Stochastic Calculus: Lecture on Probability

Fall 2025

Probability - Sample space

- In probability, the key idea is that of a random experiment.
- Here, sample space Ω collects all possible outcomes in an experiment.
- Example: In experiment of rolling a dice, $\Omega = \{1, 2, 3, 4, 5, 6\}$
- In experiment of flipping coins infinitely many times Ω_B consists of all $\omega = (\omega_1, \omega_2, \ldots)$ where each $\omega_i \in \{H, T\}$.

Assigning probability to subsets

- We need an idea of collection of subsets $\mathcal{F}$ of Ω where to each subset probability P can be assigned.
- Surprisingly, not possible to assign probabilities to each and every subset of Ω unless you limit yourself to countably many outcomes.
- A collection of subsets on Ω is an algebra if

i) $\Omega \in \mathcal{F}$

ii) If $A \in \mathcal{F}, \Rightarrow A^c \in \mathcal{F} \quad (\forall A \in \mathcal{F})$

iii) Let $A_1, A_2, \ldots, A_n \in \mathcal{F}$, then $\bigcup_{i=1}^{n} A_i \in \mathcal{F}$.

Assigning probability to algebras is usually easy

- Consider again coin toss space Ω_B. For each ν_i either H or T, consider subsets of form

$$A_n(\nu_1, \nu_2, \ldots, \nu_n) = \{\omega \in \Omega_B : (\omega_1, \omega_2, \ldots, \omega_n) = (\nu_1, \nu_2, \ldots, \nu_n)\}$$

- There are 2^n such subsets or atoms, and 2^{2^n} possible subsets formed by their power set (sets formed by all possible combination of these atoms, including the empty set). Call that collection $\mathcal{F}_n$. Check that $\mathcal{F}_n$ is an algebra.

- And so is $\mathcal{F} = \cup_{n=1}^{\infty} \mathcal{F}_n$!

- Easy to assign probability to $\mathcal{F}$ by assigning probability to each set A_n. E.g., for $p \in (0, 1)$

$$P(A_n) = p^{\#\nu_i = H}(1 - p)^{\#\nu_i = T}.$$

Algebra on $[0, 1)$

- Similarly, for $\Omega = [0, 1)$ it is easy to assign probabilities to sets of the form

$$A = [a_1, b_1) \cup [a_2, b_2) \ldots [a_n, b_n)$$

  where $0 \leq a_1 < b_1 \leq a_2 \ldots < b_n \leq 1$.

- Collect all such sets for each n and call that $\mathcal{F}$.
- $\mathcal{F}$ is an algebra. Check!
- Easy to assign measure to this algebra. Example, Lebesgue measure measures length $\mu(A) = \sum_{i=1}^{n}(b_i - a_i)$.
- But such algebras contain very few subsets to be interesting

We need σ-algebras!

- A σ-algebra $\mathcal{F}$ defined on Ω has to satisfy the following conditions:
  i) $\Omega \in \mathcal{F}$
  ii) If $A \in \mathcal{F}, \Rightarrow A^c \in \mathcal{F} \quad (\forall A \in \mathcal{F})$
  iii) Let $A_1, A_2, \dots, A_n, \dots \in \mathcal{F} \quad \forall i \neq j$, then $\bigcup_i A_i \in \mathcal{F}$.

Thus by construction, a σ-algebra on sample space Ω will contain Ω itself, complements of every set that belongs to it, and is closed under countable union of sets in it.

Note

An algebra $\mathcal{F}'$ on Ω satisfies (i), (ii), and is closed under *finite* union of its members. Every algebra is a σ-algebra, not the other way around

Some examples of σ-algebras

- Given any Ω, $\mathcal{F}_0 = \{\Omega, \emptyset\}$ is the smallest σ-algebra on Ω.
- Let $\Omega = \{\omega : \omega = \omega_1, \omega_2, \omega_3 \ldots$ where $\omega_i \in \{0, 1\}\}$. Thus, Ω is a collection of infinite sequence of 0's or 1's.
- Define $B_1 = \{\omega \in \Omega : \omega_1 = 1\}$, and $B_2 = \{\omega \in \Omega : \omega_1 = 0\}$.
- Easy to verify that $\mathcal{F}_1 = \{\emptyset, \Omega, B_1, B_2\}$ is a σ-algebra on Ω.
- Since $\Omega \in \mathcal{F}_1$. Given any set in $\mathcal{F}_1$, its complement is in $\mathcal{F}_1$. Given $A_1, A_2, \ldots$ in $\mathcal{F}_1$, we have $\cup_i A_i \in \mathcal{F}_1$.

Algebras and σ algebras

- Consider again coin toss space Ω_B. For each ν_i either H or T,

  $$A_n(\nu_1, \nu_2, \ldots, \nu_n) = \{\omega \in \Omega_B : (\omega_1, \omega_2, \ldots, \omega_n) = (\nu_1, \nu_2, \ldots, \nu_n)\}$$

- $\mathcal{F}_n$ denotes collection of all subsets formed by combining these subsets in all possible ways. Check that $\mathcal{F}_n$ is a σ algebra.

- But $\mathcal{F} = \cup_{n=1}^{\infty} \mathcal{F}_n$ is not!

- Is $\{H, H, H, H \ldots\} \in \mathcal{F}$?. Consider the set $\cap_{n=1}^{\infty} A_n(H, H, \ldots, H)$ (here H occcurs n times).

Creating σ-algebras from algebras

- Given S a collection of subsets of Ω, denote by $\sigma(S)$ to be the smallest σ-algebra that contains S.
- Above $\{H, H, H, H \ldots\} \in \sigma(\mathcal{F})$.
- Borel sigma algebra $\mathcal{B}([0, 1])$: Smallest sigma algebra that contains all open intervals in $[0, 1]$. Equivalently, open sets.
- Similarly, define $\mathcal{B}(\Re)$.

$\mathcal{B}(\Re)$

- It contains all closed intervals $[a, b]$ for $a \leq b$. Since for all n large enough it contains $\cap_n (a - 1/n, b + 1/n)$. This equals $[a, b]$.
- Hence, it contains the set of any rational numbers and set of all irrational numbers.
- It contains half lines $(-\infty, x]$ for each x. Similarly, $[x, \infty)$ for each x.

Probability Measure P and probability triplet $(\Omega, \mathcal{F}, P)$

- On a measurable space $(\Omega, \mathcal{F})$, set function P is referrred to as a probability measure if

a) $P(\Omega) = 1$

b) $\forall A \in \mathcal{F},\ P(A) \geq 0$

c) If $A_1, A_2, \cdots \in \mathcal{F}$, $A_i \cap A_j = \emptyset\ \forall i \neq j$, then

$$P(\bigcup_i A_i) = \sum_i P(A_i).$$

- **Big theorem (Cartheodory): A probability measure defined on an algebra S uniquely extends to the σ-algebra $\sigma(S)$.**

Law of Large Numbers (SLLN)

- Consider $(\Omega_B, \mathcal{F}, P)$ where for $p \in (0,1)$, H_n denotes number of heads amongst $(\omega_1, \omega_2, \ldots, \omega_n)$, for all A_n

$$P(A_n) = p^{H_n}(1-p)^{n-H_n}$$

- **Weak law of large numbers**: $P(\frac{H_n}{n} > \epsilon) \to 0$ as $n \to \infty$ for any $\epsilon > 0$.

- **Strong Law of Large Numbers**:

$$P\left(\left\{\omega : \frac{H_n(\omega)}{n} \to p\right\}\right) = 1.$$

Measures on σ algebras

- **Home work: Show that if $A = \left\{\omega : \frac{H_n(\omega)}{n} \to p\right\}$ and**

$$B = \bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \left\{\omega : \left|\frac{H_n(\omega)}{n} - p\right| \leq \frac{1}{m}\right\},$$

then $A = B$.

- For $\Omega = [0,1]$ and σ-algebra $\mathcal{B}([0,1])$, probability measure that assigns probability $b - a$ to every interval $[a,b]$ $(0 \leq a \leq b \leq 1)$ corresponds to the Lebesgue measure.

- Q: What is the Lebesgue measure assigned to the set of rationals in $[0,1]$?

Random Variables

- A random variable X defined on Ω is a map of $\Omega \to \mathbb{R}$.
- **Link to** $(\Omega, \mathcal{F}, P)$: In detail, a r.v. X on $(\Omega, \mathcal{F}, P)$ is $X : \Omega \to \mathbb{R}$, such that

$$P(a \leq X \leq b) = P(\{\omega : a \leq X(\omega) \leq b\})$$

  is well-defined $\forall [a, b]$.

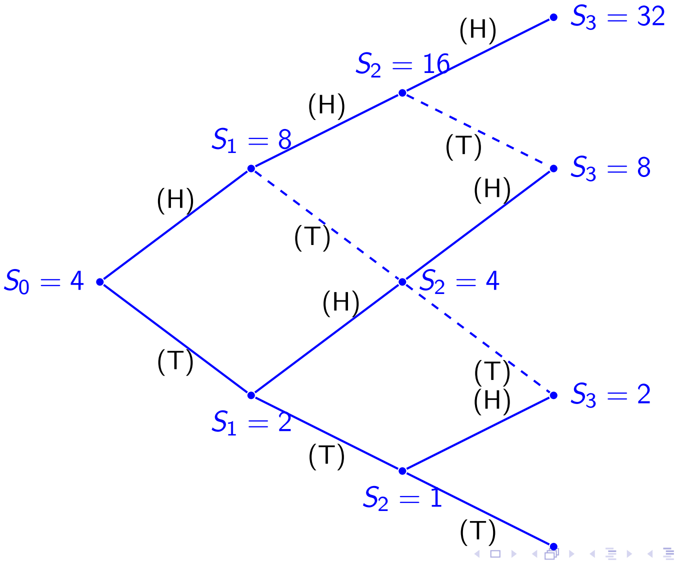
- **More generally**, X is a measurable function on $(\Omega, \mathcal{F})$ so that

$$\{\omega : X(\omega) \in B\} \in \mathcal{F}$$

  for all Borel measurable sets B. That is for all $B \in \mathcal{B}(\Re)$.

Random Variables: Example

Let $\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$.
Consider RV S_0, S_1, S_2 and S_3.



Induced Measure and Distribution Function

- A random variable X on a probability space $(\Omega, \mathcal{F}, P)$ maps the probability defined on the underlying space Ω to a probability on the sets of real line.

- Probability on a set in Ω, $\{\omega : X(\omega) \in [a, b]\}$ is assigned to a set $[a, b]$ on real line through relation

$$P(\{\omega : X(\omega) \in [a, b]\}) = \mu_X([a, b]).$$

  μ_X is the induced probability measure by r.v. X on $\mathbb{R}$.

- Define the distribution function of rv X as $F_X(x) = \mu_X((-\infty, x])$. This is a non-decreasing function such that $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

- The definition of measure μ_X can be extended beyond intervals in $\mathbb{R}$ to the space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Expectation of a Random Variable

- **Expectation**: Let's start with a simple setup where X is a r.v. assuming discrete values $\{x_1, x_2, \ldots\}$. Then

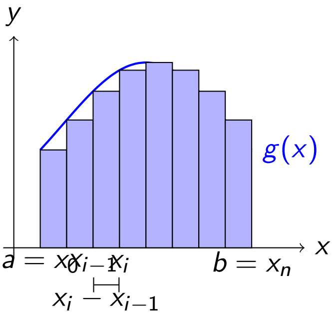
$$E[X] = \sum_k x_k P(X = x_k)$$

- For continuous random variables X, probabilities are defined through a non-negative probability density function $f_X(x) \geq 0$ for all x. Further, $\int_{-\infty}^{\infty} f_X(x)dx = 1$ and $P(X \in A) = \int_{x \in A} f_X(x)dx$.

- In that case, $E[X] = \int_{\mathbb{R}} x f_X(x)dx$. We need to define expectations for rv defined on general $(\Omega, \mathcal{F}, P)$.

Recall Riemann Integration

Riemann Integration: To compute $\int_a^b g(x)dx$ for a continuous function g.



Let $P_n = (x_0, x_1, \ldots, x_n)$ denote the partition of $[a, b]$ into n segments: $x_0 = a < x_1 < x_2 < \cdots < x_n = b$. Let $\|P_n\| = \max_i |x_i - x_{i-1}|$.

Let $L(g)$ and $U(g)$ be the Lower and Upper Darboux sums:

$$\left.\begin{array}{l} L(g) = \sum_{i=0}^{n-1} \inf_{x\in[x_i,x_{i+1}]} g(x)\cdot(x_{i+1}-x_i) \\ U(g) = \sum_{i=0}^{n-1} \sup_{x\in[x_i,x_{i+1}]} g(x)\cdot(x_{i+1}-x_i) \end{array}\right\} \begin{array}{c}\text{Upper and Lower}\\ \text{Darboux sums}\end{array}$$

If both L and U exist and their limits as $n \to \infty$ and $\|P_n\| \to 0$, are equal, then g is said to be Riemann integrable, and:

$$\int_a^b g(x)dx = \lim_{n\to\infty, \|P_n\|\to 0} \sum_{i=0}^{n-1} \inf_{x\in[x_i,x_{i+1}]} g(x)\cdot(x_{i+1}-x_i).$$

Example of a non-Riemann integrable function

Consider g defined on $[0,1]$.

$$g(x) = \begin{cases} 1 & \forall x \in [0,1] \cap \mathbb{Q}^c \\ 0 & \text{otherwise} \end{cases}$$

Since rational numbers are countable, $g(x) = 1$ almost everywhere.

We have

$$\lim_{n\to\infty} L(g) = 0 \neq \lim_{n\to\infty} U(g) = 1.$$

Then g is not Riemann integrable. Need to define integration for a richer class of functions.

Lebesgue Integral: The underlying idea

Another approach to computing definite integrals where instead of partitioning the x-area, we partition over $y = g(x)$ and then compute the area.

Lebesgue Integral: Definition

- Consider $g(x) \geq 0$ for $x \in [a, b]$.
- Let $y_0 \leq \min_{x \in [a,b]} g(x)$ and $y_n \geq \max_{x \in [a,b]} g(x)$.
- Consider partition $Q_n = (y_0, y_1, \ldots, y_n)$.
  $\|Q_n\| = \max_{1 \leq i \leq n}(y_i - y_{i-1})$.
- Let μ denote the Lebsgue measure on $([a, b], \mathcal{B}([a, b]))$,
- Lebesgue integral $\int_a^b g(x) d\mu$ is defined by:

$$\lim_{\|Q_n\| \to 0} \sum_{k=0}^{n} y_k \mu(\{x : y_k \leq g(x) < y_{k+1}\})$$

- When $g(x)$ is 0 on irrationals and 1 on rationals $\int_a^b g(x)d\mu$ equals $b-a$.
- If g is Riemann integrable, g will be Lebesgue integrable.
- On $(\Omega, \mathcal{F}, P)$ for a random variable X that takes finitely many values $x_1, x_2, \ldots, x_k$ on a partition of Ω, $A_1, \ldots, A_k$, define $EX = \int XdP$ as

$$\sum_k x_k P(A_k).$$

Lebesgue Integral for random variables

- On $(\Omega, \mathcal{F}, P)$ for general non-negative random variable X, EX is again $\int XdP$ and is defined as
- Define for $n \geq 1$ as:

$$I_n = \sum_{k=0}^{n \cdot 2^n} \frac{k}{2^n} P\left(\frac{k}{2^n} \leq X < \frac{k+1}{2^n}\right)$$

- Every term in I_n will be dominated by the sum of 2 consecutive terms in I_{n+1}.
- Then, for some I^*, $I_n \uparrow I^*$ as $n \uparrow \infty$. Set $\int_\Omega XdP$ to equal to I^*.

Extending the integral

- Any r.v. $X = X^+ - X^-$ where
$$X^+ = \max(X, 0), \quad X^- = \max(-X, 0).$$
- Then, $|X| = X^+ + X^-$.
- Further, $\int |X| dP = \int X^+ dP + \int X^- dP$.
- We say X is integrable when $\int |X| dP < \infty$,
- and set $\int X dP = \int X^+ dP - \int X^- dP$.

Some properties follow

- If $X \leq Y$ and both are integrable, then $EX \leq EY$.
- Linearity

$$\int (\alpha X + \beta Y) dP = \alpha \int X dP + \beta \int Y dP.$$

Interchanging limits and integrals

- Suppose that the sequence of rv $X_n \to X$ almost surely.
- When do we have $EX_n \to EX$? Consider a counterexample

Example

$\Omega = [0, 1]$. $P([a, b]) = b - a$.

$$X_n(\omega) = \begin{cases} n & \text{if } \omega \in [0, \frac{1}{n}) \\ 0 & \text{otherwise} \end{cases}$$

- For each $\omega > 0$, eventually $X_n(\omega) = 0$ for n large enough. (except at $\omega = 0$)
- $X_n \to X$ a.s. where $X = 0$. But $EX_n = 1 \; \forall \, n$ and $EX = 0$.

Convergence Theorems

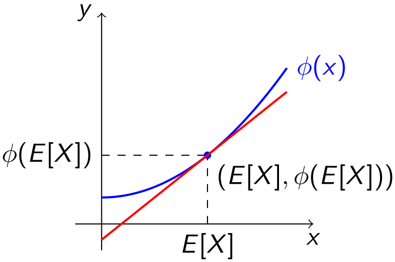
- Suppose $X_n \xrightarrow{a.s.} X$. Then when does $E[X_n] \to E[X]$?
- **Rephrasing**: When is $\lim_{n\to\infty} \int X_n dP = \int \lim_{n\to\infty} X_n dP$, i.e., when can limit and integral be interchanged?
- **Monotone Convergence Theorem (MCT)**
  If X_n's are non-negative and $X_1 \leq X_2 \leq X_3 \leq \ldots$ Then $X_n \uparrow X \Rightarrow E[X_n] \uparrow E[X]$.
- **Dominated Convergence Theorem (DCT)**
  If $\exists$ rv Y such that $|X_n| \leq Y \; \forall n \geq 1$ and $E[Y] < \infty$. Then $X_n \to X \Rightarrow E[X_n] \to E[X]$.

Jensen's Inequality

Let ϕ: convex fn on an interval containing the range of X.

$$E[\phi(X)] \geq \phi(E[X])$$

Visually:



Because ϕ is convex, the tangent to the point $(E[X], \phi(E[X]))$ will lie below $\phi(x)$.

Jensen's Inequality: Proof

- The supporting line for a convex function ϕ at $x_0 = E[X]$ is given by $y = \phi(E[X]) + m(X - E[X])$ for $m = \phi'(E[X])$.
- By convexity, the function is always above this line:

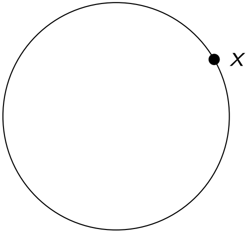
$$\phi(X) \geq \phi(E[X]) + m(X - E[X])$$

- Taking expectation on both sides, it follows that

$$\Rightarrow E[\phi(X)] \geq \phi(E[X]).$$

Example of non-measurable sets (Vitali Sets)

- Consider a circle of unit circumference, which we can identify with $[0, 1)$.



- Define an equivalence relation $x \sim y$ if $x - y \in \mathbb{Q}$. This partitions the circle into equivalence classes.

- Define $A_x = \{y \in [0, 1) : y \sim x\} = \{x + q \pmod 1 : q \in \mathbb{Q}\}$.

- Now, let B_0 be a set constructed by choosing exactly one element from each equivalence class (using the Axiom of Choice).

Example of non-measurable sets (Vitali Sets)

- For each rational $r \in \mathbb{Q} \cap [0, 1)$, let $B_r = \{b + r \pmod 1 : b \in B_0\}$.
- It can be seen that the sets B_r are disjoint and

$$\bigcup_{r \in \mathbb{Q} \cap [0,1)} B_r = [0, 1)$$

- But $P(B_r)$ can't be well defined under the standard Lebesgue measure.
  - If $P(B_0) = 0, \sum_r P(B_r) = 0 \neq 1$.
  - If $P(B_0) > 0$, then $\sum_r P(B_r) = \infty \neq 1$.

Connecting $E[h(X)] = \int h(X)dP$ to $\int h(x)f(x)dx$ when X has a pdf f

- Let's start with $h(X) = \mathbb{I}(\{\omega : X(\omega) \in B\}) = \mathbb{I}(X \in B)$.
- RV X has a probability density function f if
$$P(X \in B) = \int_B f(x)dx.$$
- Now suppose $h(X)$ is a simple rv taking finitely many values:
$$h(X(\omega)) = \sum_{k=1}^{n} \alpha_k \mathbb{I}(X \in B_k).$$

- $$\int h(X)dP = \int \sum_{k=1}^{n} \alpha_k \mathbb{I}(X \in B_k)dP$$
- so that
$$\int h(X)dP = \sum_{k=1}^{n} \alpha_k P(B_k) = \sum_{k=1}^{n} \alpha_k \int_{B_k} f(x)dx$$
- Therefore,
$$\int h(X)dP = \int \left(\sum_{k=1}^{n} \alpha_k \mathbb{I}(x \in B_k) \right) f(x)dx = \int h(x)f(x)dx.$$

$h(X)$ a general non-negative random variable

For a general non-negative r.v. $h(X)$, we can construct a sequence of simple r.v.'s that approximate it from below.

- Let $h_n(X)(\omega) = \sum_{k=0}^{n2^n} \frac{k}{2^n} \mathbb{I}_{\left\{\frac{k}{2^n} \leq h(X)(\omega) < \frac{k+1}{2^n}\right\}}$.
- It can be seen that $h_n(X)(\omega) \uparrow h(X)(\omega)$ for all ω.
- From the previous slide, we know $E[h_n(X)] = \int h_n(x)f(x)dx$.
- By the Monotone Convergence Theorem (MCT):

$$E[h(X)] = \lim_{n\to\infty} E[h_n(X)] = \lim_{n\to\infty} \int h_n(x)f(x)dx$$

- Applying MCT again to the right-hand side integral:

$$\lim_{n\to\infty} \int h_n(x)f(x)dx = \int \lim_{n\to\infty} h_n(x)f(x)dx = \int h(x)f(x)dx$$

- Thus, $E[h(X)] = \int h(x)f(x)dx$.

Notions of convergence

Almost sure convergence

$X_n \xrightarrow{a.s.} X$ if $P(\{\omega : X_n(\omega) \to X(\omega) \text{ as } n \to \infty\}) = 1$.

Convergence in probability

$X_n \xrightarrow{p} X$ if $\forall \epsilon > 0,\ P(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$.

Pointwise convergence

A sequence of functions $\{f_n\}$ on a domain D converges pointwise to f if for every $x \in D$, $\lim_{n\to\infty} f_n(x) = f(x)$.

Uniform convergence

$f_n \xrightarrow{unif} f$ if $\forall \epsilon > 0,\ \exists N \in \mathbb{N}$ s.t. $\forall n > N$ and $\forall x \in D$, we have $|f_n(x) - f(x)| < \epsilon$.

Homework

HW [II]

Show that convergence a.s. $\Rightarrow$ Convergence in prob. but the converse is not true.