TAM 598 Lecture 3 :

DISCRETE RANDOM VARIABLES

Announcements:

- HW 1 covers lectures 1-4 ; due on Feb 12
- to be submitted via CANVAS

I Some definitions and concepts:

- We condition probabilities on our current information I. Because this info is always in the background, we will not explicitly show it in our notation.

$$p(x|I) = p(x)$$

- Say we are doing an experiment. Result of the experiment depends on a bunch of things denoted by ω, which we call the state of nature.

particular state of variables

→ ω might be overcomplete, but it includes the variables that determine the result of the experiment

- The state of nature takes values in an enormous set Ω, the set of all possible $\omega_1, \omega_2, \ldots$ Since we don't know which ω will happen, we describe this uncertainty using a probability measure.

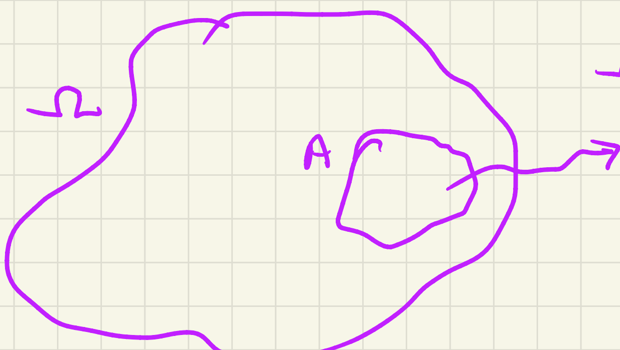
all possible ω_i $i=1,\ldots$

Definitions and concepts, cont'd:

- the probability measure is a function that takes a subset A of Ω and tells us how probable it is that the state of nature ω will be in A.

  → We denote this probability measure by $\boxed{p(A)}$

- When the set Ω is equipped with a probability measure, we call it a probability space



Ω = set of all possible ω

$A = \{\omega_1, \omega_2, \ldots\}$

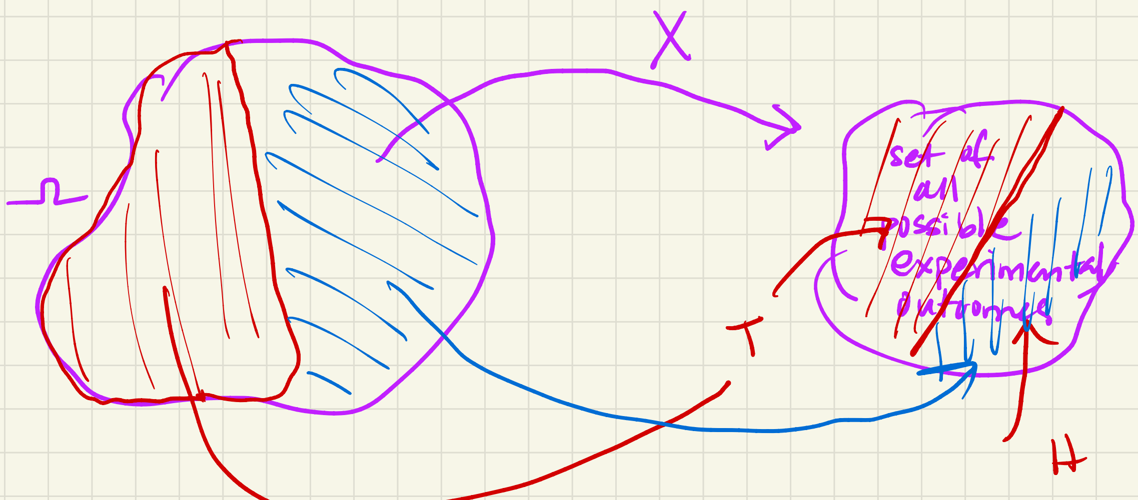
$p(A)$ = probability measure

(II) Definition of a random variable :

Let X (capital letter) be the result of an experiment, read by some measuring device.

The measuring device takes the state of nature ω and maps it to a number $X(\omega)$

So we define a random variable X as a function from Ω to some set of values



Depending on the values that X takes, we can classify it into one of several types:

- **discrete random variable** - $X(\omega)$ takes discrete values
  eg, heads or tails or 0, 1, 2, ...

- **continuous random variable** - eg, values between 0 and 1, or all positive reals

- **random vector** - when $X(\omega)$ is a vector

- **random matrix** - when $X(\omega)$ is a matrix

- **random process** - when $X(\omega)$ itself is a function
  (also "stochastic process", "random field")

Note: X, Y, Z (uppercase) are random variables
x, y, z (lowercase) are specific values of random variables

Example: coin toss from previous lecture

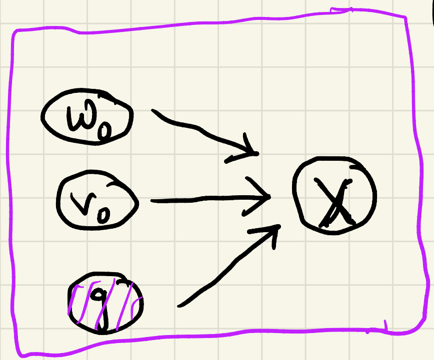
state of nature $\omega = (v_0, w_0)$

v_0: initial velocity

w_0: angular velocity

Then

$$X(v_0, w_0) = \begin{cases} T & \text{if } \dfrac{2 v_0 w_0}{g} \bmod 2\pi \in \left(\dfrac{\pi}{2}, \dfrac{3\pi}{2}\right) \\ H & \text{else} \end{cases}$$



The **probability mass function** of X is a function that gives the probability that X takes a certain value $X = x$.

- we write $p(X = x)$ or just $p(x)$
- if we want to be explicit about background info, we could write $p(x \mid I)$
- mathematicians write $\boxed{f_X(x)}$ to remember which random variable X we are concerned with

To define $p(x)$ first define the set of states of nature that yield the specific outcome $X = x$

$$(X = x) \equiv \{ \omega \in \Omega : X(\omega) = x \}$$

Then we define the probability mass function (pmf) as

$$p(x) \equiv f_X(x) := p(X = x)$$

the probability of getting $X = x$

properties of the probability mass function $p(x)$

$p(X = x)$

1) It is non-negative

$$p(x) \geq 0 \qquad \text{for all } x$$

2) It is normalized

$$\sum_x p(x) = 1$$

sum over all possible values of x

3) It is additive. For any set A of possible values of X, the probability that X takes values in A is

$$p(x \in A) = \sum_{x \in A} p(x)$$

Example: X is the outcome of a coin toss. Let $X = 0$ denote heads, and $X = 1$ denote tails.

Then for a fair coin:

heads $\quad p(X = 0) = \frac{1}{2}$

tails $\quad p(X = 1) = \frac{1}{2}$

III Functions of discrete random variables

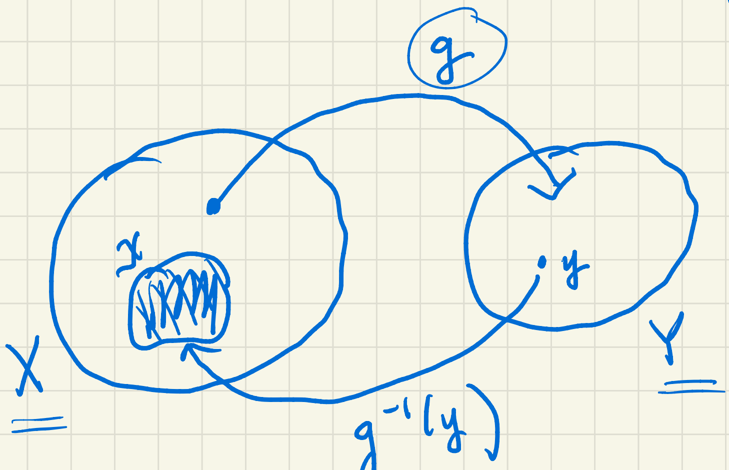
→ let X be a discrete random variable

→ let g be a function from the state space of X to some other set

→ define a new random variable $Y = g(X)$

→ let $g^{-1}(y)$ be the set of values of X that map to y through g

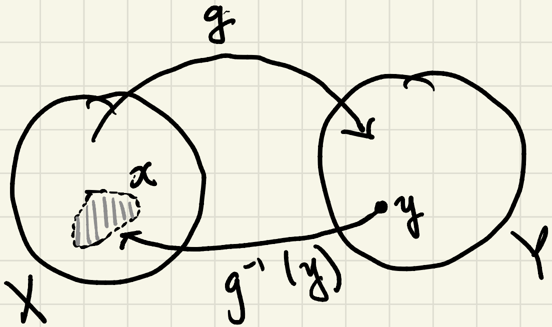
$$g^{-1}(y) = \{x : g(x) = y\}$$



Then the probability mass function of Y is

$$p(y) = P(X \in g^{-1}(y))$$

$$= \sum_{x \in g^{-1}(y)} p(x)$$



THEN: the probability mass function of Y is

$$\boxed{p(y) = p\left(X \in g^{-1}(y)\right) = \sum_{x \in g^{-1}(y)} p(x)}$$

This is the formal definition of the uncertainty propagation problem and the model calibration problem.

X are parameters of a causal model (inputs)

v_0, w_0, g

$Y = g(X)$ is the uncertain result of the causal model (output)

T, H

IV Expectation of a random variable : the value we get "on average"

$$\mathbb{E}[X] = \sum_x x\, p(x)$$

example: coin toss

$$x_0 = 0 \qquad p(x_0) = \frac{1}{2}$$

$$x_1 = 1 \qquad p(x_1) = \frac{1}{2}$$

$$\mathbb{E}[X] = \sum_x x\, p(x)$$

$$= 0 \cdot p(X=0) + 1 \cdot p(X=1)$$

$$= 0\left(\tfrac{1}{2}\right) + 1\left(\tfrac{1}{2}\right)$$

$$= \frac{1}{2}$$

Note how the expected value need not actually be a possible outcome itself.

properties of the expectation:

① $\mathbb{E}[X+c] = \mathbb{E}[X] + c$ for constant c

② $\mathbb{E}[\lambda X] = \lambda \mathbb{E}[X]$ for real numbers λ

③ For any function $g(x)$

$$\mathbb{E}[g(X)] = \boxed{\sum_{x} g(x)\, p(x)}$$

④ Jensen's Inequality. If $g(x)$ is convex

$$\underbrace{g(\mathbb{E}[X])}_{\text{function's value at the mean}} \leq \underbrace{\mathbb{E}[g(X)]}_{\text{mean of the function value}}$$

examples:

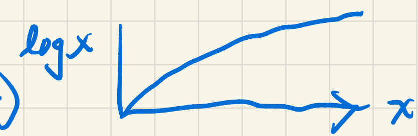
① $g(x) = x^2$

x drawn uniformly from $[-1, 1]$

$$(\mathbb{E}[X])^2 \le \underbrace{\mathbb{E}[X^2]}_{\int_{-1}^{1} x^2 \, p(x) \, dx \, = \, \frac{1}{3}} \qquad p(x) = \frac{1}{2}$$

$(\mathbb{E}[X])^2 \swarrow 0$

lead to variance decomposition



② $g(x) = \log x$ (concave, flip inequality)

$$\log(\mathbb{E}[X]) \ge \mathbb{E}[\log(X)]$$

used in information theory, "log-sum inequality"

V. Variance of random variables is the expected value of the squared deviation from its expectation

$$V[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

$$= \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])]$$

$$= \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2]$$

$$= \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2$$

$$= \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

example: variance of a coin toss

$$\left.\begin{aligned} \mathbb{E}[X^2] &= \sum_x x^2 p(x) = 0^2\left(\tfrac{1}{2}\right) + 1^2\left(\tfrac{1}{2}\right) = \tfrac{1}{2} \\ \mathbb{E}[X] &= \tfrac{1}{2} \end{aligned}\right\} \quad V[X] = \tfrac{1}{4}$$

properties of the variance :

① $\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

② $\mathbb{V}[X + c] = \mathbb{V}[X]$ for any constant c

③ $\mathbb{V}[\lambda X] = \lambda^2 \, \mathbb{V}[X]$

Let's get familiar w/ a few common pmf's that describe discrete random variables, and how to use them in python.

① Bernoulli distribution - outcomes 0 or 1

$$X = \begin{cases} 1 & \text{w/ probability } \theta \\ 0 & \text{else} \end{cases}$$

we write $\boxed{X \sim \text{Bernoulli}(\theta)}$

"distributed as", "drawn from"

② Categorical distribution aka "multinulli", K possible different outcomes

$$p(X = k) = p_k \qquad \sum_{k=1}^{K} p_k = 1$$

(3) Binomial distribution

Say you toss a coin n times, with θ = prob. of tossing heads.
Let X = the number of times you toss heads.

X is the binomial random variable

$$X \sim B(n, \theta)$$

It has PMF

$$p(X = k) = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

where $$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

(4) Poisson distribution - a model of the number of times an event occurs in an interval of space or time.

For instance:

X = # earthquakes > 6 Richter in next 100 years

= # major floods over next 100 years

= # patients arriving at the emergency room during the night shift

= # electrons hitting a detector in a specific time interval

Poisson is a good model when:

- → # of times takes discrete values 0, 1, 2, ...
- (*) → events occur independently
- (*) → the probability that an event occurs is constant per unit time, ie
  the average rate at which events occur is constant
- → events cannot occur at the same time

we say $X \sim \text{Pois}(\lambda)$ λ = rate of event

The PMF is $p(X=k) = \dfrac{\lambda^k}{k!} \exp(-\lambda)$

Then $\mathbb{E}[X] = \sum_{k=0}^{\infty} k \, p(X=k) = \lambda$

$\mathbb{V}[X] = \lambda$