Important Announcements

- assignment 1 will be released this Friday (Jan 17)
- review session by TA after due date

- <u>Countable set</u>: A set with cardinality that is finite or countably infinite
- <u>Uncountable set</u>: A set with cardinality that is uncountably infinite

Questions

(1) is $(w_1, w_2, w_3)^T$ a row or coloumn vector?

(2) $f(x) = 1/x$, $\sum_{x \in \mathcal{X}} f(x) = $?

$\mathcal{X} = \{1, 2, 3, 2\}$

$$\sum_{x \in \mathcal{X}} f(x) = \frac{1}{1} + \underbrace{\frac{1}{2}}_{\frac{3}{6}} + \underbrace{\frac{1}{3}}_{\frac{2}{6}} = \frac{11}{6}$$

(3) What does the integral of a function represent?

(4) What does the derivative of a function represent?

(5) Can you calculate the partial derivative of $f(x)$, where $x \in \mathbb{R}$

(6) Define the gradient

$f(x_1, x_2)$

$$\nabla f(x_1, x_2) = \left(\frac{\partial f(x_1, x_2)}{\partial x_1}, \frac{\partial f(x_1, x_2)}{\partial x_2} \right)$$

Motivation

Supervised Learning: Learning from a randomly sampled batch of labeled data

Dataset (tuple of tuples) $\xrightarrow{\text{Learner (function)}}$ Predictor = Model = Hypothesis (function)

$(X_{1,1}, \dots, X_{1,d})^T$

$D = ((\vec{X}_1, Y_1), (\vec{X}_2, Y_2), \dots, (\vec{X}_n, Y_n))$

$D \in (\mathcal{X} \times \mathcal{Y})^n$

$(\vec{X}_i, Y_i) \sim \mathbb{P}_{\vec{X}, Y}$ independent for all $i \in \{1, \dots, n\}$

D n feature-label pairs

$\mathcal{X}$ set of features

$\mathcal{Y}$ set of labels/targets

Ex: $\mathcal{X} = \mathbb{R}^2$, $\mathcal{Y} = \mathbb{R}$

$\boxed{f : \mathcal{X} \to \mathcal{Y}}$

f a function from features to labels

Ex $f(x) = 2x + 1$, $\mathcal{X} = \mathbb{R}$

$\boxed{\mathcal{A} : (\mathcal{X} \times \mathcal{Y})^n \to \{f \mid f : \mathcal{X} \to \mathcal{Y}\}}$

$\mathcal{A}$ a function from datasets to predictors

Ex: $\mathcal{A}(D) = f$ where $f : \mathcal{X} \to \mathcal{Y}$

$$f(x) = \begin{cases} y_i & \text{if } x = x_i \text{ for some } i \in \{1, \dots, n\} \\ 0 & \text{otherwise} \end{cases}$$

Probability

> Note: Humans have a bad intuition when it comes to randomness
> - Thinking Fast and Slow by: Daniel Kahneman

- Random variables
- Calculating probabilities using pmf and pdf
- Multivariate random variables
  - Conditional and marginal probabilities
- Representing random features, labels, and datasets
- Functions of random variables
- Expectation and variance

> **Warning:** If some things seem informal, it is likely because we would need tools from Measure Theory, which we will not cover in this course.

Experiment: A process that generates an uncertain outcome

Ex: flipping a coin, rolling a dice

Outcome Space/Set: The set of all outcomes from the experiment

Ex: $\mathcal{Y} = \{0, 1\}$ Heads, Tails — flipping a coin

$\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$ — rolling a dice

$[0, 900]$ — amount of a chemical in a wine

$\mathbb{R}$ — measurement error

The outcome space/set can be either a

1) countable set or
2) uncountable set

Event: A subset of the outcome space (imprecise)

an outcome is a single element of the outcome set

Ex: Outcome space: $\mathcal{Y} = \{0, 1\}$

Events: $\{0\}, \{1\}, \{0,1\} = \mathcal{Y}, \emptyset$

Ex: Outcome space: $\mathcal{X} = \{1,2,3,4,5,6\}$

Events: $\emptyset, \mathcal{X}, \{1\}, \{1,3,5\}, \{2,4,6\}, \{1,2,3\}, \ldots$

Ex: Outcome space: $[0, 900]$

Events: $\emptyset, [0,900], [0,4], [1,2] \cup [7,30], \ldots$

Probability Distribution: A function $\mathbb{P}$ defining the likelihood of each event (and satisfying certain properties)

a set containing all the events

$$\mathbb{P}: \underbrace{\text{event space/set}} \to [0,1]$$

A complicated set (σ-algebra) that we will not define

Properties: (imprecise)

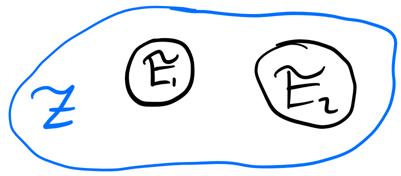
Outcome space: $\mathcal{Z}$

1. $\mathbb{P}(\mathcal{Z}) = 1$

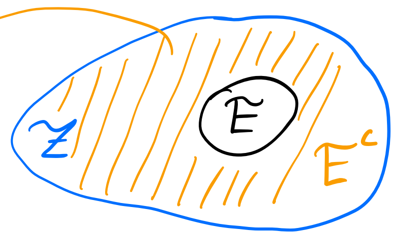
events — disjoint

2. If $\mathcal{E}_1 \subset \mathcal{Z}$, $\mathcal{E}_2 \subset \mathcal{Z}$ and $\mathcal{E}_1 \cap \mathcal{E}_2 = \emptyset$, then

$$\mathbb{P}(\mathcal{E}_1 \cup \mathcal{E}_2) = \mathbb{P}(\mathcal{E}_1) + \mathbb{P}(\mathcal{E}_2)$$



Ex: (of property 2.)



Events: $\mathcal{E}$, $\mathcal{E}^c$

$\mathcal{E} \cap \mathcal{E}^c = \emptyset$, $\mathcal{E} \cup \mathcal{E}^c = \mathcal{Z}$

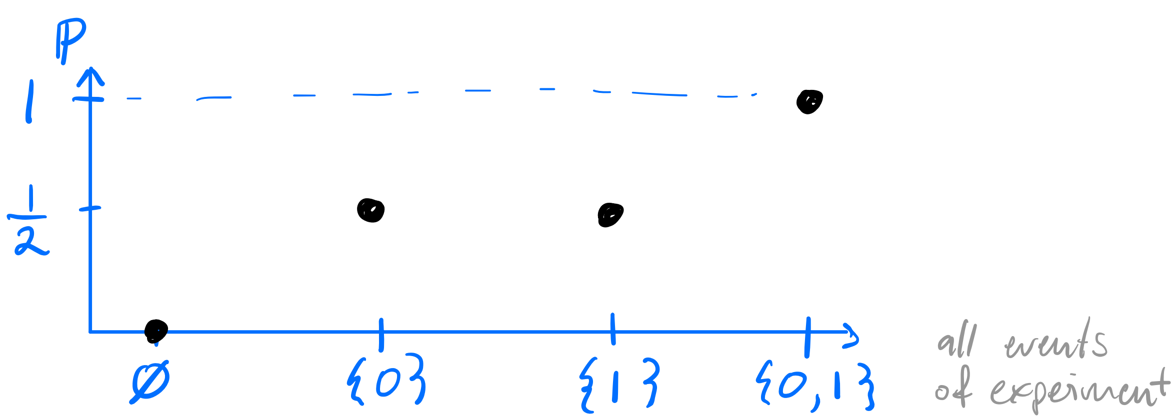
$$\underbrace{\mathbb{P}(\mathcal{E} \cup \mathcal{E}^c)}_{= \mathbb{P}(\mathcal{Z}) = 1} = \mathbb{P}(\mathcal{E}) + \mathbb{P}(\mathcal{E}^c)$$

rearranging: $\boxed{\mathbb{P}(\mathcal{E}) = 1 - \mathbb{P}(\mathcal{E}^c)}$

Ex: Outcome space: $\mathcal{Y} = \{0, 1\}$

$$\mathbb{P}(\emptyset) = 0, \quad \mathbb{P}(\mathcal{Y}) = 1, \quad \mathbb{P}(\{0\}) = \tfrac{1}{2}, \quad \mathbb{P}(\{1\}) = \tfrac{1}{2}$$

each event of an experiment is associated with a probability

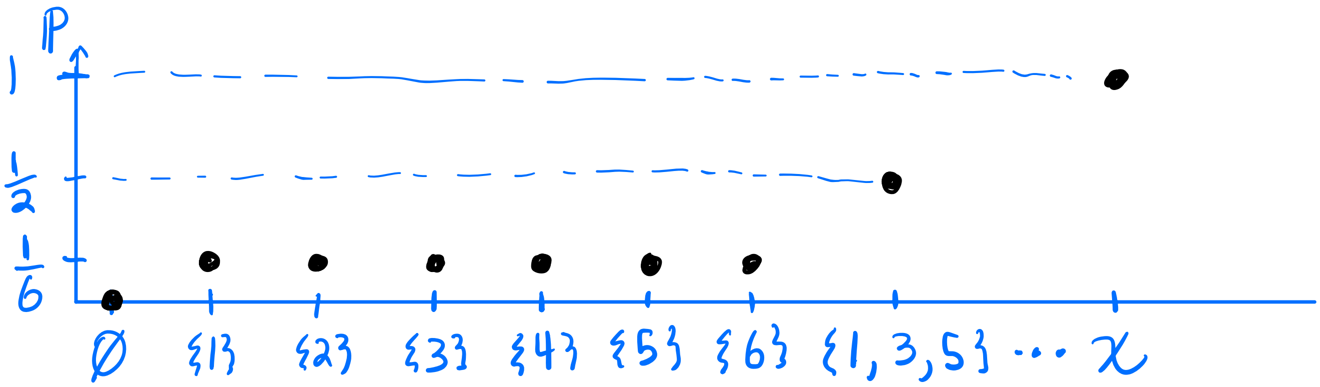


$\underline{\text{Ex}}$: Outcome space: $\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$

$$\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \cdots = \mathbb{P}(\{6\}) = \frac{1}{6}$$

$$\mathbb{P}(\{1,3,5\}) = \mathbb{P}(\{1\}) + \mathbb{P}(\{3\}) + \mathbb{P}(\{5\}) = \frac{1}{2}$$

$$\mathbb{P}(\mathcal{X}) = 1$$



Random Variables

Random Variable (r.v.): (imprecise) A variable that takes a value based on the outcome of an experiment, and is associated with a probability distribution. The value can be any random outcome.

$\underline{\text{Ex}}$: $X \in \mathcal{X} = \{1, 2, 3, 4, 5, 6\}$ with $\mathbb{P}$ from prev. example

$Y \in \mathcal{Y} = \{0, 1\}$ with $\mathbb{P}$ from prev. example

$Z \in \{H, T\}$

A random variable is actually a function (satisfying certain properties) from one outcome space to another outcome space. $\underline{\text{Ex}}$ $X(T) = 0, X(H) = 1$.
It will not be necessary to know this for this course

Probability Distributions with r.v.

Ex: Outcome space: $\mathcal{X}$ r.v.: $X \in \mathcal{X}$

$$P(\{1,3,5\}) \overset{def}{=} P(\underbrace{X \in \{1,3,5\}})$$

"the r.v. X is an element of the event $\{1,3,5\}$"

$$P(\{4,5,6\}) = P(X \in \{4,5,6\}) = P(X \geq 4)$$

$$P(\{4\}) = P(X \in \{4\}) = P(X = 4)$$

Notation: $Z \sim P$ "Z is sampled according to distribution P"

Discrete r.v.: A r.v. that takes values from:

- A countable outcome space, or
- an uncountable outcome space, but there is a countable event that has probability 1

Ex: $Y \in \mathcal{Y} = \{0,1\}$, $X \in \mathcal{X} = \{1,2,3,4,5,6\}$, $Z \in \mathbb{N}$

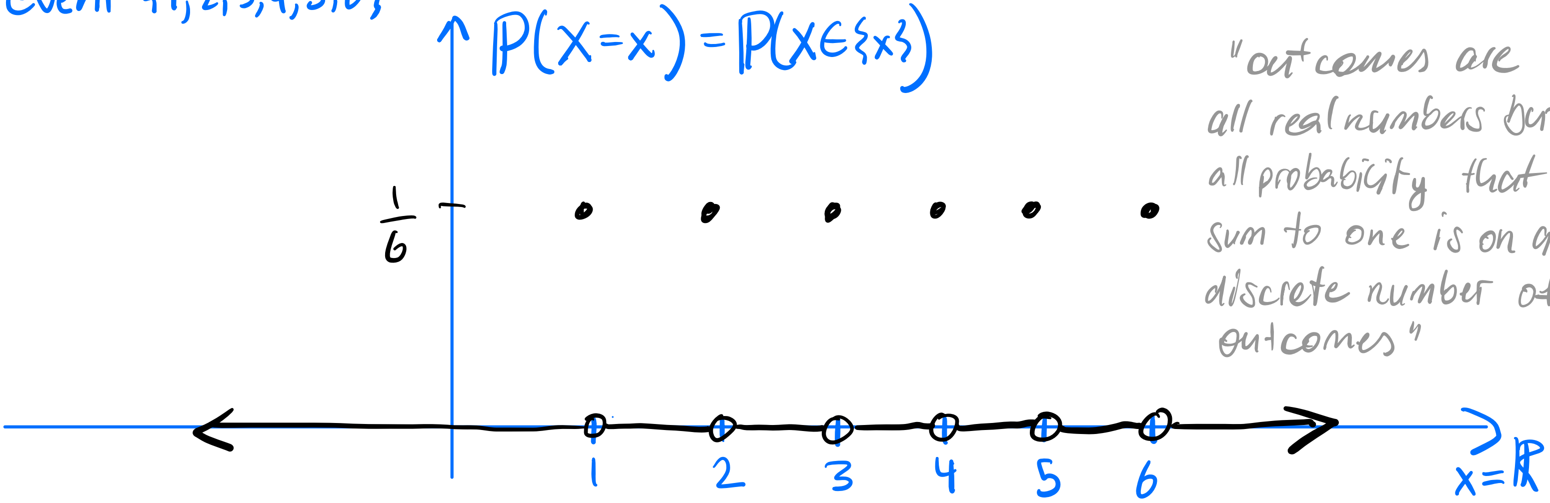
two ways to describe same discrete experiment

Ex: $X \in \mathbb{R}$ where $P(X=1) = \cdots = P(X=6) = \frac{1}{6}$

Probability 1 for countable event $\{1,2,3,4,5,6\}$

so $P(X \in \{1,2,3,4,5,6\}) = 1$

and $P(\mathbb{R} \setminus \{1,2,3,4,5,6\}) = 0$



"outcomes are all real numbers but all probability that sum to one is on a discrete number of outcomes"

$\boxed{\text{Note: You can always take a r.v. defined on a countable outcome space and define it on a larger uncountable outcome space by setting the probability of the event containing all the new outcomes to zero}}$

Continuous r.v.: A r.v. that takes values from:

- an uncountable outcome space and the probability of any single outcome is zero

Ex: $Z \in [0,900]$ and $\mathbb{P}(Z=z)=\mathbb{P}(Z\in\{z\})=0$ for all $z\in[0,900]$

but $\mathbb{P}(Z\in[0,900])=1$

Ex: $Z\in\mathbb{R}$ and $\mathbb{P}(Z=z)=\mathbb{P}(Z\in\{z\})=0$ for all $z\in\mathbb{R}$

but $\mathbb{P}(Z\in\mathbb{R})=1$

Calculating Probabilities

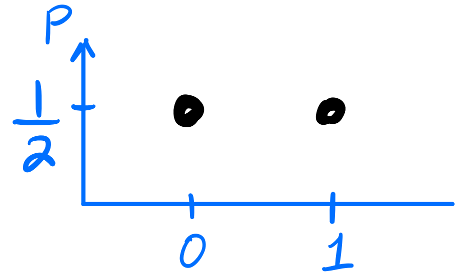
Motivation: It is hard to define the values of a probability distribution $\mathbb{P}$ for all the events

Probability Mass Function (pmf): A function $p:\mathcal{Z}\to[0,1]$ where $\mathcal{Z}$ is a discrete outcome space and $\sum_{z\in\mathcal{Z}} p(z)=1$.

The probability of an event $\mathcal{E}\subset\mathcal{Z}$ is:

$$\boxed{\mathbb{P}(Z\in\mathcal{E}) \overset{\text{def}}{=} \sum_{z\in\mathcal{E}} p(z)} \qquad \text{where } Z\in\mathcal{Z}$$

Ex: Outcome space: $\mathcal{Y}$



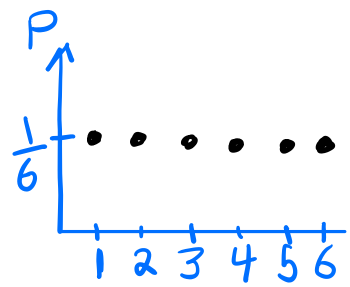
$$p(0)=\tfrac{1}{2},\ p(1)=\tfrac{1}{2}$$

$$\mathbb{P}(Y\in\{0,1\})=\sum_{y\in\{0,1\}}p(y)=p(0)+p(1)=1$$

$$\mathbb{P}(Y=0)=\mathbb{P}(Y=1)=\mathbb{P}(Y\in\{1\})=\sum_{y\in\{1\}}p(y)=p(1)=\tfrac{1}{2}$$

$$\mathbb{P}(Y\in\emptyset)=\sum_{y\in\emptyset}p(y)=0$$

Ex: Outcome space: $\mathcal{X}$



$$p(1)=p(2)=\dots=p(6)=\tfrac{1}{6}$$

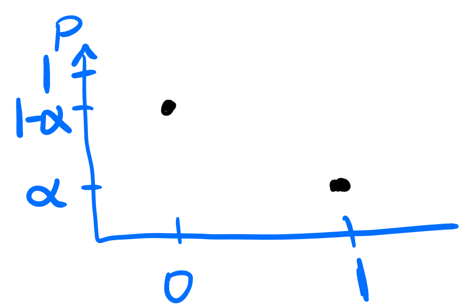
$$\mathbb{P}(X\in\{1,3,5\})=\sum_{x\in\{1,3,5\}}p(x)=p(1)+p(3)+p(5)=\tfrac{1}{2}$$

Discrete Probability Distributions with special names:

Bernoulli distribution (parameter: $\alpha\in[0,1]$):

Outcome space: $\{0,1\}$

pmf: $p(1)=\alpha,\ p(0)=1-\alpha$



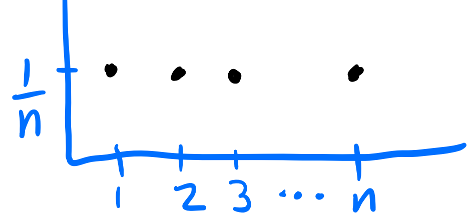
$$\boxed{\text{Distribution } \mathbb{P}=\text{Bernoulli}(\alpha)}$$

$\mathbb{P}(Z=1)=\mathbb{P}(Z\in\{1\})=p(1)=\alpha$ $\quad Z\in\{0,1\}$ is a "Bernoulli r.v."

Discrete Uniform Distribution (parameter: n):

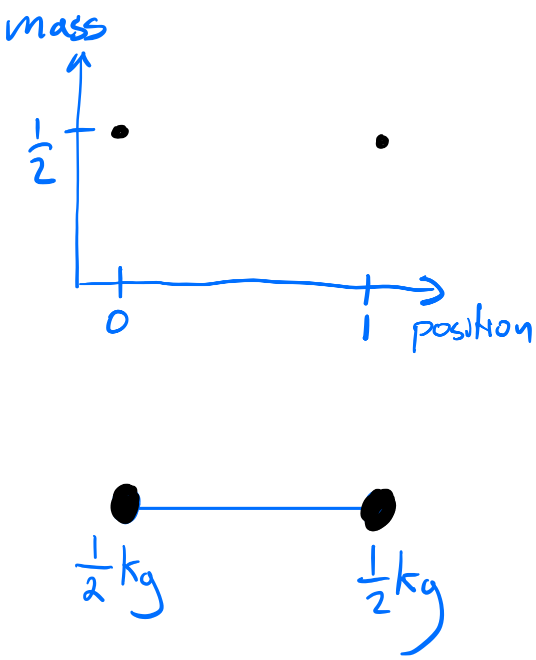
Outcome space: $\{1,2,\dots,n\}$

pmf: $p(1) = p(2) = \dots = p(n) = \frac{1}{n}$



$$\boxed{\text{Distribution } \mathbb{P} = \text{Uniform}(n)}$$

Intuition with a rod in physics



continuous $\Longrightarrow$