\forall i$ in the summations indicates summing over all attacking or defending player attributes of the vector in each instance. In addition to this possession change influence, players who have Run Phase and Kick Phase longitudinal scale attributes may affect the gain in distance that each state translates to on the pitch.

Let's first describe how we intend the Run Phase to work. Every time the match state transitions into a Run Phase, an individual player on the attacking side is chosen at random (uniformally across the team[3]) to be the the nominal 'attacker'. At the same time, an individual player on the defending side is chosen at random (again, uniformally across the team) to be the the nominal 'defender'. Once these players have been chosen (and hence the a_{run} and d_{run} parameters have been determined), the longitudinal motion update we described in Eq. (9.5) can then be applied. Note that the a_{run} and d_{run} parameters should also receive a fatigue decrement depending on the time that each player has remained on the pitch, much like the exponential factors we applied in Eqs. (9.8) and (9.9).

Finally, we turn our attention to the mechanics of a Kick Phase. For tactical kicking in the middle of play, this operates in a similar way as a Run Phase; one of the players on the attacking side with kicking parameters is chosen at random (uniformally) and their a_{kick} or p_{kick} attributes are used in the relevant equations. If a kick stays within the field of play for the attacking team to attempt to regain possession, an additional 'chaser' player on the attacking team who possesses a kick regain p_{reg} probability is randomly chosen. In the specific case of Goal kicking from a Penalty, the designated place kicker on each side uses their p_{goal} attribute to determine the success/failure of the kick at the posts.

Before moving on to actions, it will be necessary to extend the stored information in X_{t} to include more of the dynamic information about the match — some of which may be manipulable in-play by the manager. This additional state information includes the identifiers of each player on the field, the times at which they began playing t_{start} and the current values of the rates denoted by $\mathcal{T}_{\mathsf{a}\to\mathsf{b}}(X_{\mathsf{t}}, \mathsf{t})$.

9.3 Deciding on managerial actions

So how does managing a rugby match map to taking actions with our formalism? We have to start by figuring out what sorts of managerial actions exist in the real world and then specify how we should map these to changes in state.

Our model structure would suggest that the only way in which a manager can influence the state of a match is through modifying the parameters which are used by the $a_{\text{pos}}(X_{\mathsf{t}})$, $d_{\text{pos}}(X_{\mathsf{t}})$ or

[3]This uniform sampling could be refined later to associate sampling probabilities with game state and player roles.

$\mathcal{T}_{\mathsf{a}\to\mathsf{b}}(X_\mathsf{t}, \mathsf{t})$ functions. In the case of the possession attacking and defending attributes $a_{\text{pos}}(X_\mathsf{t})$ and $d_{\text{pos}}(X_\mathsf{t})$, an action that the manager can perform would be to modify which players are actively playing through substitutions. In order to map substitutions/initial squad selections to the row vector of actions $A_{\mathsf{t}+1}$, we can define the first set of 15 indices $A^{0:14}_{\mathsf{t}+1}$ as the player identifiers chosen to be on the pitch for either the Home Team or the Away Team, depending on which side the manager is in charge of. If X_t contains all of these identifiers and the possible action space is restricted by positions that each player is allowed to play (as well as whether they are currently playing or not), then, when the manager wishes to make a substitution, they need only change $A^{0:14}_{\mathsf{t}+1}$ which then gets mapped directly to the active player identifiers stored in $X_{\mathsf{t}+1}$ from this timestep onwards. The times t_{start} would also get updated accordingly.

But what about $\mathcal{T}_{\mathsf{a}\to\mathsf{b}}(X_\mathsf{t}, \mathsf{t})$? What kinds of managerial actions can change the state transition probabilities? Given that these transition probabilities mostly arise from the tactics of each team, if the tactics of either team were changed throughout the match due to managerial decisions, these actions could be mapped to $\mathcal{T}_{\mathsf{a}\to\mathsf{b}}(X_\mathsf{t}, \mathsf{t})$. In order for these actions to have a clear influence on the game, however, we need to specify how team tactics get translated into transition probabilities. To keep things as simple as possible, we're going to specify only two tactical 'axes' on which a manager has to decide a position during each moment of the match.

The first, and perhaps more obvious, tactical decision axis to dynamically manipulate is the ratio between Kick Phase and Run Phase that the team chooses when it has possession of the ball, depending on what part of the pitch they are playing in. The second axis maps how aggressively a team pursues scoring tries over any other points (even when they may be on offer from a Penalty Goal). This latter axis also only really matters for the team in possession of the ball, so we aren't planning to map out any defensive tactics in our model for now. Since both of these actions can be mapped to a single axis each, these ratios q (each defined between $1 \geq q \geq 0$) can populate the last two indices of the actions vector: $A^{15}_{\mathsf{t}+1}$ and $A^{16}_{\mathsf{t}+1}$. When either of them has been changed, this can be mapped to $X_{\mathsf{t}+1}$ using the action function $\mathcal{F}_{\mathsf{t}+1}(X_{0:\mathsf{t}}, z, A_{\mathsf{t}+1}, \mathsf{t})$ and then $X_{\mathsf{t}+1}$, will have changed ratios between transition probabilities $\mathcal{T}_{\mathsf{a}\to\mathsf{b}}(X_{\mathsf{t}+1}, \mathsf{t}+1)$ when the attacking team is making these in-play decisions in the proceeding timestep.

Before moving on to the next section, there's a quick point to make about how managerial actions can affect ball locations on the pitch. In addition to the changes that we have discussed above, let's recall that the parameters a_{run}, d_{run}, a_{kick}, p_{kick}, p_{reg} and p_{goal} can all be indirectly determined by the manager to some extent through player selection/substitutions. So, while tactical managerial actions can change the ratios of state transitions themselves (at least while in possession of the ball), the players which are chosen in the first place (or by substitution) can have quite a significant influence on the outcome of a match.

9.4 Writing the game itself

- Show which stochadex methods were called and how they were used to simulate the game.
- Give a summary of how the dashboard backend works (diagram would help) and how this connects up to the streamlit frontend via protobuf messages.

10

Optimising relief chain logistics

Concept. The idea here is

- Humanitarian aid logistics in response to flooding, fire or other natural disasters
- Routing of transportation
- Where to focus searches
- Transportation size distribution
- Supply chain logistics of resources and allocation of budget
- Example paper here with stochastic network models [58]

Bibliography

[1] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[2] "The Go Programming Language." https://go.dev/.

[3] "The C++ Programming Language." https://isocpp.org/.

[4] "The Python Programming Language." https://www.python.org/.

[5] "The TypeScript Programming Language." https://www.typescriptlang.org/.

[6] "The Worlds Of Observation Book GitHub Repository." https://github.com/umbralcalc/worlds-of-observation.

[7] "Open Source Initiative: MIT License." https://opensource.org/licenses/MIT.

[8] N. G. Van Kampen, *Stochastic processes in physics and chemistry*, vol. 1. Elsevier, 1992.

[9] H. Risken, *Fokker-planck equation*, in *The Fokker-Planck Equation*, pp. 63–95. Springer, 1996.

[10] L. Rogers and D. Williams, *Diffusions, Markov Processes and Martingales 2: Ito Calculus*, vol. 1, pp. xiv+480. Cambridge University Press, 04, 2000. 10.1017/CBO9781107590120.

[11] L. Decreusefond et al., *Stochastic analysis of the fractional brownian motion*, *Potential analysis* **10** (1999) 177–214.

[12] D. T. Gillespie, *Exact stochastic simulation of coupled chemical reactions*, *The journal of physical chemistry* **81** (1977) 2340–2361.

[13] J. Neyman and E. L. Scott, *Statistical approach to problems of cosmology*, *Journal of the Royal Statistical Society: Series B (Methodological)* **20** (1958) 1–29.

[14] A. G. Hawkes, *Spectra of some self-exciting and mutually exciting point processes*, *Biometrika* **58** (1971) 83–90.

[15] "SimPy: a process-based discrete-event simulation framework." https://gitlab.com/team-simpy/simpy/.

[16] "StoSpa: A C++ package for running stochastic simulations to generate sample paths for reaction-diffusion master equation." https://github.com/BartoszBartmanski/StoSpa.

[17] "FLAME GPU: A GPU accelerated agent-based simulation library for domain independent complex systems simulation." https://github.com/FLAMEGPU/FLAMEGPU2/.

[18] "The React Library." https://react.dev/.

[19] 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.

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

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

[22] I. Kobyzev, S. J. Prince and M. A. Brubaker, *Normalizing flows: An introduction and review of current methods*, *IEEE transactions on pattern analysis and machine intelligence* **43** (2020) 3964–3979.

[23] L. Pinheiro Cinelli, M. Araújo Marins, E. A. Barros da Silva and S. Lima Netto, *Variational autoencoder*, in *Variational Methods for Machine Learning with Applications to Deep Networks*, pp. 111–149. Springer, 2021.

[24] J. A. Nelder and R. Mead, *A simplex method for function minimization*, *The computer journal* **7** (1965) 308–313.

[25] J. Kennedy and R. Eberhart, *Particle swarm optimization*, in *Proceedings of ICNN'95-international conference on neural networks*, vol. 4, pp. 1942–1948, IEEE, 1995.

[26] Y. Shi and R. Eberhart, *A modified particle swarm optimizer*, in *1998 IEEE international conference on evolutionary computation proceedings. IEEE world congress on computational intelligence (Cat. No. 98TH8360)*, pp. 69–73, IEEE, 1998.

[27] H. Robbins and S. Monro, *A stochastic approximation method*, *The annals of mathematical statistics* (1951) 400–407.

[28] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, *arXiv preprint arXiv:1412.6980* (2014) .

[29] E. Hazan et al., *Introduction to online convex optimization*, *Foundations and Trends® in Optimization* **2** (2016) 157–325.

[30] "The MongoDB Webpage." https://www.mongodb.com/.

[31] S. A. Sisson, Y. Fan and M. Beaumont, *Handbook of approximate Bayesian computation*. CRC Press, 2018.

[32] L. F. Price, C. C. Drovandi, A. Lee and D. J. Nott, *Bayesian synthetic likelihood*, *Journal of Computational and Graphical Statistics* **27** (2018) 1–11.

[33] S. N. Wood, *Statistical inference for noisy nonlinear ecological dynamic systems*, *Nature* **466** (2010) 1102–1104.

[34] C. Drovandi and D. T. Frazier, *A comparison of likelihood-free methods with and without summary statistics*, *Statistics and Computing* **32** (2022) 42.

[35] S. T. Radev, U. K. Mertens, A. Voss, L. Ardizzone and U. Köthe, *Bayesflow: Learning complex stochastic models with invertible neural networks*, *IEEE transactions on neural networks and learning systems* **33** (2020) 1452–1466.

[36] N. Jeffrey and B. D. Wandelt, *Evidence networks: simple losses for fast, amortized, neural bayesian model comparison*, *arXiv preprint arXiv:2305.11241* (2023) .

[37] H. O. Hartley, *Maximum likelihood estimation from incomplete data*, *Biometrics* **14** (1958) 174–194.

[38] A. P. Dempster, N. M. Laird and D. B. Rubin, *Maximum likelihood from incomplete data via the em algorithm*, *Journal of the royal statistical society: series B (methodological)* **39** (1977) 1–22.

[39] C. Beeler, S. G. Subramanian, K. Sprague, N. Chatti, C. Bellinger, M. Shahen et al., *Chemgymrl: An interactive framework for reinforcement learning for digital chemistry*, *arXiv preprint arXiv:2305.14177* (2023) .

[40] M. Tipaldi, R. Iervolino and P. R. Massenio, *Reinforcement learning in spacecraft control applications: Advances, prospects, and challenges*, *Annual Reviews in Control* (2022) .

[41] B. R. Kiran, I. Sobh, V. Talpaert, P. Mannion, A. A. Al Sallab, S. Yogamani et al., *Deep reinforcement learning for autonomous driving: A survey*, *IEEE Transactions on Intelligent Transportation Systems* **23** (2021) 4909–4926.

[42] D. Garg, M. Chli and G. Vogiatzis, *Deep reinforcement learning for autonomous traffic light control*, in *2018 3rd IEEE international conference on intelligent transportation engineering (ICITE)*, pp. 214–218, IEEE, 2018.

[43] A. Q. Ohi, M. Mridha, M. M. Monowar and M. A. Hamid, *Exploring optimal control of epidemic spread using reinforcement learning*, *Scientific reports* **10** (2020) 22106.

[44] T. L. Meng and M. Khushi, *Reinforcement learning in financial markets*, *Data* **4** (2019) 110.

[45] J. Gittins, K. Glazebrook and R. Weber, *Multi-armed bandit allocation indices*. John Wiley & Sons, 2011.

[46] Y. Liu and L. Li, *A map of bandits for e-commerce*, *arXiv preprint arXiv:2107.00680* (2021) .

[47] A. S. Leong, A. Ramaswamy, D. E. Quevedo, H. Karl and L. Shi, *Deep reinforcement learning for wireless sensor scheduling in cyber–physical systems*, *Automatica* **113** (2020) 108759.

[48] T. Blau, E. V. Bonilla, I. Chades and A. Dezfouli, *Optimizing sequential experimental design with deep reinforcement learning*, in *International Conference on Machine Learning*, pp. 2107–2128, PMLR, 2022.

[49] D. P. Bertsekas et al., *Dynamic programming and optimal control 3rd edition, volume ii*, *Belmont, MA: Athena Scientific* **1** (2011) .

[50] M. Gaon and R. Brafman, *Reinforcement learning with non-markovian rewards*, in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, pp. 3980–3987, 2020.

[51] K. J. Åström, *Optimal control of markov processes with incomplete state information i*, *Journal of mathematical analysis and applications* **10** (1965) 174–205.

[52] E. Sellentin, M. Quartin and L. Amendola, *Breaking the spell of gaussianity: forecasting with higher order fisher matrices*, *Monthly Notices of the Royal Astronomical Society* **441** (2014) 1831–1840.

[53] T. Smith, N. Maire, A. Ross, M. Penny, N. Chitnis, A. Schapira et al., *Towards a comprehensive simulation model of malaria epidemiology and control*, *Parasitology* **135** (2008) 1507–1516.

[54] J. Đorđević, *A stochastic model for malaria and its behavior under insecticide-treated nets*, *Studies in Applied Mathematics* **149** (2022) 631–656.

[55] J.-P. Bouchaud, J. Bonart, J. Donier and M. Gould, *Trades, quotes and prices: financial markets under the microscope*. Cambridge University Press, 2018.

[56] H. Ye, R. J. Beamish, S. M. Glaser, S. C. Grant, C.-h. Hsieh, L. J. Richards et al., *Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling*, *Proceedings of the National Academy of Sciences* **112** (2015) E1569–E1576.

[57] "Electrofishing to assess a river's health." https://environmentagency.blog.gov.uk/2015/10/29/electrofishing-to-assess-a-rivers-health/.

[58] D. Alem, A. Clark and A. Moreno, *Stochastic network models for logistics planning in disaster relief*, *European Journal of Operational Research* **255** (2016) 187–206.