INTRODUCTION TO EVOLUTIONARY DYNAMICS AND STOCHASTIC CALCULUS

HOYIN CHU

Abstract. This paper introduces evolutionary dynamics and presents simulations done to understand the relation between the fixation probability and connectivity of a graph. The later sections introduce commonly required mathematical background for research in evolutionary dynamics.

Contents

1. Introduction 1
2. Evolutionary Dynamics 1
3. Probability Theory 8
4. Brownian Motion 13
5. Stochastic Calculus 16
References 19

1. Introduction

Evolutionary dynamics studies the mathematical principles of evolution, and a way to model the evolution of a population is by using graphs. The population we are interested in is one that consists of cooperators and defectors, in particular, how does the initial amount of connections between cooperators and defectors impact the probability of cooperator outcompeting defectors. To answer this question we will formalize the representation of cooperators and defectors in a population, and introduce background in probability theory which many of the current research in this field relies on.

2. Evolutionary Dynamics

"Survival of the fittest", as many may know, is the key to evolution. In a competition for survival, the one with the highest fitness would win. We also know one's fitness has ties to reproduction, but that could mean many things. For example, we can describe one's fitness by its number of offspring, or its probability of reproducing under some conditions. We will develop a mathematical representation of fitness by first looking at a classic example in game theory, the prisoner's dilemma.

The prisoner's dilemma describes a scenario where two suspects of a crime were separated into two interrogation rooms and were given the choice to either confess the crime, or to keep silent. If both suspects keep silent, then both would be sentenced to 1 year in jail. If both suspects confess, then both would be sentenced to 7 years in jail. If one confesses while the other keeps silent, then the one who

confesses would be let go with no jail time while the other sentences to 10 years in jail. We can represent the outcome of each scenario in a payoff matrix:

	C	D
C	-1	-10
D	0	-7

The information in the matrix can interpreted as follow: C denotes cooperator, or the behavior to cooperate with their partner, in this context it means to keep silent. D denotes Defector, in this context it means to confess. The matrix is read from row to column and the number denotes the "payoff" of each outcome. For example, if one is a cooperator, then it gets the payoff of -1 from interacting with another cooperator, or -10 from interacting with a defector.

Generalizing from this payoff matrix, we can also use it to model interactions between populations who have behaviors similar to that of cooperators and defectors. In this context, cooperators represent population that is willing to share resources, while defector represent population that is selfish.

	C	D
C	$b-c$	$-c$
D	b	0

In this payoff matrix, b represents benefit and c represents cost. For example, when two cooperators interact, a benefit is received but the act of sharing also result in a cost. When a cooperator and a defector interact, the cooperator pays the cost without receiving anything in return and the defector receives the benefit. When two defector interact, they are both selfish so nothing is gained.
Suppose the frequency of cooperators and defectors in the population is f_c and f_d, according to this matrix, the average payoff p for each type is:

$$p_c = f_c \cdot (b-c) + f_d \cdot (-c)$$
$$p_d = f_c \cdot (b) + f_d \cdot (0)$$

After some simplifications, we see:

$$p_c = f_c \cdot (b) - c \cdot (f_c + f_d)$$
$$p_d = f_c \cdot (b)$$

Because cooperators and defectors make up the entire population, $f_c + f_d = 1$ and we have:

$$p_c = f_c \cdot (b) - c$$
$$p_d = f_c \cdot (b)$$

Since payoff is relative, we don't lose any information by dividing both equations by c. This also make sense because it's not the specific benefit or cost that matters but rather their ratio. The fraction $\frac{b}{c}$ is called the *benefit to cost ratio*.

$$p_c = f_c \cdot (\frac{b}{c}) - 1$$
$$p_d = f_c \cdot (\frac{b}{c})$$

Definition 2.1. Suppose we have a population of cooperators and defectors where the ratio between the number of cooperators and the total population is f_c, and

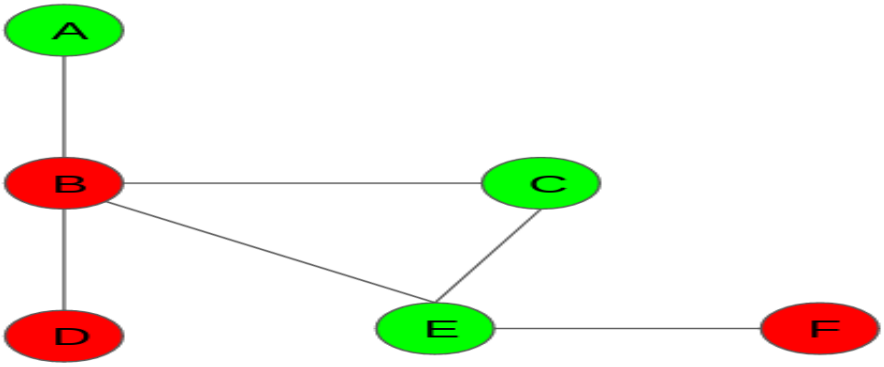

FIGURE 1. An example of a cooperator and defector graph (Co-operators: A,C,E Defectors: B,D,F)

benefit to cost ratio $\frac{b}{c}$: The average payoff of cooperators is :

$$p_c = f_c \cdot (\frac{b}{c}) - 1$$

The average payoff of defectors is:

$$p_d = f_c \cdot (\frac{b}{c})$$

Definition 2.2. Suppose we have selection constant δ and average payoff p, the fitness r is calculated by:

$$r = 1 + \delta p$$

Here is a breakdown of the definition. By having the selection constant δ, which is a non-negative real number, we are allowed to manipulate how much do payoff matter. For example, in an environment where individuals are competing for food, δ being small means food is abundant and the selective pressure is low, where δ being big means food is scarce and the selective pressure is high. When δ is small the system is subject to *weak selection*, on the other hand we call it *strong selection* when δ is big. When $\delta = 0$, it means the payoff do not matter at all and we call this *neutral drift*. By adding the 1 to the expression δp, we can get a value between 0 and 1 when the fitness of cooperators and defectors are normalized, which can represent the probability of each reproducing.

To put "probability of reproducing" in context, consider a population modeled on a graph (Figure 1), where each vertex represents a cooperator or defector. At each timestep, a vertex is randomly selected to die and one of its neighbor is chosen to reproduce and replace the original vertex, where the chance of being chosen to reproduce is proportional to its fitness with rest of the neighbors. This process is known as the *death-birth update*. When all of the vertices become defector or cooperator, we reached an *absorbing state* and the graph will no longer change.

For example, currently the payoff of each of the vertex in Figure 1 is:

$$\begin{aligned}
p_A &= f_c \cdot (\frac{b}{c}) - 1 = 0 \cdot (\frac{b}{c}) - 1 = -1 \\
p_B &= f_c \cdot (\frac{b}{c}) = \frac{3}{4} \cdot \frac{b}{c} \\
p_C &= f_c \cdot (\frac{b}{c}) - 1 = \frac{1}{2} \cdot \frac{b}{c} - 1 \\
p_D &= f_c \cdot (\frac{b}{c}) = 0 \\
p_E &= f_c \cdot (\frac{b}{c}) - 1 = \frac{1}{3} \cdot \frac{b}{c} - 1 \\
p_F &= f_c \cdot (\frac{b}{c}) = \frac{b}{c}
\end{aligned}$$

So the fitness of each of the vertex in Figure 1 is:

$$\begin{aligned}
r_A &= 1 + \delta p_A = 1 - \delta \\
r_B &= 1 + \delta p_B = 1 + \delta(\frac{3}{4} \cdot \frac{b}{c}) \\
r_C &= 1 + \delta p_C = 1 + \delta(\frac{1}{2} \cdot \frac{b}{c} - 1) \\
r_D &= 1 + \delta p_D = 1 \\
r_E &= 1 + \delta p_E = 1 + \delta(\frac{1}{3} \cdot \frac{b}{c} - 1) \\
r_F &= 1 + \delta p_F = 1 + \delta(\frac{b}{c})
\end{aligned}$$

Suppose vertex C is chosen to die, it could be replaced by vertex B or E. We can calculate the probability of each happening by:

$$\begin{aligned}
P(\text{B reproduces}) &= \frac{r_B}{r_B + r_E} = \frac{1 + \delta p_B}{(1 + \delta p_B) + (1 + \delta p_E)} = \frac{1 + \delta p_B}{2 + \delta(p_B + p_E)} \\
P(\text{E reproduces}) &= \frac{r_E}{r_B + r_E} = \frac{1 + \delta p_E}{(1 + \delta p_B) + (1 + \delta p_E)} = \frac{1 + \delta p_d}{2 + \delta(p_B + p_E)}
\end{aligned}$$

Notice when the system is under neutral drift ($\delta = 0$), the chance of each replacement happening is even at $\frac{1}{2}$, which is consistent with our assumption that when payoff doesn't matter at all replacements are purely random.

To measure the "competitiveness" of cooperators, we will consider the following quantity: given a initial configuration of a graph with k cooperators, the probability that the cooperators will take over the entire graph. More formally, suppose $s(t)$ denotes the size of cooperators at time t and there are N vertices in the graph, we are interested in:

$$P(\lim_{t \to \infty} s(t) = N) = P(\lim_{t \to \infty} \frac{s(t)}{N} = 1)$$

Since we know $s(t)$ will be 0 or N eventually, the probability of this expression is equivalent to its expected value

$$E[\lim_{t\to\infty}\frac{s(t)}{N}] = 0 + 1 \cdot P(\lim_{t\to\infty}\frac{s(t)}{N} = 1) = P(\lim_{t\to\infty}\frac{s(t)}{N} = 1)$$

To take into account that we started with k cooperators, we turn this into a conditional expectation:

$$E[\lim_{t\to\infty}\frac{s(t)}{N} \mid s(0) = k] = \frac{1}{N} \cdot E[\lim_{t\to\infty} s(t) \mid s(0) = k]$$

This is known as the *fixation probability* of the cooperators on this graph.

Definition 2.3. Given a graph with N vertices and k cooperators, a function $s(t)$ representing the size of the cooperator, the *fixation probability* of the cooperators is defined as:

$$\frac{1}{N} \cdot E[\lim_{t\to\infty} s(t) \mid s(0) = k]$$

In general, computing fixation probability is difficult but we can compute it rather easily for neutral drifts and regular graphs.

Definition 2.4. A *d-regular graph* is a graph where every vertex has d edges.

Theorem 2.5. *Under neutral drift, the fixation probability of k cooperators on a d-regular graph with N vertices is* $\frac{k}{N}$

We will give an informal proof here, interested readers can refer to Allen et al's paper from 2016 for a more precise and general result[1].

Proof. We claim that $E[s(t+h) - s(t)] = 0$ for any $t, h > 0$. In other words,

$$E[\lim_{t\to\infty} s(t)] = E[s(0)]$$

To see this, we first look at the case when $h = 1$ and write the expectation as:

$$E[s(t+1) - s(t)] = \sum_{i\to j} p(i \to j)\Delta_{i\to j}$$

where $p(i \to j)$ denotes the probability of vertex i replacing vertex j, $\Delta_{i\to j}$ denotes the change in the cooperator size resulting from the replacement. Note that $\Delta_{i\to j}$ can only be -1,0 or 1.

Since $\Delta_{i\to j} = 0$ when a vertex is replaced by a vertex of the same kind, we are only interested in the probability when it is a cooperator replacing a defector or vice versa. Suppose vertex i is a cooperator and vertex j is a defector, then the probability of i replacing j is $\frac{1}{N} * \frac{1}{d}$ because there is $\frac{1}{N}$ chance j is chosen to die and $\frac{1}{d}$ chance i is chosen to reproduce. This is also true when i is a defector and j is a cooperator.

$$\sum_{i\to j} p(i \to j)\Delta_{i\to j} = \sum_{i\to j} \frac{1}{N} \cdot \frac{1}{d} \cdot \Delta_{i\to j} = \frac{1}{Nd} \cdot \sum_{i\to j} \Delta_{i\to j}$$

Since $\Delta_{i\to j}$ can only be 1 or -1, and for every cooperator replacing defector (1) we can also find a reverse edge (-1), the sum $\sum_{i\to j} \Delta_{i\to j} = 0$ which means $E[s(t+1) - s(t)] = 0$ and $E[s(t+h) - s(t)] = 0$ for any $t, h > 0$. Given that $E[s(t+h) - s(t)] = 0$

for any $t, h > 0$, we can conclude $E[\lim_{t\to\infty} s(t)] = E[s(0)]$. Since we are given that $s(0) = k$, $E[s(0)] = E[k] = k$ and the fixation probability

$$\frac{1}{N} \cdot E[\lim_{t\to\infty} s(t) \mid s(0) = k] = \frac{k}{N}$$

□

Now that we have established a "baseline" fixation probability for neutral drift, we will consider weak selection (when δ is small). Since selection constant δ is no longer 0, we have to take into account of the payoff, more specifically, the benefit to cost ratio.

Definition 2.6. The threshold $\frac{b}{c}$ (the fraction in the payoff equation) that would give cooperators a higher fixation probability than in the neutral drift case is known as the *critical ratio*.

In a 2006 paper, Otsuki estimated that under weak selection, the critical ratio of any graph is the average degree of the graph[3]. In 2016, this estimate was made more exact. In particular, for a regular graph, $\frac{b}{c}$ would need to be larger than $\frac{n-2}{\frac{n}{d}-2}$ [1]

However, not much work have been done in the literature on strong selection, when $\lim_{\delta\to\infty}$. For ease of computation under strong selection, we adopt the fitness equation used by Altrock [5]

$$r = e^{\delta p}.$$

For example, if vertex C were to die, then the probability of each its neighbor reproducing is:

$$P(\text{B reproduces}) = \frac{e^{\delta p_B}}{e^{\delta p_B} + e^{\delta p_E}} = \frac{1}{1 + e^{\delta(p_E - p_B)}} = \frac{1}{1 + e^{\delta(-\frac{1}{2}(\frac{b}{c})-1)}}$$

$$P(\text{E reproduces}) = \frac{e^{\delta p_E}}{e^{\delta p_B} + e^{\delta p_E}} = \frac{1}{1 + e^{\delta(p_B - p_E)}} = \frac{1}{1 + e^{\delta(\frac{1}{2}(\frac{b}{c})+1)}}$$

As we can see in the denominator, as δ goes to infinity the value largely depends on whether the difference in the payoff is positive or negative, which is independent of the benefit to cost ratio but rather the frequency of cooperators in their neighbors. To better understand the fixation probability of cooperators in strong selection, we investigated in connectivity, or the amount of edges connecting cooperators and defectors.

We conjecture that in strong selection, there exist a quantity that depends on the connectivity of the graph, such that the quantity has similar effect to that of critical ratio in weak selection, where once it is reached the fixation probability of cooperators would be higher than that of neutral drift. To confirm, we started gathering empirical data on fixation probability by running simulations, starting from regular graphs and neutral drift. Our method of simulation is as follow:

(1) Choose the parameters of a regular graph (N total vertices and degree d) and k initial cooperators,
(2) Calculate the maximum possible initial connectivity for k initial cooperators
(3) For every possible initial connectivity c, calculate the fixation probability by doing the following simulations 500 times and count the amount of times cooperators win.

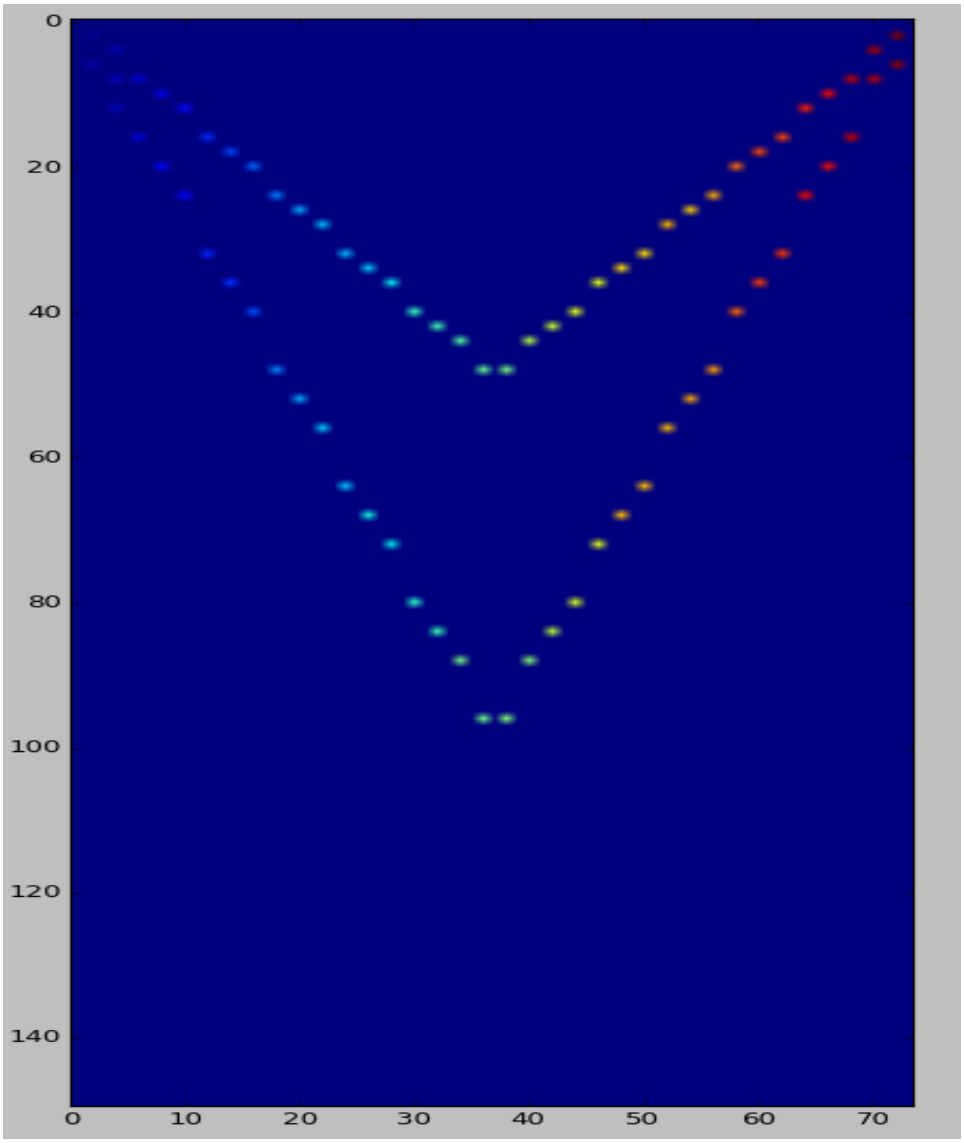

FIGURE 2. X-axis = Initial Cooperators; Y-axis = Initial Connectivity

- Create a regular graph with N total vertices, degree d, k starting cooperators and c edges between cooperators and defectors.
- Until the graph is all cooperators or all defectors, repeat the death-birth update.

We performed the above simulation on regular graphs with 74 vertices and degree 4. Due to time restriction, we only calculated fixation probability for even amount of cooperators, and for every cooperator, the fixation probability when the initial connectivity is at 1/3 and 2/3 of maximum connectivity. In Figure 2, a redder color represents a higher number. This result agrees with our analytic result that

the fixation probability of cooperators on regular graph is $\frac{k}{N}$ as we see the colors becomes more red as cooperator size increases, and the colors of same cooperator but different connectivity do not have significant difference.

We then attempted to gather fixation probability from strong selection by using the same simulation method with a different δ parameter. Unfortunately we weren't able to gather enough data as it takes a lot longer than the case with neutral drift. We suspect the increase in runtime is due to the fact that under strong selection, the amount of "pseudo-absorbing state" increases and the death-birth update may get stuck for a long time in these states before being able to converge to a true absorbing state. To tackle this problem, we plan to put a threshold on how long the system can stay at a state, and if the system stays longer than a threshold then instead of increasing the amount of win of cooperators by 1, we increase the amount by the ratio of cooperators to the total number of vertices.

During our investigation we also encountered interesting problems that is worth looking into, such as formalizing the aforementioned "pseudo-absorbing" states and optimizing simulation method to decrease run time complexity. Unfortunately due to time limitation we were not able to have a deeper look at these problems, but we hope to investigate in them in the future.

3. Probability Theory

Many current researchers in evolutionary dynamics make use of stochastic calculus[3] which has its roots in probability theory. By understanding probability theory we give ourselves more tools to tackle new problems.

Definition 3.1. Given a set Ω, a *σ-algebra* $\mathcal{F}$ is a collection of subsets of Ω with the following properties:

(1) $\emptyset \in \mathcal{F}$
(2) $\mathcal{F}$ is closed under complements
(3) $\mathcal{F}$ is closed under countable unions

Example 3.2. Suppose $\Omega_D = \{1,2,3,4,5,6\}$ and $\mathcal{F}_1 = \{\emptyset, \Omega, \{1,2,3\}, \{4,5,6\}\}$, $\mathcal{F}_2 = 2^{\Omega_D}$ (Power set of Ω_D) and $\mathcal{F}_3 = \{\emptyset, \Omega, \{1,2\}, \{3,4\}, \{5,6\}, \{1,2,3,4\}, \{1,2,5,6\}, \{3,4,5,6\}\}$, $\mathcal{F}_4 = \{\emptyset, \Omega, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$. $\mathcal{F}_1$, $\mathcal{F}_2$, and $\mathcal{F}_3$ are σ-algebras of Ω while $\mathcal{F}_4$ is not (because it is not closed under countable unions or complements).

Definition 3.3. A pair $(\Omega, \mathcal{F})$ in which $\mathcal{F}$ is a σ-algebra is called a *measurable space*. Subsets in Ω that are also in $\mathcal{F}$ are called *measurable sets*.

Definition 3.4. Given A which is a collection of subsets in Ω, the smallest σ-algebra of Ω that contains A is the called the *σ-algebra generated by A*. More formally

$$\mathcal{H}_A = \bigcap \{\mathcal{H} : \mathcal{H} \text{ is a } \sigma\text{-algebra of } \Omega, A \subseteq \mathcal{H}\}$$

$\mathcal{H}_A$ is the σ-algebra generated by A.

Example 3.5. Let $\Omega = \Omega_D$ and $A = \{\{1,2\}, \{3,4\}\}$. Then $\mathcal{F}_3$ from Example 3.2 is the σ-algebra generated by A.

Definition 3.6. The σ-algebra generated by the collection of open sets of $\mathbb{R}^n$ is called the *Borel Algebra* and is denoted by $\mathcal{B}$.

Definition 3.7. If $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ are measurable spaces, a function $X : \Omega \to \Omega'$ is called *$\mathcal{F}$-measurable* if for each measurable set in $\mathcal{F}'$, the preimage is measurable in $\mathcal{F}$. More formally:

$$X^{-1}(A) := \{\omega \in \Omega; X(\omega) \in A\}.$$

X is a measurable function if $X^{-1}(A) \in \mathcal{F}$ for all $A \in \Omega'$.

Example 3.8. Suppose $\Omega = \Omega_D, \mathcal{F} = \mathcal{F}_1$ from 3.2. $\Omega' = \mathbb{R}, \mathcal{F}' = \mathcal{B}$ and $X_D : \Omega \to \Omega'$ where $X(\{1,2,3\}) = -50, X(\{4,5,6\}) = 100, X(\emptyset) = 0, X(\Omega) = 0$. X is measurable function.

Definition 3.9. A *probability measure* P on a measurable space $(\Omega, \mathcal{F})$ is a function $\mathcal{F} \to [0,1]$ such that

(1) $P(\emptyset) = 0$
(2) $P(\Omega) = 1$
(3) $P(A_1 \cup A_2 ... \cup A_n) = \sum_{i=1}^{n} P(A_i)$, where A_i are pairwise disjoint sets.

Example 3.10. A possible probability measure P for the measurable space $(\Omega_D, \mathcal{F}_1)$ is $P(\emptyset) = 0, P(\Omega) = 1, P(\{1,2,3\}) = 0.6, P(\{4,5,6\}) = 0.4$.

Definition 3.11. The triple $(\Omega, \mathcal{F}, P)$ is called a *probability space*. We will use this notation unless otherwise stated.

Definition 3.12. In a probability space, measurable sets are called *events*.

Example 3.13. Suppose we are rolling a fair die, the probability space is $(\Omega_D, 2^{\Omega_D}, P_D)$ where $P_D(\{1\}) = P_D(\{2\}) = ... = P_D(\{6\}) = \frac{1}{6}$. The event "Rolling a number less than 3" is the measurable set $\{1,2\}$ and the event "Rolling an odd number" is the measurable set $\{1,3,5\}$.

Definition 3.14. A *random variable* X is an $\mathcal{F}$-measurable function $X : \Omega \to \mathbb{R}^n$ where Ω is from a probability space.

Example 3.15. Suppose X_D is a random variable with respect to the probability space $(\Omega_D, 2^{\Omega_D}, P_D)$ and $X_D(\{1,3,5\}) = -50$ and $X_D(\{2,4,6\}) = 100$. Then X_D can be understood as "winning after rolling a die" where one loses 50 on odd numbers and wins 100 on even numbers.

Although random variables are functions, they are usually thought of as values that depend on the events in Ω. Let ω be an element from Ω in a probability space and B a subset from $\mathcal{B}$, per convention, we usually write $X(\omega)$ as X and $P(X^{-1}(B))$ as $P(X \in B)$.

Definition 3.16. Suppose we have a probability space $(\Omega, \mathcal{F}, P)$ and Borel algebra $\mathcal{B}$. Let X be a random variable $X : \Omega \to R^n$. Then

$$\mathcal{H}_X := \{X^{-1}(B) : B \in \mathcal{B}\}$$

is a σ-algebra called the σ-algebra generated by X and is the smallest sub-σ-algebra of $\mathcal{F}$ in which X is measurable.

Definition 3.17. For a random variable X, we define its *expected value* to be

$$\mathrm{E}[X] := \int_{\Omega} X dp$$

Example 3.18. Suppose $X_D(\{1,3,5\}) = -50$ and $X_D(\{2,4,6\}) = 100$. The expected value of the random variable X_D with respect to the probability space $(\Omega_D, 2^{\Omega_D}, P_D)$ is:

$$\mathrm{E}[X_D] = 0.5 \cdot (-50) + 0.5 \cdot (100) = 25$$

Remark 3.19. The following properties of expected values follow directly from properties of integrals: If c is constant and X, Y are two random variables

(1) $\mathrm{E}[c] = c$
(2) $\mathrm{E}[X - Y] = \mathrm{E}[X] - \mathrm{E}[Y]$
(3) $\mathrm{E}[cX] = c\,\mathrm{E}[X]$

This also means expected value is a *linear operator*. We will give a precise definition of a linear operator later.

Based on expectation, we define *covariance* and *variance*. We will prove some of their properties in the following proposition, and the remaining proofs can be found in Lawrence's text[2].

Definition 3.20. The *covariance* of two random variables X, Y is defined as

$$\mathrm{Cov}[X, Y] = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$$

Proposition 3.21. *If a, b are constants and X, Y are two random variables*

(1) $\mathrm{Cov}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y]$

Proof. To see this, first expand $\mathrm{Cov}[X, Y]$ from definition:

$$\begin{aligned}
\mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])] &= \mathrm{E}[XY - X\,\mathrm{E}[Y] - Y\,\mathrm{E}[X] + \mathrm{E}[X]\,\mathrm{E}[Y]] \\
&\quad \text{by Remark 3.19 (2)} \\
&= \mathrm{E}[XY] - \mathrm{E}[X\,\mathrm{E}[Y]] - \mathrm{E}[Y\,\mathrm{E}[X]] + \mathrm{E}[\mathrm{E}[X]\,\mathrm{E}[Y]] \\
&\quad \text{E[X], E[Y] are constants, by Remark 3.19 (3)} \\
&= \mathrm{E}[XY] - \mathrm{E}[Y]\,\mathrm{E}[X] - \mathrm{E}[X]\,\mathrm{E}[Y] + \mathrm{E}[X]\,\mathrm{E}[Y] \\
&= \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] \\
&\quad \text{As required}
\end{aligned}$$

□

(2) $\mathrm{Cov}[X, Y] = \mathrm{Cov}[Y, X]$

Proof. $\mathrm{Cov}[X, Y] = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])] = \mathrm{E}[(Y - \mathrm{E}[Y])(X - \mathrm{E}[X])] = \mathrm{Cov}[Y, X]$ □

(3) $\mathrm{Cov}[aX, Y] = a\,\mathrm{Cov}[X, Y]$

Proof. $\mathrm{Cov}[aX, Y] = \mathrm{E}[aXY] - \mathrm{E}[aX]\,\mathrm{E}[Y] = a(\mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y]) = a\,\mathrm{Cov}[X, Y]$ □

(4) $\mathrm{Cov}[X, a] = 0$

Proof. $\mathrm{Cov}[X, a] = \mathrm{E}[aX] - \mathrm{E}[X]\,\mathrm{E}[a] = a\,\mathrm{E}[X] - a\,\mathrm{E}[X] = 0$ □

(5) $\mathrm{Cov}[X + Y, Z] = \mathrm{Cov}[X, Z] + \mathrm{Cov}[Y, Z]$

Definition 3.22. The *variance* of a random variable X is defined as

$$\operatorname{Var}[X] = \operatorname{E}[(X - \operatorname{E}[X])^2]$$

Remark 3.23. If c is constant and X, Y are two random variables

(1) $\operatorname{Var}[c] = 0$
(2) $\operatorname{Var}[X] = \operatorname{Cov}[X, X]$
(3) $\operatorname{Var}[X] = \operatorname{E}[X^2] - \operatorname{E}[X]^2$

Proof.

$$\begin{aligned}\operatorname{Var}[X] &= \operatorname{E}[(X - \operatorname{E}[X])^2] \\ &= \operatorname{E}[X^2 - 2X \operatorname{E}[X] + \operatorname{E}[X]^2] \\ &= \operatorname{E}[X^2] - \operatorname{E}[2X \operatorname{E}[X]] + \operatorname{E}[\operatorname{E}[X]^2] \\ &= \operatorname{E}[X^2] - 2 \operatorname{E}[X]^2 + \operatorname{E}[X]^2 \\ &= \operatorname{E}[X^2] - \operatorname{E}[X]^2\end{aligned}$$

□

(4) $\operatorname{Var}[cX] = c^2 \operatorname{Var}[X]$

Proof.

$$\begin{aligned}\operatorname{Var}[cX] &= \operatorname{E}[(cX - \operatorname{E}[cX])^2] \\ &= \operatorname{E}[(cX - c \operatorname{E}[X])^2] \\ &= \operatorname{E}[(c(X - \operatorname{E}[X]))^2] \\ &= \operatorname{E}[c^2 (X - \operatorname{E}[X])^2] \\ &= c^2 \operatorname{E}[(X - \operatorname{E}[X])^2] \\ &= c^2 \operatorname{Var}[X]\end{aligned}$$

□

(5) $\operatorname{Var}[X + Y] = \operatorname{Var}[X] + \operatorname{Var}[Y] + 2 \operatorname{Cov}[X, Y]$

Proof. $\operatorname{Var}[X + Y] = \operatorname{E}[(X + Y)^2] - \operatorname{E}[X + Y]^2$
$\operatorname{Var}[X + Y] = \operatorname{E}[X^2 + 2XY + Y^2] - \operatorname{E}[X + Y]^2$
$\operatorname{Var}[X + Y] = \operatorname{E}[X^2] + 2 \operatorname{E}[XY] + \operatorname{E}[Y^2] - (\operatorname{E}[X] + \operatorname{E}[Y])^2$
$\operatorname{Var}[X + Y] = \operatorname{E}[X^2] + 2 \operatorname{E}[XY] + \operatorname{E}[Y^2] - (\operatorname{E}[X]^2 + 2 \operatorname{E}[X] \operatorname{E}[Y] + \operatorname{E}[Y]^2)$
$\operatorname{Var}[X + Y] = (\operatorname{E}[X^2] - \operatorname{E}[X]^2) + (\operatorname{E}[Y^2] - \operatorname{E}[Y]^2) - 2(\operatorname{E}[XY] - \operatorname{E}[X][Y])$
$\operatorname{Var}[X + Y] = \operatorname{Var}[X] + \operatorname{Var}[Y] + 2 \operatorname{Cov}[X, Y]$

□

(6) $\operatorname{Var}[X + c] = \operatorname{Var}[X]$

Definition 3.24. Let X be a random variable, we define

$$F_X(x) := P(X \leq x) \text{ for all x } \in R^n.$$

to be the *Cumulative Distribution Function (CDF)*. It is also referred to as the distribution function.

Example 3.25. Let the probability space be $(\Omega, 2^{\Omega_D}, P_D)$ and random variable X_D defined by $X_D(\{1,3,5\}) = -50$ and $X_D(\{2,4,6\}) = 100$. The CDF of X_D has the following values:

$$F_{X_D}(-100) = 0, F_{X_D}(0) = 0.5, F_{X_D}(99) = 0.5, F_{X_D}(101) = 1.$$

Definition 3.26. Suppose X is a random variable and F_X is its CDF. If there exist a function $f : R^n \to \mathbb{R}$ such that

$$F_X(x) = F_X(x_1, x_2, ..., x_n) = \int_{-\infty}^{x_1} ... \int_{-\infty}^{x_n} f(y_1, ..., y_n) dy_n ... dy_1$$

then f is called the probability density function (PDF) of X. It is also referred to as the density function.

This give us a way to compute

$$P(X \in B) = \int_B f(x)dx \text{ for all B} \in \mathcal{B}.$$

Definition 3.27. If a random variable X has mean μ, variance σ^2, and a density function of

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Then we say X has a *normal (Gaussian) distribution* with mean μ and variance σ^2. Denoted $X \sim N(\mu, \sigma^2)$. If $\mu = 0$ and $\sigma^2 = 1$ then we say X has a *standard normal distribution*.

Definition 3.28. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Two events A, B are called *independent* if

$$P(A \cap B) = P(A)P(B).$$

Example 3.29. Suppose we have probability space $(\Omega_D, 2^{\Omega_D}, P_D)$ Let A denote the event "Rolling a number less than 3" (the measurable set $\{1,2\}$) and B denote the event "Rolling an even number" (the measurable set $\{2,4,6\}$). $P(A \cap B) = P(\{2\}) = \frac{1}{6}, P(A)P(B) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$. Hence the two events are independent.

Definition 3.30. Let $A_1, A_2..A_n$ be events. These events are independent if for all choices of $1 \leq k_1 < k_2 < ... < k_m$, we have

$$P(A_{k_1} \cap A_{k_2} \cap ... \cap A_{k_m}) = P(A_{k_1})P(A_{k_2}) \cdots P(A_{k_m})$$

Definition 3.31. Let $\mathcal{V}_i$ be sub σ-algebras of $\mathcal{F}$ for $i = 1, \cdots n$. We say that these σ-algebras are independent if for all choices of $1 \leq k_1 < k_2 < ... < k_m$, and of events $A_{k_i} \in \mathcal{V}_{k_i}$ we have

$$P(A_{k_1} \cap A_{k_2} \cap ... \cap A_{k_m}) = P(A_{k_1})P(A_{k_2}) \cdots P(A_{k_m})$$

Definition 3.32. Let $X_i : \Omega \to \mathbb{R}^n$ be random variables for $i = 1, \cdots m$, we say these random variables are *independent* if the sigma algebras $\mathcal{H}_{X_i}$ are independent.

Theorem 3.33. *(Central Limit Theorem) Let $X_1, X_2, ..., X_n$ be i.i.d (independent and identically distributed) random variables each with finite expected value μ and finite variance σ^2. Then the random variable*

$$Z_n = \frac{X_1 + X_2 + ... + X_n - n\mu}{\sqrt{n}\sigma} = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

converges in distribution to the standard normal distribution as n goes to infinity. That is

$$\lim_{n\to\infty} P(Z_n \leq x) = \Phi(x)$$

where $\Phi(x)$ *is the standard normal CDF.*

Definition 3.34. Let Y be a random variable. Then $E[X|Y]$ is any $\mathcal{H}_Y$ measurable random variable such that

$$\int_A X dP = \int_A E[X|Y] dP \text{ for all A } \in \mathcal{H}_Y.$$

Example 3.35. Let the probability space be $(\Omega_D, 2^{\Omega_D}, P_D)$ and random variable $X(\{1,2,3\}) = -100, X(\{4,5,6\}) = 100$ and random variable $Y(\{1,3,s5\}) = 50, Y(\{2,4,6\}) = -50$.
$\mathcal{H}_Y = \{\emptyset, \Omega_D, \{1,3,5\}, \{2,4,6\}\}$
$\int_A X dP = X(\{1,3,5\}) \cdot P(\{1,3,5\}) + X(\{2,4,6\}) \cdot P(\{2,4,6\}) = \frac{1}{2} \cdot (-100) + \frac{1}{2} \cdot (100) = 0$
$\int_A \mathrm{E}[X|Y] dP = E[X|Y = 50] + E[X|Y = -50] = E[X|\{1,3,5\}] + E[X|\{2,4,6\}] = \frac{2}{3} \cdot (-100) + \frac{1}{3} \cdot (100) + \frac{2}{3} \cdot (100) + \frac{1}{3} \cdot (-100) = 0$

Definition 3.36. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{V}$ be a σ-algebra such that $\mathcal{V} \subseteq \mathcal{F}$. If $X : \Omega \to \mathbb{R}^n$ is an integrable random variable, then we define:

$$E[X|\mathcal{V}]$$

to be any random variable on Ω such that

(1) $E[X|\mathcal{V}]$ is V-measurable
(2) $\int_A X dp = \int_A E[X|\mathcal{V}] dp$ for all $A \in \mathcal{V}$

The following properties about conditional expectations are true. The proof can be found on page 31 in Lawrence's text[2]

Proposition 3.37.
(1) $E[X|Y] = E[X|\mathcal{H}_Y]$
(2) $E[E[X|\mathcal{V}]] = E[X]$
(3) *If X is V-measurable,* $E[X|\mathcal{V}] = X$
(4) *If a,b constant,* $E[aX + bY|V] = aE[X|\mathcal{V}] + bE[Y|\mathcal{V}]$
(5) *If X is* $\mathcal{V}$*-measurable and XY integrable,* $E[XY|\mathcal{V}] = XE[Y|\mathcal{V}]$
(6) *If X is independent of* $\mathcal{V}$*,* $E[X|\mathcal{V}] = E[X]$

4. Brownian Motion

We can describe the size of the cooperators in our model overtime as a stochastic process. In this section we present Brownian Motion which is the building block of many stochastic processes. Due to time limitation, we will not provide detailed example in this section and it serves more as a road map for further reading. Readers can find the definition and proofs from [4] and [2] where much of this section came from.

Definition 4.1. A collection of random variables $\{X(t)|t \geq 0\}$ is called a stochastic process

Definition 4.2. For each point $\omega \in \Omega$, the mapping $t \to X(t, \omega)$ is the corresponding sample path

Definition 4.3. A stochastic process is called a Brownian Motion or Wiener Process if

(1) $B(0) = 0$
(2) $B(t) - B(s) \sim N(0, t-s)$ for all $t \geq s \geq 0$
(3) for all times $t_1 < t_2 < ... < t_n$, the random variables $B(t_1), B(t_2) - B(t_1)...B(t_n) - B(t_{n-1})$ are independent
(4) With probability one, the function $t \to B(t)$ is a continuous function of t

Definition 4.4. If $B(t)$ is a Brownian Motion, then $Y(t) = \sigma B(t) + mt$ is also a Brownian Motion with drift m and variance σ^2. When $\sigma^2 = 1$ and $m = 0$ the process is called Standard Brownian Motion.

Theorem 4.5. *Brownian Motion exist*

Theorem 4.6. *(Re-scaling Property) Suppose $B(t)$ is a standard Brownian Motion and $a > 0$, and*

$$Y(t) = \frac{B(at)}{\sqrt{a}}$$

Then $Y(t)$ is also a standard Brownian Motion

Theorem 4.7. *With probability one, Brownian Motion is nowhere differentiable*

Definition 4.8. Let X be a stochastic process, then the σ-algebra generated by the random variable $X(s)$ for $0 \leq s \leq t$ is called the history of the process until time t. More formally:

$$\mathcal{H}(t) := \mathcal{H}(X(s)|0 \leq s \leq t)$$

Where $\mathcal{H}(t)$ is the history of X until time t

Definition 4.9. Let X be a stochastic process such that $E[|X|] < \infty$ for all $t \geq 0$. If

$$X(s) = E[X(t)|\mathcal{H}(s)] \text{ for all } 0 \leq s \leq t$$

Then X is called a Martingale

We can rewrite this definition as

$$E[X(t) - X(s)|\mathcal{H}(s)] = 0$$

Intuitively, we can interpret a Martingale as a "fair game" where we expect to neither lose or gain anything in the future.

Theorem 4.10. *A Standard Brownian Motion is a Martingale*

Proof. Let $B(t)$ be a Standard Brownian Motion, $0 \leq s \leq t$

$E[B(t)|\mathcal{H}_{B(s)}] = E[B(t)-B(s)|\mathcal{H}_{B(s)}]+E[B(s)|\mathcal{H}_{B(s)}] = E[B(t)-B(s)]+B(s) = B(s)$

□

Definition 4.11. If $\mathcal{V}$ is a σ-algebra, $\mathcal{V} \subseteq \mathcal{F}$, then we define the conditional probability of A given $\mathcal{V}$ by:

$$P(A|\mathcal{V}) := E[1_A|\mathcal{V}] \text{ for all } A \in \mathcal{F}$$

where 1_A is the indicator function for A

Definition 4.12. A stochastic process is called a Markov Process if

$$P(X(t) \in B|\mathcal{H}(s)) = P(X(t) \in B|X(s)) \text{ for all } 0 \leq s \leq t \text{ and } B \in \mathcal{B}$$

Intuitively, a stochastic process is a Markov Process if the future of the process depends only on the current state and is independent of the past.

Theorem 4.13. *A Standard Brownian Motion is a Markov process*

Proof. Since a standard Brownian Motion has independent increments, let f be a bounded measurable function $f : R \to R$ and $0 \leq s \leq t$

$$E[f(B(t))|\mathcal{H}(s)] = E[f(B(t))|B(s)]$$

In particular, let $f = 1_B$, where $B \in \mathcal{B}$, then

$$P(B(t) \in B|\mathcal{H}(s)) = P(B(t) \in B|B(s))$$

□

Definition 4.14. A (linear) operator is a function L from (a linear subspace of) functions to functions satisfying $L(af+bg) = aL(f)+bL(g)$ where a, b are constants and f, g are functions. We also use the notation L_x to indicate the operator acting on the variable x

Theorem 4.15. *Suppose B_t is a standard Brownian motion with drift μ and variance σ^2. Then the transition density $p_t(x, y)$ satisfies the heat equation*

$$\partial_t p_t(x, y) = L^*_y p_t(x, y)$$

with initial condition $p_0(x, .) = \partial_x(.)$ Here L^* is the operator on functions

$$L^* f(y) = -\mu f'(y) + \frac{\sigma^2}{2} f''(y)$$

Theorem 4.16. *Suppose B_t is a standard Brownian motion with drift μ and variance σ^2. Then the transition density $p_t(x, y)$ satisfies the heat equation*

$$\partial_t p_t(x, y) = L_x p_t(x, y)$$

with initial condition $p_0(., y) = \partial_y(.)$ Here L is the operator on functions

$$Lf(y) = \mu f'(y) + \frac{\sigma^2}{2} f''(x)$$

Definition 4.17. The (infinitesimal) generator of a Markov Process X_t is the operator L on functions defined by

$$Lf(x) = \lim_{t \to 0} \frac{E[f(X_t)|X_0 = x]f(x) - f(x)}{t}$$

Definition 4.18. The generator of Brownian motion with drift m and variance σ^2 is defined by

$$Lf(x) = mf'(x) + \frac{\sigma^2}{2} f''(x)$$

Theorem 4.19. *The Chapman-Kolmogorov Equation is*

$$p_{s+t}(x) = \int_{-\infty}^{\infty} p_s(y) p_t(x - y) dy$$

5. Stochastic Calculus

As a complement to ordinary calculus, stochastic calculus allow us to study instantaneous changes in random processes and it has proven its usefulness in the 2006 Ohtsuki paper where diffusion was used as tool to estimate the fixation probability of cooperators[3]. Due to time limitation, we will not provide detailed example in this section and it serves more as a road map for further reading. Readers can find the definition and proofs from [4] and [2] where much of this section came from.

Definition 5.1. A stochastic process A_t is called a simple process if there exist times $0 = t_0 < t_1 < ... < t_n < \infty$ and random variables $Y_j, j = 0, 1, ...n$ that are $\mathcal{H}_{t_j}-$ measurable such that $A_t = Y_j$ for all $t \in [t_j, t_{j+1})$. Intuitively, we can think of A_t as a step function, or a process where you are only allowed to "change bet" at certain times

Definition 5.2. If $A(t)$ is a simple process and we define

$$Z_{t_j} = \sum_{i=0}^{j-1} Y_i[B(t_{i+1}) - B(t_i)]$$

and

$$Z_t = Z_{t_j} + Y_j[B(t) - B(t_j)] \text{ if } t_j \leq t \leq t_{j+1}$$

then we define

$$Z_t = \int_0^t A_s dB_s$$

Theorem 5.3. *Suppose A_t is a process with continuous paths adapted to $\mathcal{H}_t$, and suppose with probability one $|A_t| < \infty$ for all t,then there exist a sequence of simple process $A_t^{(n)}$ such that for all t*

$$\lim_{n\to\infty} \int_0^t E\left[|A_s - A_s^{(n)}|^2\right] ds = 0$$

Proof. See Lawler □

Informally, this states we can approximate any continuous process with a sequence of simple processes, and we define

$$Z_t = \lim_{n\to\infty} \int_0^t A_s^{(n)} dBs = \int_0^t A_s dBs$$

Definition 5.4. We write the stochastic differential

$$dX_t = A_t dB_t$$

to mean X_t satisfies

$$X_t = X_0 + \int_0^t A_s dB_s$$

Definition 5.5. (Quadratic Variation) If $X(t)$ is a stochastic process, the quadratic variation is defined by

$$\langle X \rangle_t = \lim_{n\to\infty} \sum_{j\leq tn} \left[X(\frac{j}{n}) - X(\frac{j-1}{n})\right]^2$$

where the sum is over all j with $\frac{j}{n} \leq t$

Theorem 5.6. *The quadratic variation of a standard Brownian Motion is t*

$$\langle B\rangle_t = t$$

Proof. Define Q_n as

$$Q_n = \sum_{j=1}^{n} \left[B(\frac{j}{n}) - B(\frac{j-1}{n})\right]^2$$

where $B(t)$ is a standard Brownian motion. And define $Y_j = Y_{j,n}$ as

$$Y_{j,n} = \left[\frac{B(\frac{j}{n}) - B(\frac{j-1}{n})}{\frac{1}{\sqrt{n}}}\right]^2$$

Notice the random variables $Y_1, ..., Y_n$ are independent random variables each with distribution of Z^2 where Z is a standard normal. In particular, $E[Y_j] = 1$
Now we can rewrite Q_n in terms of Y_j

$$Q_n = \frac{1}{n}\sum_{j=1}^{n} Y_j$$

and rewrite $\langle B\rangle_t$ as

$$\langle B\rangle_t = t \lim_{n\to\infty} Q_n$$

Since Y_j are i.i.d.random variables, by the strong law of large number

$$P\left(\lim_{n\to\infty} Q_n = E[Y_j] = 1\right) = 1$$

In other words, with probability 1,

$$\langle B\rangle_t = t$$

□

This result is often expressed as $dt = (dB)^2$

Theorem 5.7. *If W_t is a Brownian motion with drift μ and variance σ^2, then $\langle W\rangle_t = \sigma^2 t$*

Theorem 5.8. *(Ito's formula) Suppose f is a twice differentiable function and $B(t) = B_t$ is a standard Brownian motion. Then for every t*

$$f(B_t) = f(B_0) + \int_0^t f'(B_s)dB_s + \frac{1}{2}\int_0^t f''(B_t)dt$$

or, in differential form:

$$df(B_t) = f'(B_t)dB_t + \frac{1}{2}f''(B_t)dt$$

Proof. We can first look at the difference between $f(B_1)$ and $f(B_0)$

$$f(B_1) - f(B_0) = \sum_{j=1}^{n}[f(B_{j/n}) - f(B_{(j-1)/n})]$$

Using Taylor approximation, and let $\Delta_{j,n} = B_{j/n} - B_{(j-1)/n}$ we can write the summation terms as

$$f(B_{j/n}) - f(B_{(j-1)/n}) = f'(B_{(j-1)/n})\Delta_{j,n} + \frac{1}{2}f''(B_{(j-1)/n})\Delta_{j,n}^2 + o(\Delta_{j,n}^2)$$

where $o(\Delta_{j,n}^2)/\Delta_{j,n}^2 \to 0$ as $\Delta_{j,n} \to 0$
Hence $f(B_1) - f(B_0)$ is equal to sum of the following three limits

$$\lim_{n\to\infty} \sum_{j=1}^{n} f'(B_{(j-1)/n})[B_{j/n} - B_{(j-1)/n}]$$

To be continued? □

In stochastic calculus, we often throw away terms that are of smaller order than dt, which means we can simplify our calculations using these rules:

$$(dt)^2 = 0, (dB_t)(dt) = 0$$

Definition 5.9. A function is referred to as being C^n if it is $n-$times differentiable.

Theorem 5.10. *(Ito's Formula 2). Suppose $f(t,x)$ is a function that is C^1 in t and C^2 in x. If B_t is a standard Brownian motion, then*

$$f(t,B_t) = f(0,B_0) + \int_0^t \partial_x f(s,B_s)dB_s + \int_0^t \left[\partial_s f(s,B_s) + \frac{1}{2}\partial_{xx} f(s,B_s)\right]ds$$

or in differential form

$$df(t,B_t) = \partial_x f(t,B_t)dB_t + \left[\partial_t f(t,B_t) + \frac{1}{2}\partial_{xx} f(t,B_t)\right]dt$$

Proof.

$$df(t,B_t) = \partial_t f(t,B_t)dt + \partial_x f(t,B_t)dB_t + \frac{1}{2}\partial_{xx} f(t,B_t)^2(dB_t)^2 + o(dt) + o((dt)(dB_t)) + o((dB_t)^2)$$

$$df(t,B_t) = \partial_t f(t,B_t)dt + \partial_x f(t,B_t)dB_t + \frac{1}{2}\partial_{xx} f(t,B_t)^2(dt) + (0) + (0) + (0)$$

□

Theorem 5.11. *Suppose X_t satisfies*

$$dX_t = R_t dt + A_t dB_t$$

which is equivalent to

$$X_t = X_0 + \int_0^t R_s ds + \int_0^t A_s dB_s$$

where R_t and A_t are adapted process with continuous paths.

$$d\langle X\rangle_t = A_t^2 dt$$

Proof.

$$\begin{aligned}(dX_t)^2 &= (R_t dt + A_t dB_t)^2 \\ &= R^2(dt)^2 + 2R_t A_t(dt)(dB_t) + A_t^2(dB_t)^2 = A_t^2 dt\end{aligned}$$

□

Theorem 5.12. *(Ito's Formula 3). Suppose X_t satisfies $dX_t = R_t dt + A_t dB_t$ and $f(t,x)$ is C^1 in t and C^2 in x. Then*

$$\begin{aligned}df(t,X_t) &= \partial_t f(t,X_t)dt + \partial_x f(t,X_t)dX_t + \frac{1}{2}\partial_{xx} f(t,X_t)d\langle X\rangle_t \\ &= \left[\partial_t f(t,X_t) + R_t\partial_x f(t,X_t) + \frac{A_t^2}{2}\partial_{xx} f(t,X_t)\right]dt + A_t\partial_x f(t,X_t)dB_t\end{aligned}$$

References

[1] Yu-Ting Chen Babak Fotouhi Naghmeh Momeni Shing-Tung Yau Benjamin Allen, Gabor Lippner and Martin A. Nowak. Evolutionary dynamics on any population structure. 2016.

[2] Lawrence C. Evans. *An Introduction to Stochastic Differential Equations Version 1.2*.

[3] Erez Lieberman Hisashi Ohtsuki, Christoph Hauert and Martin A. Nowak. A simple rule for the evolution of cooperation on graphs and social networks. 2006.

[4] Gregory F. Lawler. *Stochastic Calculus: An Introduction with Applications*. 2014.

[5] A. Traulsen P. M. Altrock and M. A. Nowak. Evolutionary games on cycles with strong selection. 2017.