IWOMB 2021 Prereading: Stochastics

Alastair Jamieson-Lane and Alejandra D. Herrera Reyes

March 2021

1 Introduction

You've signed up for the "Introduction to numerical methods for stochastic models" workshop at the IWOMB conference. Awesome! We're looking forward to having you.

Everyone coming into the workshop will have different levels of expertise. This document is here to make sure everyone is up to speed with the basics of what you need to know coming in. If you're unfamiliar, then the information here should paint a general picture of what you'll be dealing with. If you've seen this all before, the next couple pages should act as a brief reminder.

If you are unfamiliar with the information listed under "Basic probability", then it is probably important to read up. For all the later sections, the information provided here should be enough to get you up to speed for the workshop, links are provided for anyone who wishes to read deeper. Feel free to contact us (alejandra.herrera@nottingham.ac.uk , alastair_JL@xtra.co.nz) if you need further references.

2 Probability Basics

This workshop assumes that you are familiar with the basics of probability theory. In particular, you should be familiar with:

- **Continuous** and **Discrete** **Random variables**. In this workshop, random variables have names like X and Y.
- Probability functions: **Probability Mass Function** and probability density function; **Cumulative Distribution Function** (c.d.f)
- Numeric random variables have an **Expected Value**. This is denoted $\mathbb{E}(X)$ or μ.
- Numeric random variables have an **Variance**. This is denoted $Var(X)$ or σ^2.
- **Independence of random variables**.

In theory, you should be comfortable with all the concepts so far. If these concepts *aren't* familiar, then get reading! You can find an open textbook covering this stuff **HERE**.

3 The Central limit Theorem

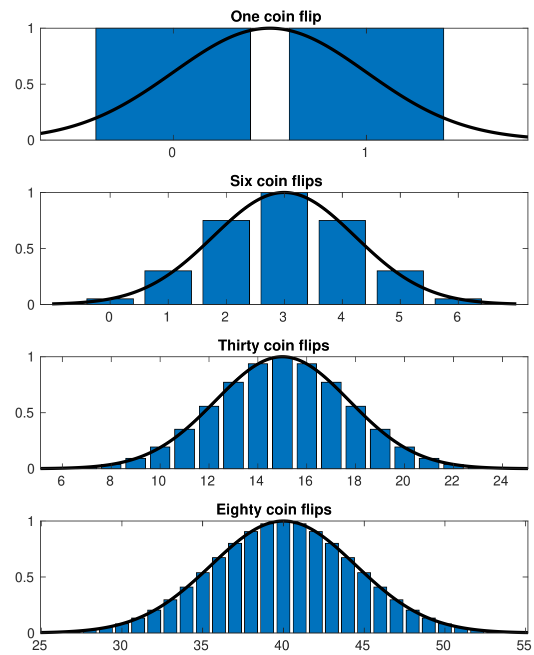

Figure 1: *Suppose we toss some number of fair coins, and count the number of heads. For a very large number of coins, our distribution starts looking more and more like a normal distribution. This happens not only for coin flips, but for any collection of independent identical random variables with finite variance and mean.*

Suppose you have a random variable X that is found by taking the sum of a large number of smaller random variables $X = \sum Y_i$. The exact processes governing all the individual Y_i might be complex and unknowable, but we are mostly interested in the total effect X.

For example, imagine a bucket full of dice, that you might pour across the floor and add up. Each dice is an individual Y_i, while their sum is X. In an epidemic, the infection and recovery of individual people add up to tell us the story of the overall epidemic. In your body, an individual strand of DNA will bounce about according to its collisions with hundreds and thousands of water molecules. Each collision is complex and unknowable (Y_i), but the sum can be observed (X).

The Central Limit Theorem states that whenever we add up a large pile of random numbers, the total value of those numbers will tend towards a **normal distribution** (bell curve):

$$f_X(x) \to f_N(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right].$$

Where μ is the expected value of our final distribution, and is found by adding up all the expected values of Y_i, and σ^2 is the variance of X, and is found by adding up the variance of all the Y_i. See Figure 1 for a graphic example.

Perhaps the most important point to note here is that this limit holds *regardless* of the underling shapes of the distributions Y_i. So long as Y_i are independent of one another, and have bounded finite mean and variance, it is possible to add them all up and get something approaching a normal distribution. We encourage you to check the details of the theorem (chapter 9 of **this textbook**) but they are not necessary for the workshop.

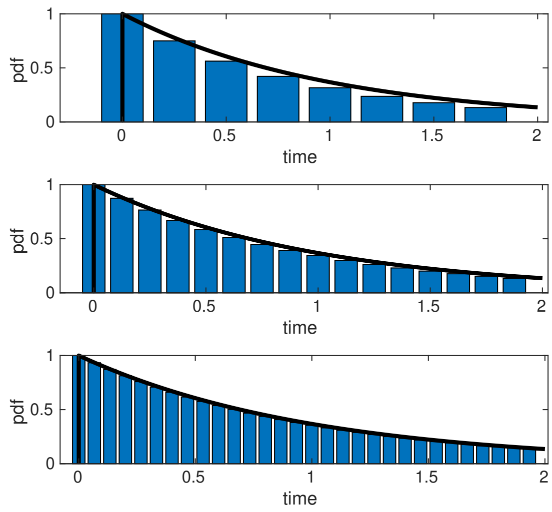

Figure 2: *The exponential distribution can be approximated by faster and faster dicerolls... with a smaller and smaller chance of 'stopping' with each roll (notice how the 'steps' get progressively smaller),*

4 Exponential Distributions

The second key distribution you will need to be familiar with is the **exponential distribution**.

The exponential distribution is a continuous probability distribution, and is most often used to represent the time until an event occurs.

Its density function is

$$f_T(t) = \lambda exp(-\lambda t), \quad \forall t > 0,$$

and its cumulative distribution function

$$F_T(t) = 1 - exp(-\lambda t), \quad \forall t > 0.$$

The expected value of the exponential distribution is $1/\lambda$.

To get an intuition for what the exponential distribution actually **is**, imagine that you have a 6 sided dice, and you roll is 6 times per minute, then record the time the first time you roll '1'. On average it'll take you about 1 minute to roll a 1, but it might be faster, or it might be slower. Now imagine you have a 100 sided dice, and you are rolling it 100 times per minute, or a 1000 sided dice being rolled 1000 times per minute. As the number of sides goes up, the chances of rolling a '1' go steadily down, but your number of dice rolls increases. Overall, the amount of time until you roll a 1 stays the same. The exponential distribution is the limit of this; *searching for a very rare event, but doing it very quickly.* A real world example of this type of process is the time take for a particular gene to mutate somewhere in your body.

The exponential distribution has two important properties for this workshop. The first, and most important property is that it is a 'memoryless' distribution. What that means is that if you are waiting for some exponentially distributed event with parameter λ to occur, and then I tell you that the event **has not** occurred by time $t = 3$... then the remaining time until the event occurs will **still** be an exponentially distributed random variable, with the same mean $1/\lambda$. The random variable doesn't remember how long its been waiting for. It is not intuitive if you think it like that, but let's think it with an example and go back to our dice rolls up above; If I've rolled seventeen fives on my dice, none of those dice rolls make it more or less likely that my next dice roll will be a 1. So my waiting time for the next 1 will not be affected by whatever had happened so far.

The second important property of the exponential distribution is what happens when you have two 'racing' exponential distributions. Suppose T_1 and T_2 are two exponentially distributed random variables, with mean values $1/\lambda_1, 1/\lambda_2$ and we want to get the distribution of the first of the two to occur. In mathematical notation, this is equivalent to calculate the distribution of the minimum of T_1, T_2, and guess what? the minimum of two exponential distribution is **also** exponentially distributed, with mean value $1/(\lambda_1 + \lambda_2)$.

The probability that $T_1 < T_2$ is precisely $\lambda_1/(\lambda_1 + \lambda_2)$.

5 Stochastic Processes

A Stochastic process is a series of random variables. Random processes can exists in discrete time in which case we have $X_0, X_1, X_2...$, or in continuous time, in which case we have $X(t)$.

An example of a random process is the random walk. Imagine we have a person at position zero at time zero ($X_0 = 0$). At $t = 1$ they flip a coin, and take a step up if the get heads, and a step down if they get tails;

hence $X_1 = \pm 1$. At time $t = 2$ they repeat the process again, and again at $t = 3$... and so on. So the next step will be whatever they are right now plus or minus one, $X_{i+1} = X_i \pm 1$. Three different simulations of this random walk can be seen in Figure 3 Each X_i in the random walk is a random variable. If we know where the walker is at time $t = 15$, this gives us some information about where they are likely to be at other times, but not perfect information.

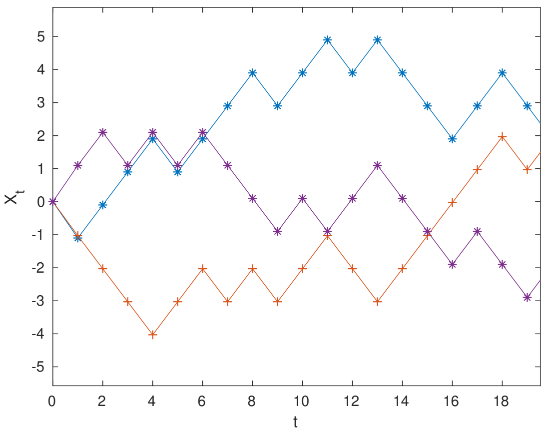

Figure 3: *Three example trajectories for Random Walks.*

6 Markov Chains

An important type of stochastic process is the Markov chain. A Markov chain is a system, composed of a number of 'states' along with a set of transition rates saying how we move from one state to the next. In order for a random process to be a **Markov** chain, transitions must depend only on the *current* state of the system, not on the *past* state. This is known as the 'Memoryless' property. Stated formally we say $P(X_{i+1}|X_i) = P(X_{i+1}|X_0, X_1, ...X_i)$. Markov chains can exist in continuous time or discrete time.

For example, a simplistic model of the whether might describe the weather as 'sunny', 'cloudy', 'rainy' and 'stormy', and give probabilities of changing from one weather pattern to the next in the memoryless manner (see bottom panel in Figure 4). The random walk described above can also be formulated as a Markov chain: the states of the system are the integer numbers, starting at zero, the probability of moving in either direction is 1/2 (top panel in Figure 4).

Markov chains can exist either in discrete time or continuous time. The Random Walk is an example of Markov chain in discrete time, while the simplified weather model described above is an example of a continuous time Markov chain: the weather might be cloudy for ten minutes or ten days, and there are no distinct 'time steps'.

In order to describe a **discrete time Markov chain**, we need to know the *probability* of moving to state i from state j, for each possible pair of states. This information is often stored as a **transition probability matrix**:

$$T_{ij} = \mathbb{P}(X_{t+1} = i|X_t = j).$$

Because the system always has to transfer *somewhere*, the columns of a T add to 1.

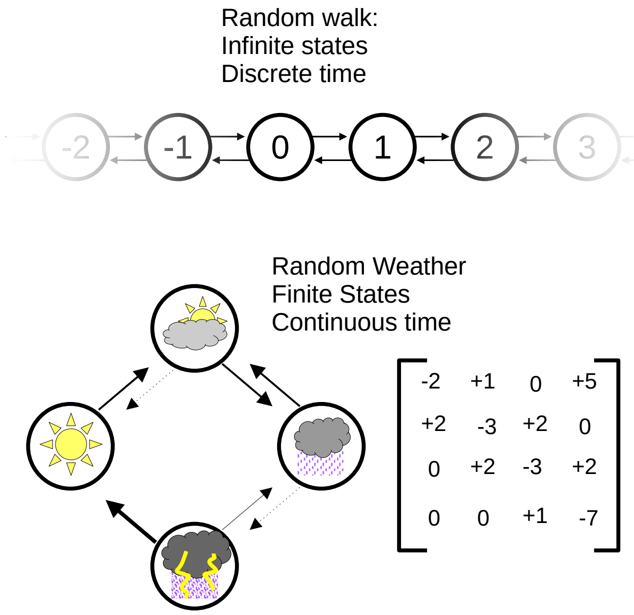

Figure 4: *(Top) The Markov chain for a random walk. This random walk happens in discrete space and time, and has an infinite number of possible states (all of $\mathbb{Z}$). (Bottom) The Markov chain for a stochastic 'weather model'. There is a limited number of states, and transitions can be represented using a transition matrix. Transitions happen in continuous time.*

Continuous time Markov chains are instead described in terms of their **transition rate matrix** Q. For continuous time Markov chains the time taken to transition to state i from state j is an exponentially distributed random variable with rate constant Q_{ij}. The time taken to transition *somewhere* from state j is an exponentially distributed random variable with rate equal to $\sum_{i \neq j} Q_{ij}$. It is usually stated that $Q_{jj} = -\sum_{i \neq j} Q_{ij}$, hence $-Q_{jj}$ is the transition rate *out* of state state j, and columns of Q sum to zero. The probability of moving to state i from state j is $-Q_{ij}/Q_{jj}$.

Markov chains with a finite number of states (like the weather system above) can be described cleanly by their transition matrix (either T or Q, depending on discrete/continuous time). For Markov chains with an infinite or impractically large number of states, T and Q are too large to be stated explicitly. In this case, we generally define transition rates in terms of some rule or algorithm, and never attempt to construct the full transition matrix explicitly. The random walk described above is an example of this sort of Markov chain: it is easy to form the rule, but no transition matrix can be formed because the number of states is infinite.

Epidemics and disease spread are often described using Markov chains where the total number of states is large. Hence, we use some rule to determine the transition rate between states.

Often, Markov chains of this kind are easy to simulate, but difficult or impossible to 'solve' analytically.

7 Brownian Motion

Brownian motion is a particular type of Stochastic process, similar to the random walk. Unlike the random walk which takes discrete steps at discrete time intervals, Brownian motion exists in continuous space and time. Brownian motion is (in some sense) what you get if you have a random walk with very small steps happening very quickly.

Brownian motion is often used for modelling diffusion: the motion of small things in the real world, for example, the movement of a particle, or a value on stock market.

Brownian motion obeys the rule:

$$X(t_2) = X(t_1) + N(0, |t_2 - t_1|)$$

That is to say, the difference in value between two time points is normally distributed, and the variance of this normal distribution is equal to the time difference between the two points. Also, if $t_1 < t_2 \leq t_3 < t_4$, then the increment in X from t_1 to t_2 will be INDEPENDENT of the increment from t_3 to t_4. If the intervals $[t_1, t_2]$ and $[t_3, t_4]$ overlap, than the increments will NOT be independent (to see this imagine $t_1 = t_3$ and $t_2 = t_4$). By default it is assumed that $X(0) = 0$.

This leads to a process that looks like Figure 5

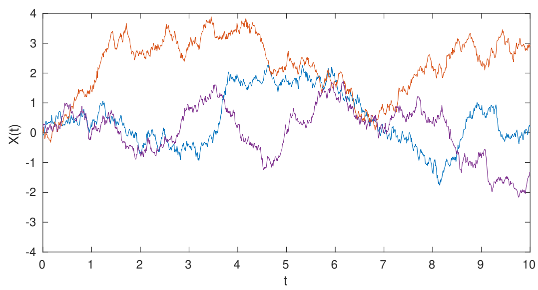

Figure 5: *Three example trajectories for Brownian Motion.*

Here we have given only the most basic and intuitive definition of Brownian motion. More detailed introductory notes can be found **here**.

8 Exercises and further reading

Based on the reading you have done so far, you can expect to be prepared for the workshop. If you would like to slightly more prepared, you can read over **this article by Linda J.S. Allen, from 2017**. This article is nicely written, and gives a preview of pretty much everything we are going to cover.

If you'd prefer more hands on work, rather than more reading try out some of the exercises below:

- Suppose you have 10 apples on a tree. Each of them falls off after an exponentially distributed amount of time, with mean 10 days. How long (on average) until the first apple falls from the tree? What is the expected time until all 10 apples fall from the tree? Can you calculate this analytically? Can you simulate this apple tree?

- When talking about Markov chains, we use the example of a simple weather model as a Markov chain. Do you believe weather in the real world can be modeled using a Markov chain? Why / why not?

- Using your favorite programming language, program up a for loop that starts at zero, and moves 3 steps up 40% of the time, and 2 steps down 60%. If you run the loop 1000 times, how often does it reach -5 (or lower) *before* it reaches $+10$ (or higher).

- Using your favorite programming language, write a program that generates a Brownian motion, and reproduce the kind of plot you see in figure 5.