Reintroduction to Probability

James Guo

Johns Hopkins University

April 26th, 2025

Contents:

What is Probability?

Often, we use the word **probability** to describe how likely an *event* will happen.

Examples of Probability

- The probability of sun rising tomorrow is (*hopefully*) 100%.
- The probability of a fair coin flip being head is 50%.
- The probability of getting a 1 by rolling a even, 6-sided dice is $16.\overline{6}\%$.

What is Probability?

Often, we use the word **probability** to describe how likely an *event* will happen.

Examples of Probability

- The probability of sun rising tomorrow is (*hopefully*) 100%.
- The probability of a fair coin flip being head is 50%.
- The probability of getting a 1 by rolling a even, 6-sided dice is $16.\overline{6}\%$.

Here, **probability** tells us how likely something will occur, and it is a quantitative description, *i.e.*, you know it is more likely to get a head in coin flip than getting a 1 from a dice roll.

Probability of Finite Case

For finitely many equally likely cases, the probability of something happening is the number of desired cases over the number of all cases.

Example: Rolling Dices

Let's assume that we have some even, 6-sided dices.

We can compute the following probabilities:

$$\mathrm{P}(\text{rolling a even number}) =$$

Example: Rolling Dices

Let's assume that we have some even, 6-sided dices.

We can compute the following probabilities:

$$\mathrm{P}(\text{rolling a even number}) = \frac{3}{6} = \frac{1}{2}.$$

$$\mathrm{P}(\text{rolling a 1 first time and a 2 second time}) =$$

Example: Rolling Dices

Let's assume that we have some even, 6-sided dices.

We can compute the following probabilities:

$$\mathrm{P}(\text{rolling a even number}) = \frac{3}{6} = \frac{1}{2}.$$

$$\mathrm{P}(\text{rolling a 1 first time and a 2 second time}) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}.$$

Events and Joint Probability

In the previous example, we can consider *rolling a even number* or *rolling a 1 and a 2* as **events**, which represents something that we want to investigate how likely it will happen.
Typically, we denote an event with a capital letter, such as $A, B, \cdots$

Joint Probability

In particular, you should notice that the probability of rolling a 1 and a 2 can be characterized as the joint probability of two events, that is:

$$\mathrm{P}(A \wedge B) \text{ or } \mathrm{P}(A, B),$$

where A denotes the event of rolling a 1, and B denotes the event of rolling a 2.

Independent Events

For the previous example, you might observe that rolling a 1 for the first time and rolling a 2 for the second time are not related to each other, and in this case, we have the **joint probability** simply as:

$$\mathrm{P}(A, B) = \mathrm{P}(A) \times \mathrm{P}(B).$$

Independent Events

Two events A and B are **independent** if:

$$\mathrm{P}(A, B) = \mathrm{P}(A) \times \mathrm{P}(B).$$

Conditional Probability

Sometimes, we concern the probability of an event when conditioned on the certain other events.

Examples of Dependent Event

Consider a black box containing 5 red balls and 1 blue balls.

- The probability of first drawing a red ball is $5/6$ (Event A).
- The probability of first drawing a blue ball is $1/6$ (Event B).
- The probability of drawing a red ball in the second draw is dependent on what has happened:
  - Conditioned on event A, the probability to draw a blue ball is $1/5$,
  - Conditioned on event B, the probability to draw a blue ball 0.

Notation-wise, we write the probability of an event B conditioned on event A as:

$$\mathrm{P}(B \mid A).$$

Bayes' Law

For conditional probability, we can write it in terms of joint probability:

$$\mathrm{P}(B \mid A) = \frac{\mathrm{P}(B, A)}{\mathrm{P}(A)}.$$

Bayes' Law

For conditional probability, we can write it in terms of joint probability:

$$\mathrm{P}(B \mid A) = \frac{\mathrm{P}(B, A)}{\mathrm{P}(A)}.$$

By some simple arithmetic deductions, we can get Bayes' Law:

Bayes' Law

Given events A and B, their conditional probability satisfies that:

$$\mathrm{P}(A \mid B) \cdot \mathrm{P}(B) = \mathrm{P}(B \mid A) \cdot \mathrm{P}(A).$$

Here, we invite our audiences to attempt to justify Bayes' Law by themselves.

Why Probability Theory?

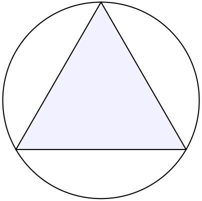
Isn't the previous "probability" good, why **probability theory**?

Some Uncovered Issues with Previous Probability

- We defined how to compute probability in finite number of events, what if there are **infinitely many** events (countable or even uncountable).
- If we can manually assign probability to certain events, what are the requirements of assigning a legit **probability distribution**.
- How can we model something that is **completely random**?

Bertrand's Paradox

Before getting into the theory, let's take a look at another problem called the **Bertrand's Paradox**.
Consider an equilateral triangle inscribed in a circle.

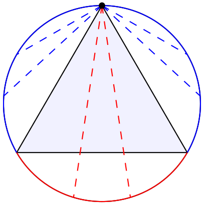


Now, suppose that we are picking a chord, *randomly*, on the circle, what is the probability that the selected chord is longer than the side length of the equilateral triangle?

Bertrand's Paradox (Continued)

Here, we consider the problem in three approaches (Denoted event A).

Random Endpoints: Have one endpoint of the chord fixed, and have the other endpoint free on the circle.

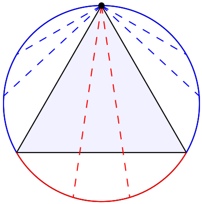


$$\mathrm{P}(A) = \frac{1}{3}.$$

Bertrand's Paradox (Continued)

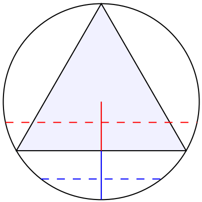
Here, we consider the problem in three approaches (Denoted event A).

Random Endpoints: Have one endpoint of the chord fixed, and have the other endpoint free on the circle.



$$\mathrm{P}(A) = \frac{1}{3}.$$

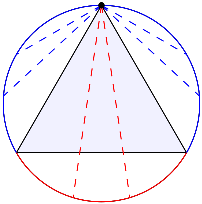
Random Radial Point: Fix a radius of the circle, and look for the chords perpendicular to that radius.



$$\mathrm{P}(A) = \frac{1}{2}.$$

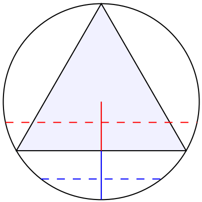
Here, we consider the problem in three approaches (Denoted event A).

Random Endpoints: Have one endpoint of the chord fixed, and have the other endpoint free on the circle.



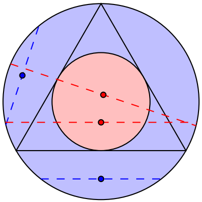
$$\mathrm{P}(A) = \frac{1}{3}.$$

Random Radial Point: Fix a radius of the circle, and look for the chords perpendicular to that radius.



$$\mathrm{P}(A) = \frac{1}{2}.$$

Random Midpoint: Chord is longer if and only if midpoint lies on the inscribed circle of the equilateral triangle.



$$\mathrm{P}(A) = \frac{1}{4}.$$

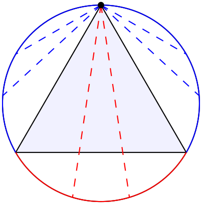
Bertrand's Paradox (Continued)

Here, we consider the problem in three approaches (Denoted event A).

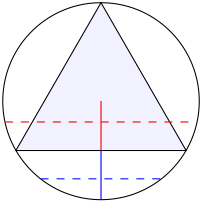
Random Endpoints: Have one endpoint of the chord fixed, and have the other endpoint free on the circle.

Random Radial Point: Fix a radius of the circle, and look for the chords perpendicular to that radius.

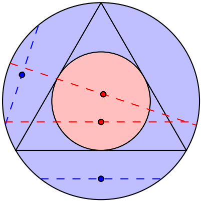
Random Midpoint: Chord is longer if and only if midpoint lies on the inscribed circle of the equilateral triangle.



$$\mathrm{P}(A) = \frac{1}{3}.$$



$$\mathrm{P}(A) = \frac{1}{2}.$$



$$\mathrm{P}(A) = \frac{1}{4}.$$

Here, the three methodologies give distinct results since the "*randomness*" are defined differently, *i.e.*, the distribution is not at random in each case with respect to the other ones.

Set Theoretics

Before defining **probability** rigorously, we get to be familiar with **sets**.

Set

A **set** is a collection of things (called **elements**).

Set Theoretics

Before defining **probability** rigorously, we get to be familiar with **sets**.

Set

A **set** is a collection of things (called **elements**).

For the sake of this presentation, we need to use the following notations:

- **Set building notation**: We may use $\{\ldots\}$ to represent a set. *E.g.* $S = \{1, 2, 3\}$.
  If a set contains nothing, it is empty set ($\emptyset$).
- **Belongs to**: We use $\in$ to represents an element in a set. *E.g.* $1 \in S$. If an element is not in the set, we use $\notin$. *E.g.* $0 \in S$.
- **Subset notation**: We use $\subset$ to represent that a set A is a subset of another set S, *i.e.*, all the elements in A belongs to S.
- **Union**: The union (denoted $\cup$) of two sets contains all the elements in any of the two sets. *E.g.* $\{1, 2, 3\} \cup \{3, 4, 5\} = \{1, 2, 3, 4, 5\}$.

Set Theoretics (Continued)

- **Intersection**: The intersection (denoted $\cap$) of two sets contains the elements that are in both two sets. *E.g.* $\{1, 2, 3\} \cap \{3, 4, 5\} = \{3\}$.
- **Set Minus**: For sets A, B, $A \setminus B := \{a \in A : a \notin B\}$.

Set Theoretics (Continued)

- **Intersection**: The intersection (denoted $\cap$) of two sets contains the elements that are in both two sets. *E.g.* $\{1,2,3\} \cap \{3,4,5\} = \{3\}$.
- **Set Minus**: For sets A, B, $A \setminus B := \{a \in A : a \notin B\}$.

Of course, the set theory is more complicated with the **ZFC axioms**, so it avoids some paradoxes.

Barber Paradox

A barber claims to shave everyone who does not shave themselves, or:

$$S := \{T : T \notin T\},$$

where $T \notin T$ means that T does not belongs to itself.

An example of set containing sets is the **power set**.

Power Set

A **power set** of S, denoted $\mathcal{P}(S)$, is the set of all subsets of S.

σ-Algebra

Then, we need just one more component to define **probability** rigorously.

σ-algebra

Let S be a set, a σ-algebra $\mathcal{F}$ on S is a set of subsets of S such that:

1. $\emptyset \in \mathcal{F}$,
2. For any $F \in \mathcal{F}$, we have $F^c \in \mathcal{F}$, where $F^c = S \setminus F$, and
3. For any $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$, we have $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

σ-Algebra

Then, we need just one more component to define **probability** rigorously.

σ-algebra

Let S be a set, a σ-algebra $\mathcal{F}$ on S is a set of subsets of S such that:

1. $\emptyset \in \mathcal{F}$,
2. For any $F \in \mathcal{F}$, we have $F^c \in \mathcal{F}$, where $F^c = S \setminus F$, and
3. For any $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$, we have $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Here are some quick observations:

- $S \in \mathcal{F}$ by (1) and (2).
- $\mathcal{P}(S)$ is a σ-algebra of S.

Probability Measure

Now, we define **probability** as a **measure**.

Probability Measure Space

The pair $(\Omega, \mathcal{F})$ of σ-algebra together with a probability measure $\mathbb{P} : \mathcal{F} \to [0, 1]$ forms a **probability measure space**, where $\mathbb{P}$ satisfies the following properties:

1. $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega) = 1$,
2. For any $A, B \in \mathcal{F}$ such that $A \subset B$, $\mathbb{P}(A) \leq \mathbb{P}(B)$, and
3. For any $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$, we have:

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Probability Measure

When we look into the definition, it is reasonable, or we can say that it aligns with our intuition with probability.

1. The probability of nothing happening is 0, the probability of anything happening is 1.
2. If the desired event list contains another list of events, it should not be less likely to happen.
3. This is a more general version of (2), that is, we can obtain (2) from this condition.

Probability Measure

When we look into the definition, it is reasonable, or we can say that it aligns with our intuition with probability.

1. The probability of nothing happening is 0, the probability of anything happening is 1.
2. If the desired event list contains another list of events, it should not be less likely to happen.
3. This is a more general version of (2), that is, we can obtain (2) from this condition.

Carathéodory Measurable Sets

Technically, we can lift the requirement for a σ-algebra $\mathcal{F}$ and still define an **exterior measure**, where $\mathcal{F}$ will be the smallest σ-algebra of Ω.

Random Variables

In elementary probability, you would be told that a **random variable** is a map to an outcome (typically numbers) with different probabilities.

Rolling a fair 6-sided Dice

We can let X denote the outcome of the dice, so:

$$\mathrm{P}(X = i) = \begin{cases} \frac{1}{6} & \text{if } i = 1, 2, 3, 4, 5, 6; \\ 0 & \text{otherwise.} \end{cases}$$

Random Variables

In elementary probability, you would be told that a **random variable** is a map to an outcome (typically numbers) with different probabilities.

Rolling a fair 6-sided Dice

We can let X denote the outcome of the dice, so:

$$\mathrm{P}(X = i) = \begin{cases} \frac{1}{6} & \text{if } i = 1, 2, 3, 4, 5, 6; \\ 0 & \text{otherwise.} \end{cases}$$

Since elementary probability often assumes the outcome as a real number, which in the formal language, the output σ-algebra pair is:

$$(\mathbb{R}, \mathcal{B}),$$

where $\mathcal{B}$ is the Borel set, or the σ-algebra containing all open intervals.

Random Variables (Continued)

But, we are not satisfied as we have not really give a definition of what is random variable, and what is not a random variable.

Random Variables

For probability measure spaces $(\Omega, \mathcal{F}, \mathbb{P})$ and σ-algebra pair $(E, \mathcal{E})$, a random variable is a function $X : \Omega \to E$ such that for any $F \in \mathcal{E}$:

$$X^{-1}(S) := \{x \in \Omega : X(x) \in S\} \in \mathcal{F}.$$

That is, the preimage of all sets in the σ-algebra of E is in the σ-algebra of the probability measure space.

Random Variables (Continued)

But, we are not satisfied as we have not really give a definition of what is random variable, and what is not a random variable.

Random Variables

For probability measure spaces $(\Omega, \mathcal{F}, \mathbb{P})$ and σ-algebra pair $(E, \mathcal{E})$, a random variable is a function $X : \Omega \to E$ such that for any $F \in \mathcal{E}$:

$$X^{-1}(S) := \{x \in \Omega : X(x) \in S\} \in \mathcal{F}.$$

That is, the preimage of all sets in the σ-algebra of E is in the σ-algebra of the probability measure space.

Here, we remark on the following:

- The σ-algebra is often called the **measurable** sets, so the function has preimage of measurable sets still measurable, and such functions are called **measurable** functions.
- The probability extends naturally as $\mathbb{P}(X \in S) = \mathbb{P}\big(X^{-1}(S)\big)$.

Example: Bernoulli Distribution

A common known random variable is the Bernoulli Distribution, which calculates the probability of success $p \in [0, 1]$. Often, we denote:

$$X \sim \texttt{Bernoulli}(p).$$

Example: Bernoulli Distribution

A common known random variable is the Bernoulli Distribution, which calculates the probability of success $p \in [0, 1]$. Often, we denote:

$$X \sim \texttt{Bernoulli}(p).$$

Particularly, lets more rigidly define X as a **random variable**. Here, our σ-algebra pair can be:

$$\big(\{\text{Success}, \text{Fail}\}, \big\{\emptyset, \{\text{Success}\}, \{\text{Fail}\}, \{\text{Success}, \text{Fail}\}\big\}\big).$$

For simplicity, we define our probability measure space trivially:

$$\big(\{S, F\}, \big\{\emptyset, \{S\}, \{F\}, \{S, F\}\big\}, \mathbb{P}\big),$$

where we have $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\{S\}) = p$, $\mathbb{P}(\{F\}) = 1 - p$, and $\mathbb{P}(\{S, F\}) = 1$. (Check by yourself that this is a measure space.)
In this way, X is a random variable, and the extension of the probability naturally aligns with the probability of the Bernoulli distribution.

Example: Normal Distribution

Another important distribution in statistics is the normal distribution, based on parameters $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_{\geq 0}$. We denote these distribution as:

$$X \sim \texttt{Normal}(\mu, \sigma^2).$$

Here we can consider the probability density function:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Example: Normal Distribution

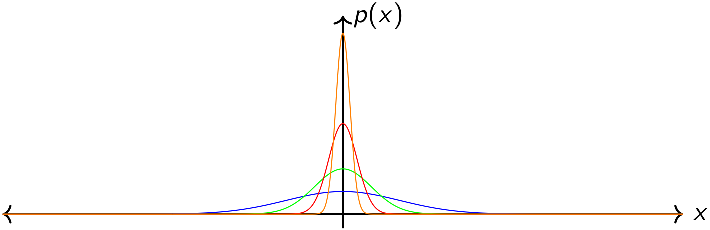
Another important distribution in statistics is the normal distribution, based on parameters $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_{\geq 0}$. We denote these distribution as:

$$X \sim \texttt{Normal}(\mu, \sigma^2).$$

Here we can consider the probability density function:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Here, are some illustrations of the density function for $\mu = 0$ and $\sigma = 1$ (blue), $\sqrt{2}/2$ (green), $1/2$ (red), $\sqrt{2}/4$ (orange).



Example: Normal Distribution (Continued)

Here, our σ-algebra pair is $\left(\mathbb{R}, \mathcal{B}\right)$, and we define our probability measure space as $\left(\mathbb{R}, \mathcal{B}, \mu\right)$, where the measure $\mu : \mathcal{B} \to \mathbb{R}$ is defined as:

$$\mu(S) = \int_S p(x)dx.$$

Example: Normal Distribution (Continued)

Here, our σ-algebra pair is $(\mathbb{R}, \mathcal{B})$, and we define our probability measure space as $(\mathbb{R}, \mathcal{B}, \mu)$, where the measure $\mu : \mathcal{B} \to \mathbb{R}$ is defined as:

$$\mu(S) = \int_S p(x)dx.$$

Lebesgue Measure? Calculus?

The integral is Lebesgue integral, but for simplicity, consider $S = [a, b]$:

- If you have learned calculus, you can simply consider it as:

$$\mu([a, b]) = \int_a^b p(x)dx.$$

- If you have not learned any calculus, consider it as the area under the curve from a to b on the previous figure.

We leave the check that this is a **random variable** to diligent audiences.

Conditional Expectation

Now, we can define **conditional expectation**, which is another expectation.

Conditional Expectation

Given two random variables X and Y, we can define the conditional expectation of X on Y as $\mathbb{E}[X \mid Y]$, which is another random variable that satisfied:

$$\int_A X d\mathbb{P} = \int_A \mathbb{E}[X \mid Y] d\mathbb{P}$$

for all A as the σ-algebra in the random variable Y.

Conditional Expectation

Now, we can define **conditional expectation**, which is another expectation.

Conditional Expectation

Given two random variables X and Y, we can define the conditional expectation of X on Y as $\mathbb{E}[X \mid Y]$, which is another random variable that satisfied:

$$\int_A X d\mathbb{P} = \int_A \mathbb{E}[X \mid Y] d\mathbb{P}$$

for all A as the σ-algebra in the random variable Y.

In fact, you should observe that Y does not matter as a random variable. The key characterization shall be the σ-algebra associated in Y, so conditional expectation is often written as:

$$\mathbb{E}[X \mid \mathcal{H}_Y].$$

Martingale

For the sake of depth, we will just discuss the case for a discrete martingale.

Discrete Martingale

Let $\{X_j\}_{j=1}^{\infty}$ be random variables such that $\mathbb{E}[|X_j|] < \infty$. The the sequence $\{X_j\}_{j=1}^{\infty}$ is discrete martingale if $X_k = \mathbb{E}[X_j \mid X_1, \cdots, X_k] = \mathbb{E}[X_j \mid \mathcal{F}_k]$ for almost all $x \in X_j$ for all $j \geq k$.

Martingale

For the sake of depth, we will just discuss the case for a discrete martingale.

Discrete Martingale

Let $\{X_j\}_{j=1}^{\infty}$ be random variables such that $\mathbb{E}[|X_j|] < \infty$. The the sequence $\{X_j\}_{j=1}^{\infty}$ is discrete martingale if $X_k = \mathbb{E}[X_j \mid X_1, \cdots, X_k] = \mathbb{E}[X_j \mid \mathcal{F}_k]$ for almost all $x \in X_j$ for all $j \geq k$.

In one sentence, we can interpret martingale as:

The best expectation of a future instance is the current.

Application in the Stock Market

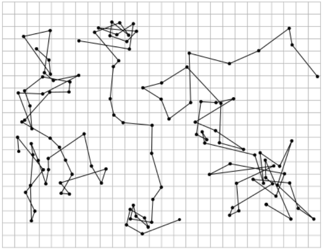
In fact, the stock market follows the martingale style, since the most accurate expectation of the future can only depend on what all we have currently.

Brownian Motion

Now, we have basically *and quickly* gone through the core concepts of **probability theory** while omitting a lot of details. Here, we consider one important feature as the **Brownian motion**.

Brownian Motion

Invented by Robert Brown, Brownian motion is used to model the motion of a particle's motion in fluid, which can be considered random.



One-dimensional Brownian Motion

The one-dimensional Brownian motion can be characterized as follows:

One-dimensional Brownian Motion

Let X be a random process on probability space $(\Omega, \mathcal{F}, \mathbb{P})$. It is a Brownian motion if X satisfies that:

1. With probability 1, $X(0) = 0$ and $X(t)$ is a continuous function of t.
2. For all $t \geq 0$ and $h > 0$, the increment $X(t+h) - X(t) \sim \texttt{Normal}(0, h)$, that is:
$$\mathbb{P}\big(X(t+h) - X(t) \leq x\big) = (2\pi h)^{-1/2} \int_{-\infty}^{x} \exp\left(-\frac{s^2}{2h}\right) ds,$$
3. If $0 \leq t_1 \leq t_2 \leq \cdots \leq t_{2m}$, the increments $X(t_2) - X(t_1)$, $X(t_4) - X(t_3)$, $\cdots$, $X(t_{2m}) - X(t_{2m-1})$ are independent.

One-dimensional Brownian Motion (Continued)

Simulation of Brownian Motion

Brownian motion is an abstract concept, but it can be simulated.

One-dimensional Brownian Motion (Continued)

Simulation of Brownian Motion

Brownian motion is an abstract concept, but it can be simulated.

- **Random walk approximation**. Values of 1 and -1 are assigned by randomness to Y_i for $1 \leq i \leq m$, where m is large and τ is small.

One-dimensional Brownian Motion (Continued)

Simulation of Brownian Motion

Brownian motion is an abstract concept, but it can be simulated.

- **Random walk approximation**. Values of 1 and -1 are assigned by randomness to Y_i for $1 \leq i \leq m$, where m is large and τ is small.
- **Random midpoint displacement**. Consider function $X : [0,1] \to \mathbb{R}$, we define the values of $X(k2^{-j})$ where $0 \leq k \leq 2^j$ by induction on j. We first choose $X(0) = 0$, then set $X(1)$ as a random number of $\texttt{Normal}(0,1)$, then choosing $X(1/2)$ as a random number of $\texttt{Normal}\big((X(0) + X(1))/2\big)$, etc.

Simulation of Brownian Motion

Brownian motion is an abstract concept, but it can be simulated.

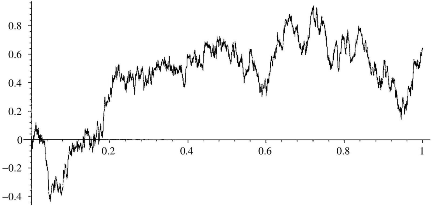
- **Random walk approximation**. Values of 1 and -1 are assigned by randomness to Y_i for $1 \leq i \leq m$, where m is large and τ is small.
- **Random midpoint displacement**. Consider function $X : [0, 1] \to \mathbb{R}$, we define the values of $X(k2^{-j})$ where $0 \leq k \leq 2^j$ by induction on j. We first choose $X(0) = 0$, then set $X(1)$ as a random number of $\texttt{Normal}(0, 1)$, then choosing $X(1/2)$ as a random number of $\texttt{Normal}\big((X(0) + X(1))/2\big)$, etc.
- **Fourier series representation**. Brownian motion on $[0, \pi]$ can be:

$$X(t) = \frac{1}{\sqrt{\pi}} C_0 t + \sqrt{\frac{2}{\pi}} \sum_{k=1}^{\infty} C_k \frac{\sin kt}{k},$$

where $C_k \sim \texttt{Normal}(0, 1)$ as independent models.

One-dimensional Brownian Motion (Continued)

An example of a one-dimensional Brownian motion simulation can be viewed as follows:



Also, you may observe the some remarks on the Brownian motion.

- The Brownian motion is not differentiable almost anywhere.
- The graph of 1-dimensional Brownian motion has Hausdorff dimension of 1.5.

Hausdorff dimension is defined using **Hausdorff measure**, if you do not know it, you can interpret this as an **extension to dimension** that aligns with your conventional intuitions on dimensions.

Kolmogrov's Continuity Theorem

The Brownian motion is particularly important **extension theorem**.

Kolmogrov's Continuity Theorem

Suppose a process $X = \{X_t\}_{t \geq 0}$ satisfies that for all $T > 0$, there exists α, β, D such that:

$$\mathbb{E}[|X_t - X_s|^\alpha] \leq D \cdot |t - s|^{1+\beta} \text{ for } 0 \leq s, t \leq T.$$

Then there exists a continuous version of X.

Does this apply to the Brownian motion?

Kolmogrov's Continuity Theorem

The Brownian motion is particularly important **extension theorem**.

Kolmogrov's Continuity Theorem

Suppose a process $X = \{X_t\}_{t \geq 0}$ satisfies that for all $T > 0$, there exists α, β, D such that:

$$\mathbb{E}[|X_t - X_s|^\alpha] \leq D \cdot |t-s|^{1+\beta} \text{ for } 0 \leq s, t \leq T.$$

Then there exists a continuous version of X.

Does this apply to the Brownian motion? Yes.
Note that with Brownian motion, we have:

$$\mathbb{E}[|B_t - B_s|^4] = n(n+2)|t-s|^2,$$

so with this characterization, the there always exists a continuous version of the Brownian motion.

Multidimensional Brownian Motion

What about extending the Brownian motion to multiple dimensions?

Multidimensional Brownian Motion

For higher dimensions, we defined the Brownian motion in $\mathbb{R}^n$ so that the coordinate components are independent 1-dimensional Brownian motions, that is:

$$B : [0, \infty) \to \mathbb{R}^n \text{ as } x \mapsto \big(B_1(t), \cdots, B_n(t)\big),$$

where $B_1, \cdots, B_n$ are one-dimensional Brownian motion.

Multidimensional Brownian Motion

What about extending the Brownian motion to multiple dimensions?

Multidimensional Brownian Motion

For higher dimensions, we defined the Brownian motion in $\mathbb{R}^n$ so that the coordinate components are independent 1-dimensional Brownian motions, that is:

$$B : [0, \infty) \to \mathbb{R}^n \text{ as } x \mapsto \big(B_1(t), \cdots, B_n(t)\big),$$

where $B_1, \cdots, B_n$ are one-dimensional Brownian motion.

- Particularly, when you have a multidimensional Brownian motion and you project it towards any dimension, it will still be a Brownian motion.
- For all higher dimensional Brownian motion graphs, their Hausdorff dimension is 2.

Stochastic Differential Equations

Stochastic differential equations can be characterized via Brownian motion, a typical formulation can be written as:

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dB_t.$$

When explicitly, it can be written as:

$$X_t = X_0 + \int_0^t \mu(X_t, t)dt + \int_0^t \sigma(X_t, t)dB_t.$$

Stochastic Differential Equations

Stochastic differential equations can be characterized via Brownian motion, a typical formulation can be written as:

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dB_t.$$

When explicitly, it can be written as:

$$X_t = X_0 + \int_0^t \mu(X_t, t)dt + \int_0^t \sigma(X_t, t)dB_t.$$

Particularly, some terms can be derived using the Itô formula.

Itô Lemma

Let X_t be a Itô process, and $dX_t = udt + vdB_t$. Let $g(t, x)$ be twice differentiable on $[0, \infty) \times \mathbb{R}$, then for any $Y_{t(\omega)} = g(t, X_{t(\omega)})$, it is a Itô process and:

$$dY_t = \frac{\partial g}{\partial t}(t, X_t)dt + \frac{\partial g}{\partial x}(t, X_t)dx_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t, X_t)(dX_t)^2,$$

in which $dt \cdot dt = dt \cdot dB_t = 0$ and $dB_t \cdot dB_t = dt$.

Example: Machine Learning

Another popular field of SDEs application is in machine learning. A very popular method of finding minimum is called **gradient descent**, namely updating the parameter θ via the loss function ℓ:

$$\theta_{t+1} = \theta_t - \alpha \nabla \ell(\theta_t; \mathcal{D}),$$

where $\mathcal{D}$ is the full data set, and α is the learning rate.

Example: Machine Learning

Another popular field of SDEs application is in machine learning. A very popular method of finding minimum is called **gradient descent**, namely updating the parameter θ via the loss function ℓ:

$$\theta_{t+1} = \theta_t - \alpha \nabla \ell(\theta_t; \mathcal{D}),$$

where $\mathcal{D}$ is the full data set, and α is the learning rate.

Stochastic Gradient Descent

When the amount of data is too large, we typically use a (random) smaller subset of the data to compute the gradient, called a **batch** and denoted $\mathcal{B} \subset \mathcal{D}$, so the θ update becomes:

$$\theta_{t+1} = \theta_t - \alpha \nabla \ell(\theta_t; \mathcal{B}).$$

When α is arbitrary small, we can model the learning problem as a SDE:

$$d\theta_t = -\nabla \underbrace{\mathbb{E}[\ell(\theta_t; \mathcal{B})]}_{\text{deterministic}} dt + \underbrace{\sqrt{2\sigma(\theta_t)}}_{\text{drift}} dB_t.$$

Example: Finance Model

A very direct application is with financial models.

- The market is modeled by a $(n+1)$-dimensional process $\big(X_0(t), X_1(t), \cdots, X_n(t)\big)$ that follows:

$$dX_0 = \rho(t,\omega)X_0(t)dt \text{ and } X_0(0) \equiv 1,$$

$$dX_i = \mu_i(t,\omega)dt + \sum_{j=1}^{m} \sigma_{i,j}(t,\omega)dB_j \text{ for } 1 \le i \le n.$$

- The market is **normalized** if $X_0(t) \equiv 1$.

Example: Finance Model

A very direct application is with financial models.

- The market is modeled by a $(n+1)$-dimensional process $\big(X_0(t), X_1(t), \cdots, X_n(t)\big)$ that follows:
$$dX_0 = \rho(t,\omega)X_0(t)dt \text{ and } X_0(0) \equiv 1,$$
$$dX_i = \mu_i(t,\omega)dt + \sum_{j=1}^{m} \sigma_{i,j}(t,\omega)dB_j \text{ for } 1 \leq i \leq n.$$
- The market is **normalized** if $X_0(t) \equiv 1$.
- The **portfolio** is an $(n+1)$-dimensional (t,ω)-measurable and adapted stochastic process:
$$\theta(t) = \big(\theta_0(t), \theta_1(t), \cdots, \theta_n(t)\big).$$
- The **value** at time t of a portfolio $\theta(t)$ is defined by:
$$V(t,\omega) = \theta(t) \cdot X(t) = \sum_{i=0}^{n} \theta_i(t)X_i(t).$$

Example: Finance Model (Continued)

Self-financing Portfolio

A portfolio $\theta(t)$ is self-financing if:

$$\int_0^T \left[\left| \theta_0(s)\rho(s)X_0(s) + \sum_{i=1}^n \theta_i(s)\mu_i(s) \right| + \sum_{j=1}^m \left(\sum_{i=1}^n \theta_i(s)\sigma_{i,j}(s) \right)^2 \right] ds$$

is finite almost surely and satisfied the Itô relation:

$$V(t) = V(0) + \int_0^t \theta(s) \cdot dX(s) \text{ for } t \in [0, T].$$

For more information of this topic, please check on Chapter 12 of Bernt Øksendal's book Stochastic Differential Equations.

This is end of this talk.

We hope that this reintroduction to probability might have innovate you through bases of probability theory.

Thank you for listening and being here at JHMT-2025 with us today.