# Lecture 1: <u>Measure Theoretic Probability Theory</u>

<u>Why Do We Care About This</u>? Ergodic theory describes the stochastic behavior of deterministic dynamical systems, and <u>it uses measure theoretic language.</u>

## <u>Setup of Naïve Probability Theory:</u>

- <u>finite</u> sample space $\Omega = \{\omega_1, \dots, \omega_N\}$.
- each point has <u>probability</u> $\boxed{\mathrm{Prob}(\omega_i) = p_i}$
- An <u>event</u> is a subset $E \subseteq \Omega$. It has probability

$$\boxed{\mathrm{Prob}(E) = \sum_{\omega \in \Omega} \mathrm{Prob}(\omega)} = \sum_{i=1}^{N} p_i \underbrace{1_E(\omega_i)}_{\text{indicator} = \begin{cases} 1 & \omega_i \in E \\ 0 & \text{else} \end{cases}}$$

But this doesn't work in models where $\Omega$ is uncountable* (e.g. $\Omega = [0,1]$, $\mathbb{R}^d$ etc) and each $\omega \in \Omega$ has prob. zero

$\longrightarrow$ that's what measure theory is for.

---

* A set $A$ is <u>countable</u> (בן מניה) if it can be put in the form $A = \{a_1, \dots, a_n\}$ or $\{a_1, a_2, a_3, \dots\}$.
Many sets ($[0,1]$, $\mathbb{R}$, $\{\text{irrationals}\}$, ...) are <u>uncountable</u>.

# Measure Theoretic Probability Theory (Kolmogorov, '33)

A <u>probability space</u> is $(\Omega, \mathcal{F}, \mu)$ where

(1) $\Omega$ is a general (perhaps uncountable) set

(2) $\mathcal{F}$ is a collection of subsets of $\Omega$, called "<u>measurable sets</u>", with the <u>σ-algebra axioms</u>:

- $\emptyset, \Omega \in \mathcal{F}$
- if $E \in \mathcal{F}$, then $E^c := \{\omega \in \Omega : \omega \notin \Omega\} \in \mathcal{F}$
- if $E_1, E_2, \ldots$ is a <u>countable</u> collection of sets in $\mathcal{F}$, then $\bigcup_{i=1}^{\infty} E_i, \bigcap_{i=1}^{\infty} E_i \in \mathcal{F}$. *

(3) $\mu$, called the <u>probability measure</u>, is a function $\mu : \mathcal{F} \longrightarrow [0,1]$ s.t. $\boxed{\mu(\Omega) = 1}$ and with the <u>σ-additivity property</u>:

If $E_i$ is a <u>countable</u> collection of measurable sets and $E_i \cap E_j = \emptyset$ for all $i \neq j$, then

$$\boxed{\mu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i)}$$

---

* $\bigcup_{i=1}^{\infty} E_i = \{\omega \in \Omega : \omega \in E_i \text{ for } \underline{\text{some}}\ i\}$

$\bigcap_{i=1}^{\infty} E_i = \{\omega \in \Omega : \omega \in E_i \text{ for } \underline{\text{all}}\ i\}$

[Why do we not simply take $\mathcal{F} = \{ \substack{\text{all subsets} \\ \text{of } \Omega} \}$?
Because there are (many) important cases when we cannot define a $\sigma$-additive measure with given symmetries on all subsets.

Solovay's Thm ('70): The paradoxical "non measurable sets" which cause this problem cannot be constructed without the axiom of choice

$\Rightarrow$ physicists don't need to worry about non measurable sets (but mathematicians do) ]

Random Variables:

- A random variable (or an "$\mathcal{F}$-measurable function") is a function $f: \Omega \to \mathbb{R}$ s.t.

$$[f \le t] := \{\omega \in \Omega : f(\omega) \le t\} \in \mathcal{F} \quad \text{for all } t.$$

- Allows to define the distribution of $f$

$$F(t) = \mu[f \le t] \qquad (t \in \mathbb{R})$$

- For example, a Gaussian random variable on $\Omega$ is a measurable $f: \Omega \to \mathbb{R}$ s.t. $\mu[f \le t] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{t} e^{-s^2/2\sigma^2} ds$ for all $t$.

If $f:\Omega\to\mathbb{R}$ is bounded and measurable, we can also define its <u>expectation</u>, or <u>integral</u>

$$\boxed{\mathbb{E}_\mu(f)=\int_\Omega f\,d\mu}$$

## How to Define the Integral:

(1) Case 1 ("simple function"): Suppose $f$ has finitely many values, i.e.

$$f(\omega)=\sum_{i=1}^{n} y_i\,1_{[f=y_i]}(\omega).$$ Then

$$\boxed{\int_\Omega f\,d\mu := \sum_{i=1}^{n} y_i\,\mu\{\omega\in\Omega : f(\omega)=y_i\}}$$

(2) Case 2 (measurable functions $|f|\le M$):

- Such functions are uniform limits of simple functions:

$$f(\omega)=\sum_{k=-Mn}^{Mn}\frac{k}{n}\,1_{\left[\frac{k-1}{n}<f\le\frac{k}{n}\right]}(\omega)\pm\frac{1}{n}$$

- $$\boxed{\int_\Omega f\,d\mu := \lim_{n\to\infty}\sum_{k=-Mn}^{Mn}\frac{k}{n}\,\mu\left[\frac{k-1}{n}<f\le\frac{k}{n}\right]}$$

**Stochastic Processes**: A stochastic process (in discrete time) is a sequence of measurable functions $f_i : \Omega \to \mathbb{R}$ on the same probability space.

The joint distribution is

$$\mathrm{Prob}\left[ a_i < f \le b_i \quad (i = 1, \dots, n) \right]$$

$$:= \mu \{ \omega \in \Omega : a_i < f(\omega) \le f(b_i) \, , \; i = 1, \dots, n \}$$

$$\left( = \mu \left( \bigcap_{i=1}^{n} [f \le b_i] \setminus [f \le a_i] \right) . \right)$$

## Examples

(I) Bernoulli Processes

- Informally: A sequence of independent random variables $X_1, X_2, \dots$, each taking $N$ possible values $s_1, \dots, s_N$ with prob. $p_1, \dots, p_N$.

- Formally: The sample space of $(X_1, X_2, \dots)$ is

$$\Omega = \{ \underline{x} = (x_1, x_2, x_3, \dots) : x_i \in S \}$$

where $S = \{ s_1, \dots, s_N \}$.

A <u>cylinder set</u> is a set of the form

$$[a_1, \dots, a_n] := \{ \underline{x} \in \Omega : x_i = a_i \ (i=1,\dots,n) \}.$$

We'd like to have

$$\mu[a_1, \dots, a_n] = p_{a_1} \cdots p_{a_n}.$$

<u>Thm</u>. There exists a $\sigma$-algebra $\mathcal{F}$ and a $\sigma$-additive $\mu : \mathcal{F} \to [0,1]$ s.t.:

(a) $\mathcal{F}$ contains all the cylinders

(b) $\mu[a_1, \dots, a_n] = p_{a_1} \cdots p_{a_n}$ for every cylinder.

The coordinate functions $X_i(\underline{x}) = x_i$ form a stochastic process, called a <u>Bernoulli process</u>. $\boxed{X_i = \text{"}i\text{th outcome"}}$

Why do we need $\Omega$ and $\mathcal{F}$? Because it allows us to study events which involve <u>all of $X_i$ at the same time</u>, e.g.

$$\left\{ \underline{x} \in \Omega : \frac{f(x_1) + \dots + f(x_n)}{n} \xrightarrow[n \to \infty]{} \int_\Omega f \, d\mu \right\}.$$

(II) <u>Markov Chains</u>: Suppose

- $G$ is a directed graph with finite or countable collection of vertices $S$
- $(p_i)_{i\in S}$ is a prob. vector
- $(p_{ij})_{S\times S}$ is a stochastic matrix s.t

  $p_{ij} > 0 \iff i \to j$ is an edge in the graph.

<u>Sample Space</u>:

$$\Omega = \left\{ (x_1, x_2, \dots) : \begin{array}{l} x_i \in S, \ x_i \to x_{i+1} \\ \text{is an edge of } G \end{array} \quad (i \in \mathbb{N}) \right\}$$

We'd like $[a_1, a_2, \dots, a_n] = \{\underline{x} : x_i = a_i \ (i = 1, \dots, n)\}$ to have probability $p_{a_1} p_{a_1 a_2} \cdots p_{a_{n-1} a_n}$.

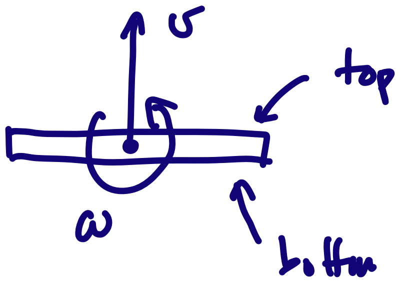
<u>Thm</u>. There exists a $\sigma$-algebra $\mathcal{F}$ which contains all cylinders and a $\sigma$-additive $\mu : \mathcal{F} \to [0,1]$ s.t. for every cylinder

$$\mu[a_1, \dots, a_n] = p_{a_1} p_{a_1 a_2} \cdots p_{a_{n-1} a_n} .$$

Again, $X_i : \Omega \to \mathbb{R}$, $X_i(\underline{x}) = x_i$ is a stochastic process

$$\boxed{X_i = \text{"position of the chain at time } i\text{"}}$$

(3) "Real" Coin Tossing: What's the probability that a tossed coin will fall with the original top side up?




- Sample space: $\{(v,\omega) : v > 0,\ \omega \in \mathbb{R}\}$

  $v$ = vertical velocity

  $\omega$ = angular velocity

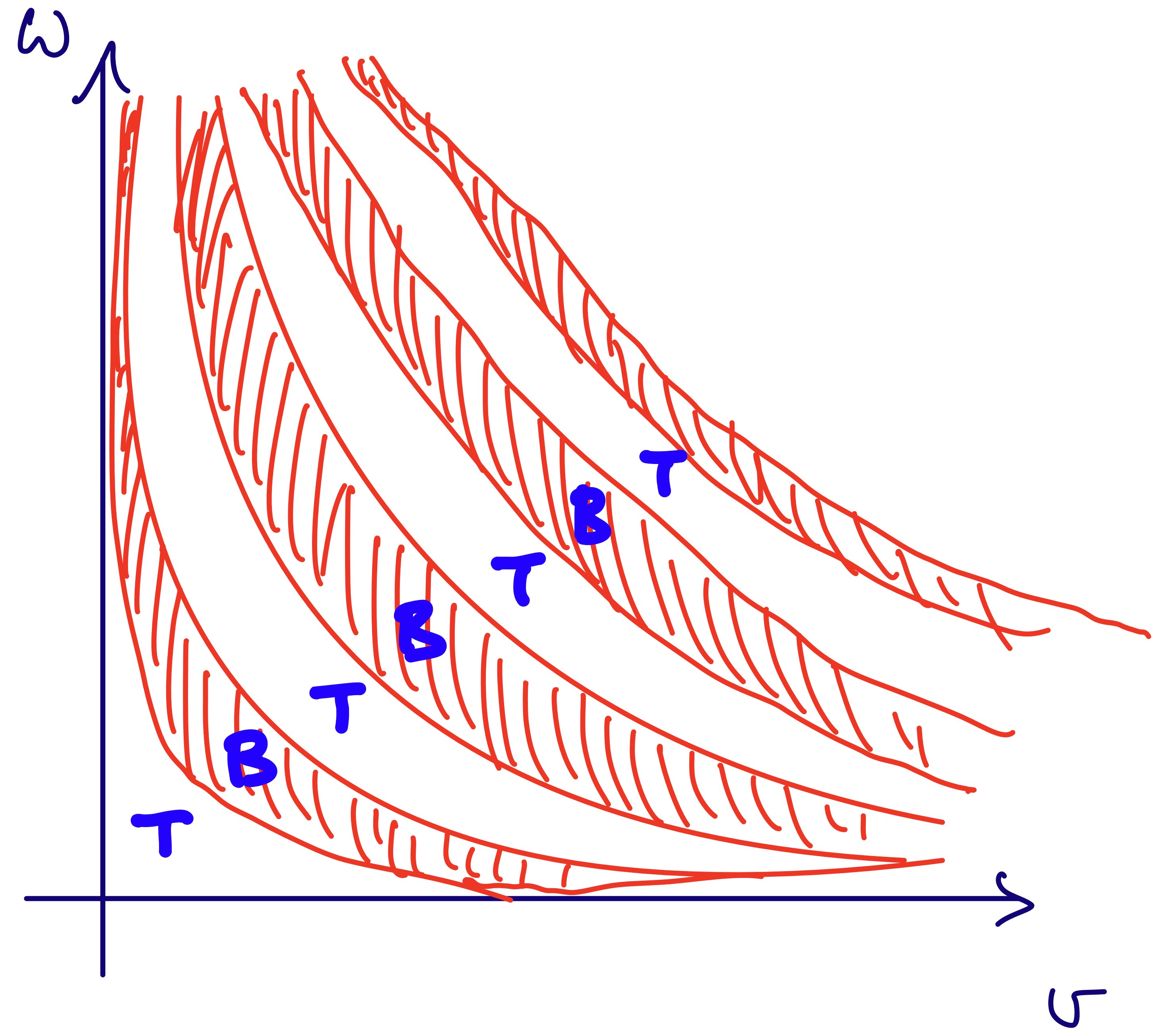
- Random Variable: $X(v,\omega) = \begin{cases} T & \text{number of turns in the air is even} \\ B & \text{is odd} \end{cases}$

  - time in the air: $v - g(t/2) = 0$

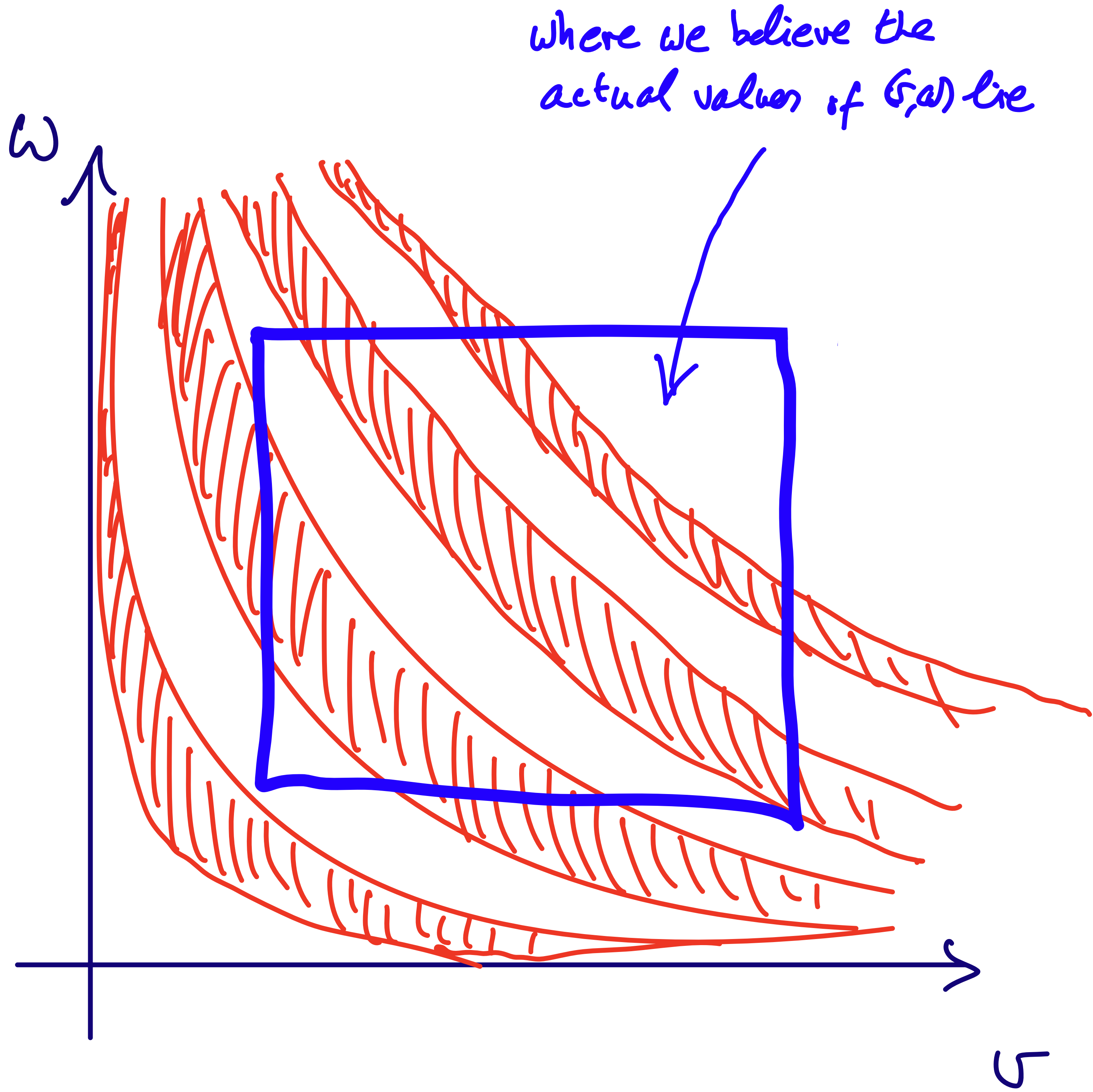
    $t = 2v/g$

  - number of flips: $N = \lfloor \omega t / \pi \rfloor = \lfloor \frac{2\omega v}{\pi g} \rfloor$

$$X(v,\omega) = \begin{cases} T & 2k \le \frac{2\omega v}{\pi g} < 2k+1 \\ B & 2k+1 \le \frac{2\omega v}{\pi g} < 2k+2 \end{cases} \quad (k \in \mathbb{Z}).$$



ω
T
B
T
B
T
B
T
T
σ




<u>Heuristic</u>: For many measures which model our uncertainty as to the <u>precise</u> value of $(\sigma,\omega)$, the probability of "T" is <u>roughly</u> $1/2$.