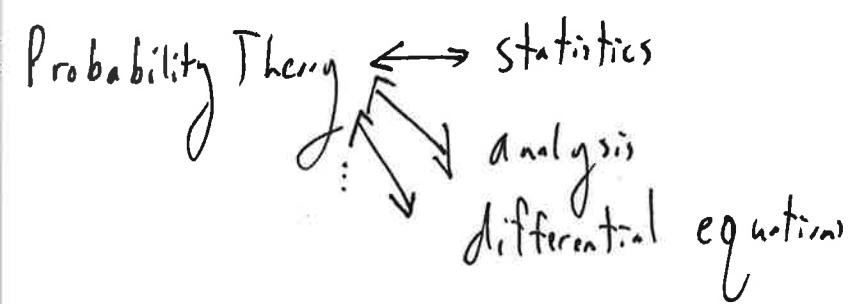


Lecture 1

- Foundational subject in computer science, electrical engineering, economics, ...
- We will examine different quantifications of randomness and how to analyze them. Lots of formalism, but also intuition matters.

Defn: Let Ω be a discrete set. A function $m: \Omega \rightarrow \mathbb{R}$ is a discrete probability distribution if m satisfies:

- (1) Non-negativity: $\forall w \in \Omega, m(w) \geq 0$.
- (2) Sum to 1: $\sum_{w \in \Omega} m(w) = 1$.

We may write, for $E \subseteq \Omega$, $m(E) = \sum_{w \in E} m(w)$. We call a subset $E \subseteq \Omega$ an event. We call Ω a sample space.

Remark: To really make the idea of a probability distribution and sample space precise (i.e. to delimit possible Ω), measure theory required.

The way will formalize the intuitive notion of "random thing" (e.g. a

roll of a die) is through the abstract notion of random variable. ②

Defn.: Let Ω be a discrete set. We call ~~$X: \Omega \rightarrow E$~~ E -valued random variable for a space E .

Remark: Again, to really describe what E can be requires more math than this class requires. Best to figure E is something like \mathbb{R}^d , $d \geq 1$, or $\mathbb{Z} = \{-2, -1, 0, 1, 2, \dots\}$.

// This is all very abstract. But random variables and probability distributions are everywhere.

ex.: Let \bar{X} correspond to the flip of an unbiased coin. Then we can think there are two outcomes in my sample space, heads (H) and tails (T). So, $\Omega = \{H, T\}$. So, we can define \bar{X} to determine if heads was achieved:

$$\bar{X}(w) = \begin{cases} 1, & w = H \\ 0, & w = T \end{cases}$$

Then \bar{X} is, \mathbb{R} -valued (even $\{0, 1\}$ -valued) r.v. on the sample space

I can define an associated distribution function on $\{0, 1\}$ through

$$\begin{aligned} X: p_X(x) &= P(X=x) \\ &= P(\{w \mid X(w)=x\}) \end{aligned}$$

Note that p_X is a distribution on the output of X (on E), while \textcircled{A} is based on a probability distribution on the sample space Ω !

In our coin-flip case, the statement that the coin is "unbiased" tells us how the sample space works: $P(H) = P(T) = \frac{1}{2}$; it is trivial to see this $\&$ function P on Ω satisfies the properties of a probability distribution.

So, what is $p_X(x)$? Looking at how $P(H), P(T)$ work, we

$$\text{See } p_X(x) = \begin{cases} \frac{1}{2}, & x=0 \quad (\text{tails is observed}) \\ \frac{1}{2}, & x=1 \quad (\text{heads is observed}) \\ 0, & \text{else} \quad (\text{no other possibilities}) \end{cases}$$

Remark. One can just forget about the coin and define a r.v. X as

$$p_X(x) = \begin{cases} \frac{1}{2}, & x=0 \\ \frac{1}{2}, & x=1 \\ 0, & \text{else.} \end{cases} \quad \text{In fact, this is what is done! The physical}$$

phenomenon (coin flip) gives intuition, but then we can just define a r.v. (and associated distribution) as mathematical function.

From this abstract formulation, I can see an easy way to generalize to a notion of "biased" coin: consider a r.v. \bar{X} with distribution function

$$P_{\bar{X}}(x) = \begin{cases} 1-p, & x=0 \\ p, & x=1 \\ 0, & \text{else} \end{cases} \quad \text{for some } p \in [0,1].$$

I think of this as a coin with $P(\bar{X}=H) = p$. This is the Bernoulli r.v. with parameter p.

Recall that for $E \subseteq \Omega$, we let $P(E) := \sum_{\omega \in E} m(\omega)$, where m is a ~~distribution~~ function. What kind of properties does P have as a function on subsets of Ω ?

Theorem (Properties of P): Let $P(E) := \sum_{\omega \in E} m(\omega)$ for a distribution m on

Ω . Then:

$$(1.) \forall E \subseteq \Omega, P(E) \geq 0$$

$$(2.) P(\Omega) = 1$$

$$(3.) \forall E, F \subseteq \Omega \text{ such that } E \subseteq F, P(E) \leq P(F)$$

$$(4.) \text{Suppose } E, F \text{ are disjoint: } E \cap F = \emptyset. \text{ Then } P(E \cup F) = P(E) + P(F).$$

$$(5) \text{ Let } A^c := \Omega \setminus A \\ = \{w \in \Omega \mid w \notin A\}.$$

Then $\forall A \subseteq \Omega, P(A^c) = 1 - P(A)$.

Proof:

$$(1) \quad P(E) = \sum_{w \in E} \underbrace{m(w)}_{\substack{\geq 0 \\ \forall w}} \geq 0 \quad \square$$

$$(2) \quad P(\Omega) = \sum_{w \in \Omega} m(w) = 1, \text{ by definition of } m \text{ being a distribution.}$$

(3) Suppose $E \subseteq F$. Then let $G = \{w \in F \mid w \notin E\}$, which is well-defined and empty iff $E = F$. Then:

$$P(E) = \sum_{w \in E} m(w) \quad \underbrace{\quad}_{\text{non-negative}}$$

$$\leq \sum_{w \in E} m(w) + \sum_{w \in G} m(w)$$

$$= \sum_{w \in E \cup G} m(w)$$

$$= \sum_{w \in F} m(w)$$

$$= P(F). \quad \square$$

$$(4.) \quad \mathbb{P}(E \cup F) = \sum_{\omega \in E \cup F} m(\omega)$$

$$= \sum_{\omega \in E} m(\omega) + \sum_{\omega \in F} m(\omega)$$

$$= \mathbb{P}(E) + \mathbb{P}(F).$$

(5.) \mathbb{P} Follows directly from (4), since $\forall A \subseteq \Omega, A \cap A^c = \emptyset$. ■

Remark : The proof of (4.) is illuminating. We see

$$\sum_{\omega \in E \cup F} m(\omega) = \sum_{\omega \in E} m(\omega) + \sum_{\omega \in F} m(\omega) \quad \text{requires in general } E \cap F = \emptyset.$$

to avoid double-counting. From this intuition, the following is motivated

Theorem (Union bound): Let $A, B \subseteq \Omega$ and let \mathbb{P} be a probability on Ω associated to distribution m .

$$(a.) \quad \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

$$(b.) \quad \mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B).$$

Proof : (a.) $\mathbb{P}(A \cup B) = \sum_{\omega \in A \cup B} m(\omega)$

(7)

$$= \sum_{\omega \in A} m(\omega) + \sum_{\omega \in B} m(\omega) - \sum_{\omega \in A \cap B} m(\omega)$$

$$= P(A) + P(B) - P(A \cup B),$$

(b.) Follows from (a.) by noting that $P(A \cap B) \geq 0.$

The bound $P(A \cup B) \leq P(A) + P(B)$ is the so-called union bound.