

Important Announcements

- assignment 1 is due this Friday (Jan 24)

- review session by TA on Monday

- notation : random variable $\rightarrow$ upper case, X

outcome space / set of a r.v. $\rightarrow$ calligraphic upper case, $\mathcal{X}$

outcome of a r.v. $\rightarrow$ lower case, x

$$X \in \mathcal{X} = \{x_1, x_2, \dots, x_n\}$$

$X \sim P \rightarrow$ "The r.v. X is sampled according to the distribution P "

typo: RV section $Y \in \mathcal{Y} \in \{0, 1\}$

- two ways to describe a discrete random experiment

- cont. r.v. statements

- question in past lecture: infinite outcome set possible?

Questions

(1) What is the difference btw. an event & outcome space?

(2) Property of a distribution: if $E_1 \subset \mathcal{X}$ & $E_2 \subset \mathcal{X}$
and $E_1 \cap E_2 = \{\}$

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

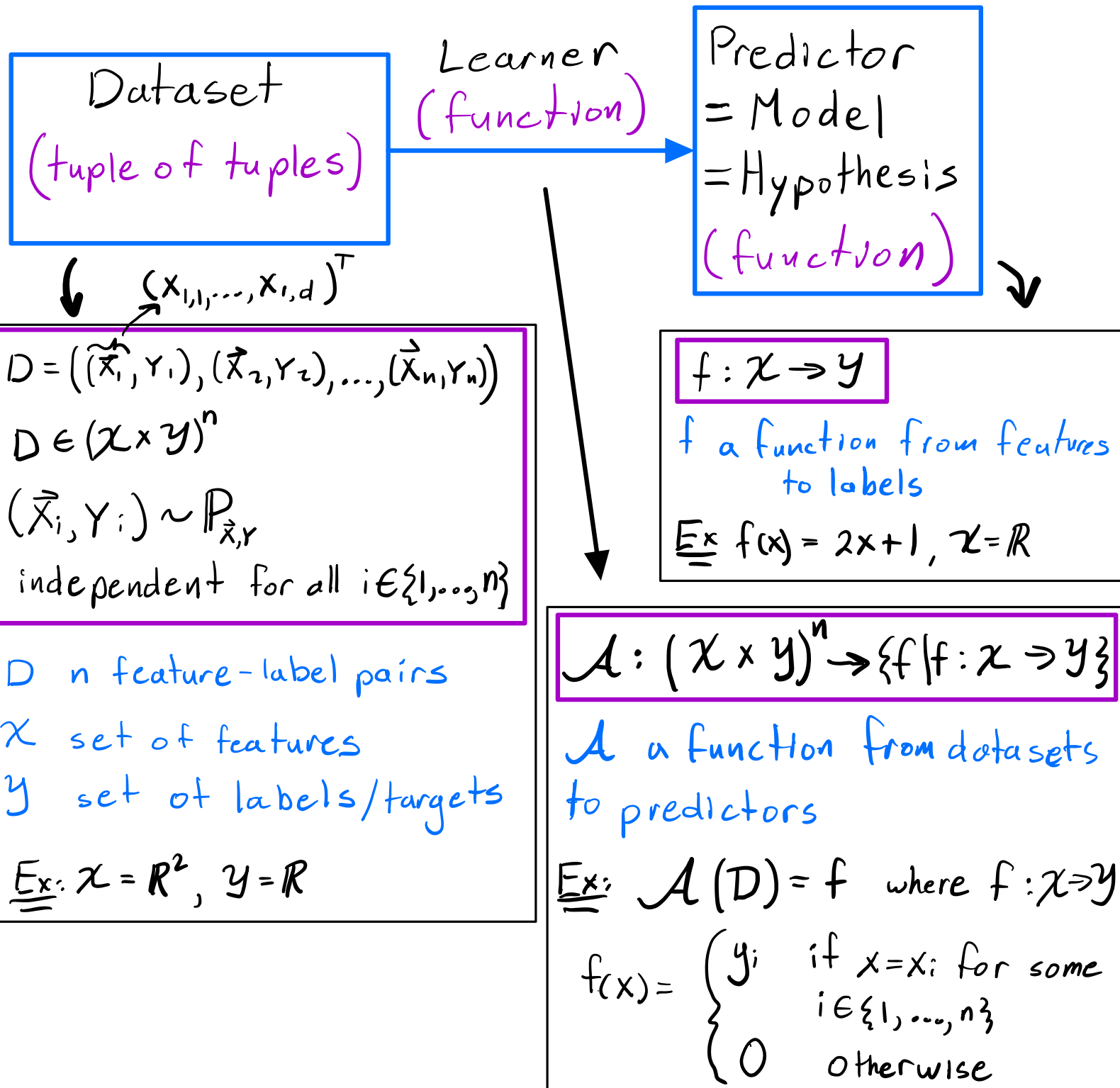
(3) What if $E_2 = E_1^c \rightarrow P(E_1 \cup E_2) = 1$

(4) Define a random variable.

"A variable that takes a value based on the outcome of an experiment & this value is associated with a probability distribution"

Motivation

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



Probability

Note: Humans have a bad intuition when it comes to randomness

- 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
- Thinking Fast and Slow
by: Daniel Kahneman

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: $y = \{0, 1\}$ ^{Heads} _{Tails} flipping a coin

$$X = \{1, 2, 3, 4, 5, 6\} \text{ rolling a dice}$$
 $[0, 900]$ $\mathbb{R}$

amount of a chemical in a wine
measurement error

The outcome space/set can be either a

- 1) countable set or
- 2) uncountable set

→ cardinality finite or countably infinite

→ cardinality is uncountably infinite

Event: A subset of the outcome space (imprecise)

Ex: Outcome space: $\mathcal{Y} = \{0, 1\}$ an outcome is a single element of the outcome set

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

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

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

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

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

Probability Distribution: A function P defining the likelihood of each event (and satisfying certain properties)

$P: \underbrace{\text{event space/set}}_{\text{a set containing all the events}} \rightarrow [0, 1]$

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

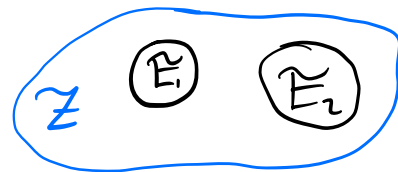
Properties: (imprecise)

Outcome space: $\mathcal{Z}$

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

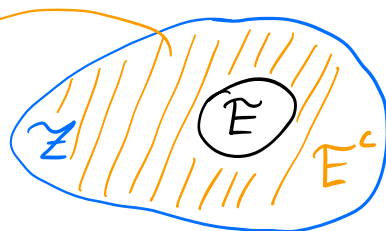
2. If $E_1 \subset \mathcal{Z}$, $E_2 \subset \mathcal{Z}$ and $E_1 \cap E_2 = \emptyset$, then

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$



E_x : (of property 2.)

Events: E , E^c



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

$$P(E \cup E^c) = P(E) + P(E^c)$$

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

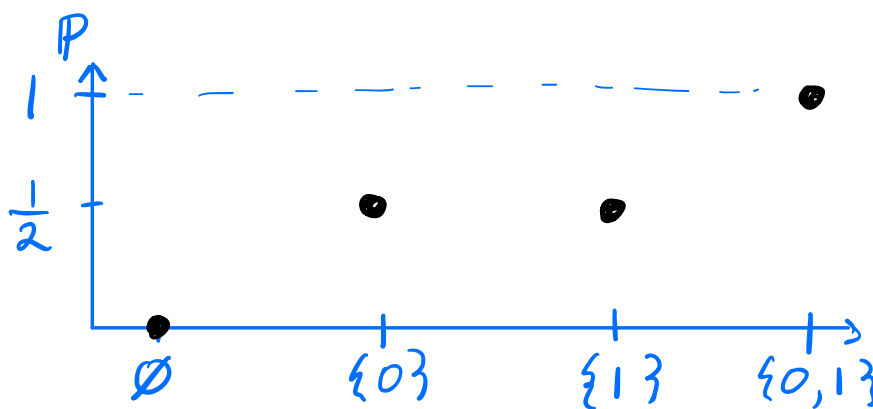
rearranging:

$$P(E) = 1 - P(E^c)$$

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

$$P(\emptyset) = 0, P(\mathcal{Y}) = 1, P(\{0\}) = \frac{1}{2}, P(\{1\}) = \frac{1}{2}$$

each event of
an experiment
is associated with
a probability



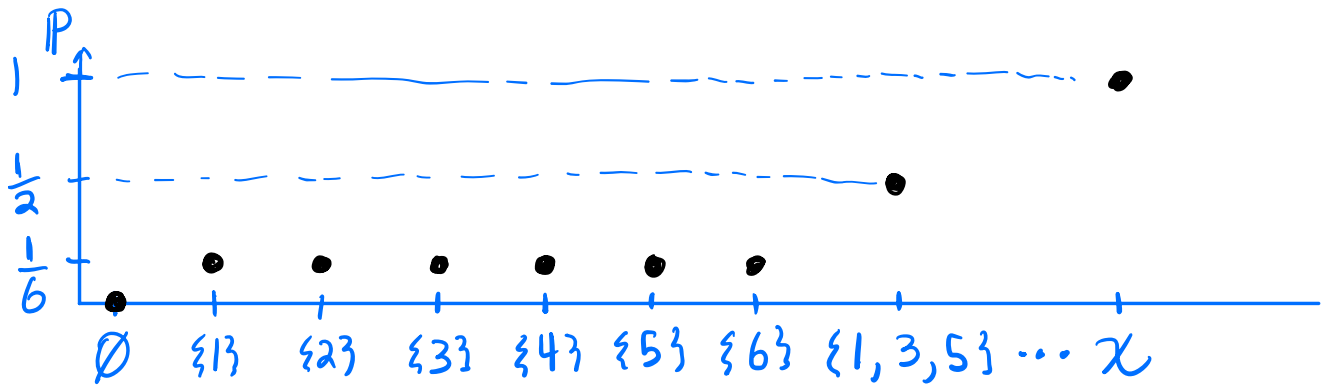
all events
of experiment

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

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

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

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



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.

Ex: $X \in \mathcal{X} = \{1, 2, 3, 4, 5, 6\}$ with IP from prev. example
 $Y \in \mathcal{Y} = \{0, 1\}$ with IP 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. 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.: $X \in \mathcal{X}$

$$P(\{1, 3, 5\}) \stackrel{\text{def}}{=} P(X \in \{1, 3, 5\})$$

"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 $P(X=1) = \dots = P(X=6) = \frac{1}{6}$ ← same discrete experiment

Probability 1 so $P(X \in \{1, 2, 3, 4, 5, 6\}) = 1$
for countable event $\{1, 2, 3, 4, 5, 6\}$ and $P(\mathbb{R} \setminus \{1, 2, 3, 4, 5, 6\}) = 0$

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

$\frac{1}{6}$



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

$x \in \mathbb{R}$

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: $Z \in [0, 100]$ and $P(Z=z) = P(Z \in \{z\}) = 0$ for all $z \in [0, 100]$
but $P(Z \in [0, 100]) = 1 =$

Ex: $Z \in \mathbb{R}$ and $P(Z=z) = P(Z \in \{z\}) = 0$ for all $z \in \mathbb{R}$
but $P(Z \in \mathbb{R}) = 1 =$

Calculating Probabilities

Motivation: It is hard to define the values of a probability distribution P for all the events

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

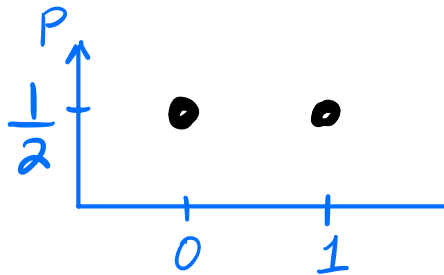
The probability of an event $E \subset \mathcal{Z}$ is:

$$P(Z \in E) \stackrel{\text{def}}{=} \sum_{z \in E} p(z)$$

where $Z \in \mathcal{Z}$

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

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



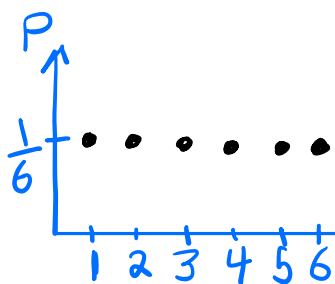
$$P(Y \in \{0, 1\}) = \sum_{y \in \{0, 1\}} p(y) = p(0) + p(1) = 1$$

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

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

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

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



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

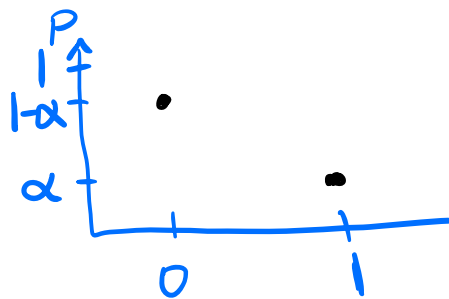
Discrete Probability Distributions with special names:

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

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

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

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



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

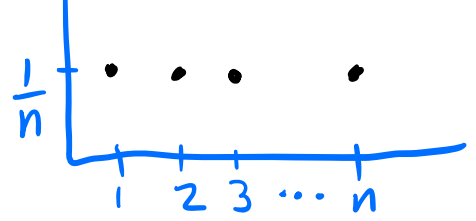
Discrete Uniform Distribution (parameter: n):

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

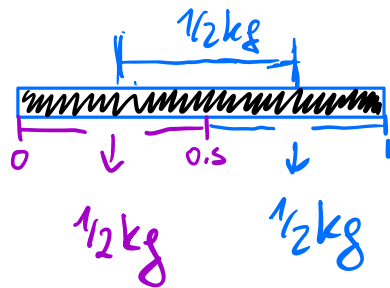
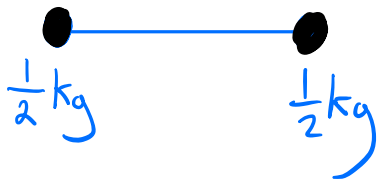
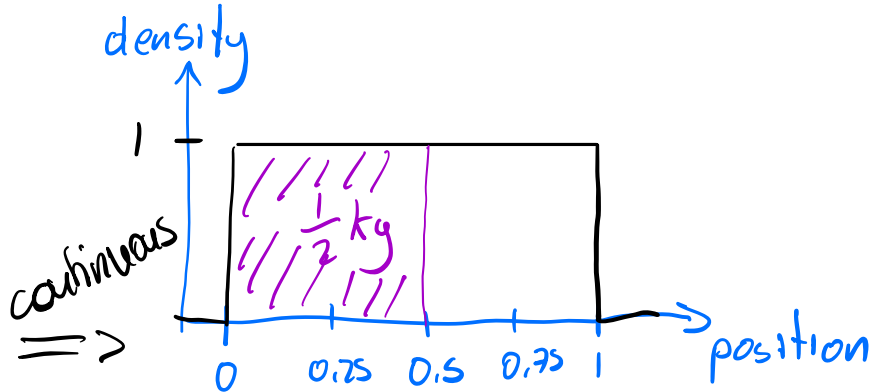
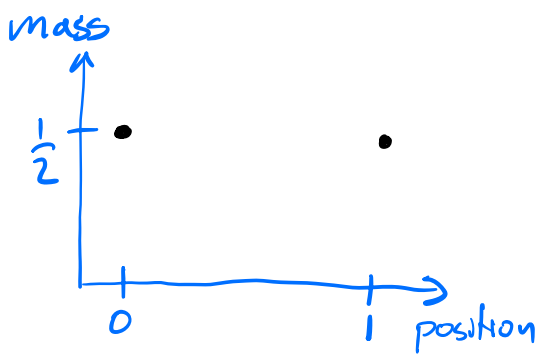
$P \uparrow$

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

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



Intuition with a rod in physics



Probability Density Function (pdf): a function $p: \mathbb{Z} \rightarrow [0, \infty]$ where $\mathbb{Z}$ is an uncountable outcome space and $\int_{\mathbb{Z}} p(z) dz = 1$

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

$$P(Z \in E) \stackrel{\text{def}}{=} \int_E p(z) dz$$

$$P(Z = z) = P(Z \in \{z\}) = 0 \quad \text{where } z \in \mathbb{Z}$$

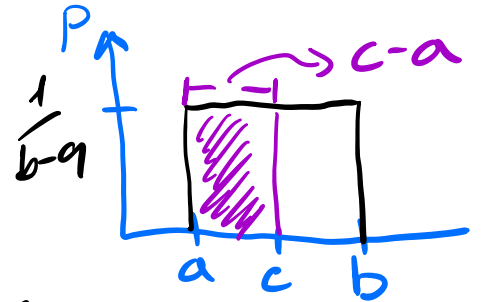
Continuous Probability Distributions with special names:

Continuous Uniform Distribution (parameters: $a \in \mathbb{R}, b \in \mathbb{R}$):

Outcome space: $[a, b]$

pdf: $p(z) = \frac{1}{b-a}$

Distribution $P = \text{Uniform}(a, b)$



$$P(a \leq Z \leq c) = P(Z \in [a, c]) = \int_a^c p(z) dz = \frac{c-a}{b-a}$$

where $a \leq c \leq b$

Gaussian/Normal Distribution (parameters: $\mu \in \mathbb{R}, \sigma^2 > 0$):

Outcome space: $\mathbb{R}$

standard normal dist.
 $\rightarrow \mu = 0, \sigma^2 = 1$

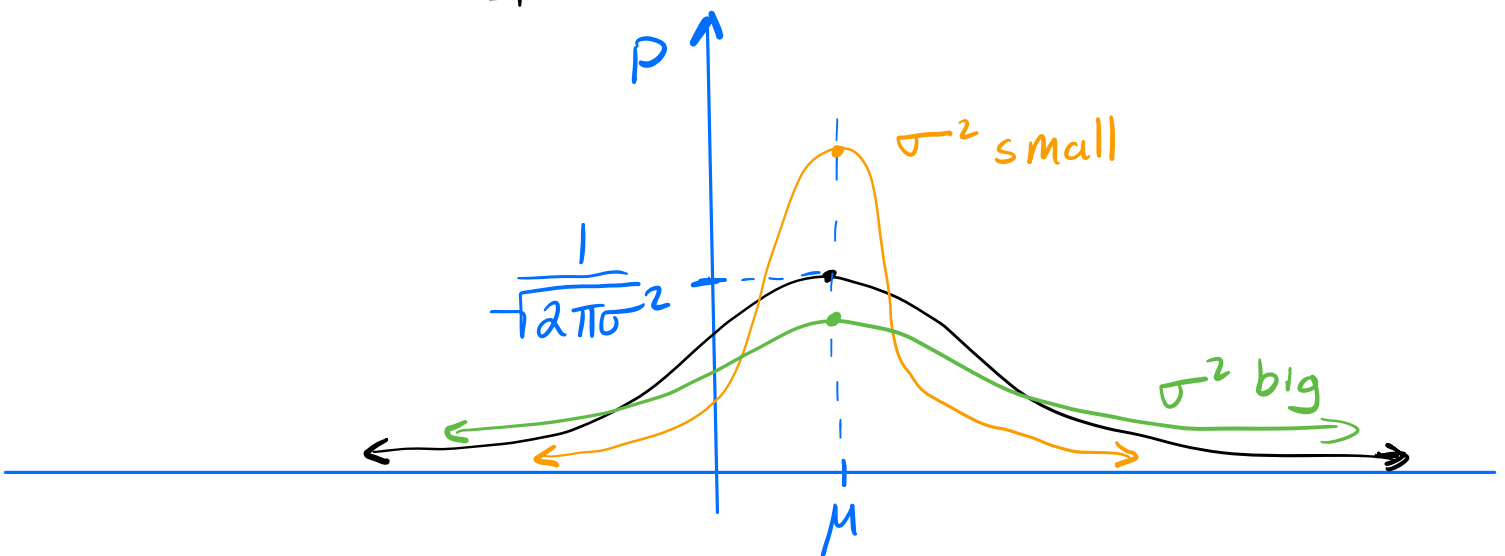
pdf: $p(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(z-\mu)^2\right)$

Distribution $P = \mathcal{N}(\mu, \sigma^2) = \text{Gaussian}(\mu, \sigma^2)$

mean
variance

$\sigma \rightarrow$ standard deviation

Ex $P(-1 \leq Z \leq 1) = \int_{-1}^1 p(z) dz$

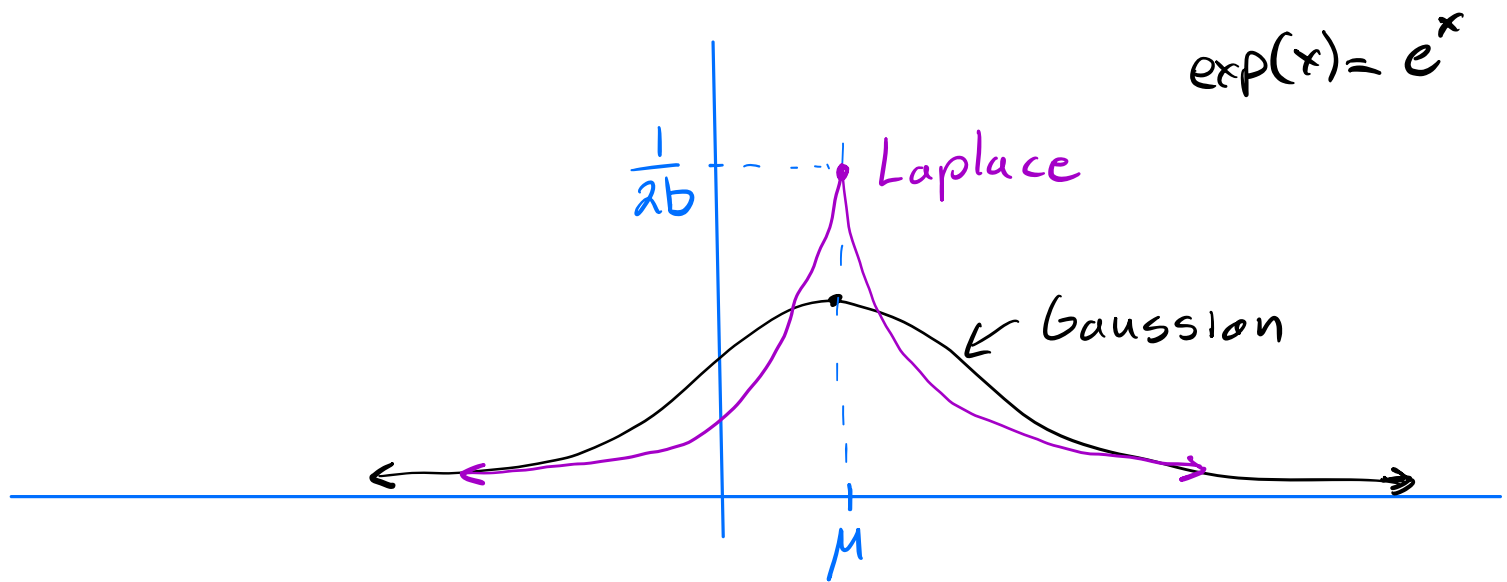


Laplace Distribution (parameters: $\mu \in \mathbb{R}, b > 0$):

Outcome space: $\mathbb{R}$

pdf: $p(x) = \frac{1}{2b} \exp(-\frac{1}{b}|x-\mu|)$

Distribution $P = \text{Laplace}(\mu, b)$



Multivariate Random Variables

Motivation: To be able to talk about the probability of different types of events (outcomes of different experiments) at the same time!

Ex: The probability of getting heads and rolling a 3

The probability of a wine containing 2.5mg of one chemical and 4mg of another chemical

The probability of a house having 4 rooms and 2 washrooms and being less than 10min from a university

The probability of being young and having arthritis

Multi variate Random Variable: A tuple of more than one random variable

Ex: (Flipping 2 coins)

Outcome space: $\mathcal{X} = \{0,1\} \times \{0,1\} = \{(0,0), (0,1), (1,0), (1,1)\}$

r.v.: $X = (X_1, X_2) \in \mathcal{X}$

(Collecting the info of one house (ex: # of rooms, age))

Outcome space: $\mathcal{X} = \mathbb{N} \times [0, \infty)$

r.v.: $X = (X_1, X_2) \in \mathcal{X}$

(Collecting the info of one house and its price)

Outcome space: $\mathcal{Z} = (\mathbb{N} \times [0, \infty)) \times [0, \infty)$

r.v.: $Z = (X, Y)$
 $= ((X_1, X_2), Y)$

Calculating Joint Probabilities

Ex: (If you have arthritis and if you are young or old)

Outcome space: $\mathcal{Z} = \{0,1\} \times \{0,1\} = \{(0,0), (0,1), (1,0), (1,1)\}$

r.v.: $Z = (X, Y) \in \mathcal{Z}$

pmf: $p: \mathcal{Z} \rightarrow [0, 1]$

Not based on real data $\left\{ \begin{array}{l} p(0,0) = p(0,1) = \frac{1}{2} \\ p(1,0) = \frac{1}{10}, p(1,1) = \frac{39}{100} \end{array} \right.$

		Y (arthritis)	
		0	1
X (age)	0	$\frac{1}{2}$	$\frac{1}{100}$
	1	$\frac{1}{10}$	$\frac{39}{100}$

$\frac{1}{2}$	$\frac{1}{10} (1,0)$
$(0,0)$	$\frac{39}{100} (1,1)$
	$\frac{1}{100} (0,1)$

$$\sum_{z \in \mathcal{Z}} p(z) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) = \frac{1}{2} + \frac{1}{100} + \frac{1}{10} + \frac{39}{100} = 1 \quad \left(\text{for a valid pmf} \right)$$

What is the probability of being young (i.e. $X=0$)?

$$\mathcal{E} = \{(0,0), (0,1)\} = \{0\} \times \{0,1\} \subset \mathcal{Z}$$

$$\begin{aligned} P(X=0, Y \in \{0,1\}) &= P(Z \in \mathcal{E}) = \sum_{z \in \mathcal{E}} p(z) \\ &= \sum_{x \in \{0\}} \sum_{y \in \{0,1\}} p(x,y) \\ &= p(0,0) + p(0,1) \\ &= \frac{1}{2} + \frac{1}{100} = \frac{51}{100} \end{aligned}$$

Marginal Distribution: The distribution over a subset of random variables

$\mathbb{E}_x$: Continuing with the arthritis example

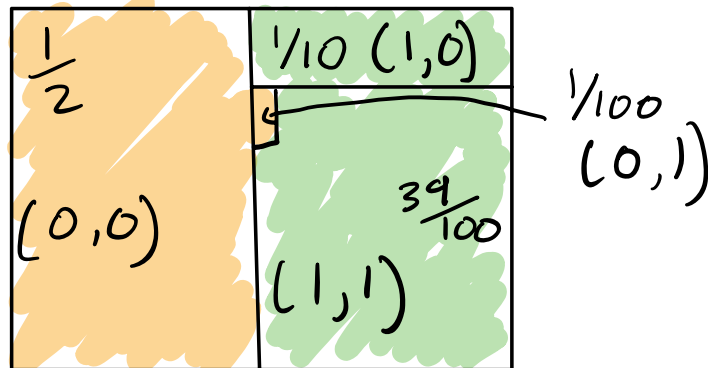
Marginal Distribution: $P_x(X \in E_x)$ where $E_x \subset \mathcal{X}$

Marginal pmf: $p_x: \mathcal{X} \rightarrow [0,1]$, $p_x(x) = \sum_{y \in \mathcal{Y}} p(x,y)$

$$P_x(X=0) \stackrel{P(X \in \{0\})}{=} p_x(0) = \sum_{x \in \{0\}} \sum_{y \in \mathcal{Y}} p(x,y) = \frac{51}{100}$$

$$P_x(X=1)$$

$$= \frac{1}{10} + \frac{39}{100} = \frac{49}{100}$$



Discrete r.v. $X_1, \dots, X_d$

$X = (X_1, \dots, X_d) \in \mathcal{X}_1 \times \dots \times \mathcal{X}_d = \mathcal{X}$, $p: \mathcal{X} \rightarrow [0,1]$

$P_{x_i}: \mathcal{X}_i \rightarrow [0,1]$, $i \in \{1, \dots, d\}$

"Sum over every element except x_i "

Marginal pmf:

$$p_{x_i}(x_i) \stackrel{\text{def}}{=} \sum_{x_1 \in \mathcal{X}_1} \dots \sum_{x_{i-1} \in \mathcal{X}_{i-1}} \sum_{x_{i+1} \in \mathcal{X}_{i+1}} \dots \sum_{x_d \in \mathcal{X}_d} p(x_1, \dots, \cancel{x_i}, \dots, x_d)$$

$$P_{x_i}(X_i \in E_i) = \sum_{x_i \in E_i} p_{x_i}(x_i)$$

where $E_i \subset \mathcal{X}_i$

Continuous r.v. $X_1, \dots, X_d$

$X = (X_1, \dots, X_d) \in \mathcal{X}_1 \times \dots \times \mathcal{X}_d = \mathcal{X}$, $p: \mathcal{X} \rightarrow [0, \infty)$

$$P_{X_i}: \mathcal{X}_i \Rightarrow [0, \infty), \quad i \in \{1, \dots, d\}$$

Marginal pdf:
$$p_{X_i}(x_i) = \int \dots \int_{\mathcal{X}_1} \int \dots \int_{\mathcal{X}_{i-1}} \int \dots \int_{\mathcal{X}_{i+1}} p(x_1, \dots, x_d) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_d$$

Distribution:
$$P_{X_i}(X_i \in \mathcal{E}_i) = \int_{\mathcal{E}_i} p_{X_i}(x_i) dx_i \quad \text{where } \mathcal{E}_i \subset \mathcal{X}_i$$

Conditional Distribution: Probability of a r.v. given info about another r.v.

Ex: Probability that I have arthritis given I am young

Let r.v. = $Y \in \mathcal{Y}, X \in \mathcal{X}$

Discrete Y for any $x \in \mathcal{X}$ that $p_X(x) \neq 0$

$$P_{Y|X=x}: \mathcal{Y} \Rightarrow [0, 1], \quad P_{Y|X=x}(Y) = P_{Y|X}(Y|x)$$

conditional pmf:
$$P_{Y|X}(y|x) \stackrel{\text{def}}{=} \frac{p(y, x)}{p_X(x)} \quad \text{implies} \quad \sum_{y \in \mathcal{Y}} P_{Y|X}(y|x) = 1$$

Distribution:
$$P_{Y|X}(Y \in \mathcal{E}_Y | X=x) \stackrel{\text{def}}{=} \sum_{Y \in \mathcal{E}_Y} P_{Y|X}(Y|x) \quad \text{where } \mathcal{E}_Y \subset \mathcal{Y}$$

Continuous Y for any $x \in \mathcal{X}$ that $p_X(x) \neq 0$

$$P_{Y|X=x}: \mathcal{Y} \Rightarrow [0, \infty), \quad P_{Y|X=x}(Y) = P_{Y|X}(Y|x)$$

conditional
pdf:

$$P_{Y|X}(y|x) \stackrel{\text{def}}{=} \frac{p(y, x)}{p_X(x)}$$

$$\text{implies } \int_Y P_{Y|X}(y|x) dy = 1$$

Distribution:

$$P_{Y|X}(Y \in E | X=x) \stackrel{\text{def}}{=} \int_E P_{Y|X}(y|x) dy \quad \text{where } E_Y \subset Y$$

Product Rule:

$$p(x, y) = p(x|y)p(y) = p(y|x)p(x)$$

More generally:

$$p(x_1, x_2, x_3) = p(x_1, x_2 | x_3) p(x_3) = p(x_1 | x_2, x_3) p(x_2 | x_3) p(x_3)$$

for d-dimensional r.v. for 3-dimensional r.v.

$$p(x_1, x_2, \dots, x_d) = p(x_d | x_1, \dots, x_{d-1}) \dots p(x_3 | x_1, x_2) p(x_2 | x_1) p(x_1)$$

Bayes' Rule:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

Note: Sometimes the subscripts are not used for marginal and conditional distributions when it is clear from the context

$$p(x, y) = p_{x,y}(x, y)$$

$$p(x) = p_X(x)$$

$$p(y|x) = p_{Y|X}(y|x)$$

E_x : Probability that I have arthritis given I am young

$$P(Y=1 | X=0) = \sum_{Y \in \{1\}} P_{Y|X}(Y|0)$$

Arthritis Young

$\frac{1}{2}$ $(0,0)$	$\frac{1}{100}(1,0)$ $\frac{39}{100}(1,1)$
--------------------------	---

$\frac{1}{100}(0,1)$

condition on being young

$Y=0$ $\frac{50}{51}$

$\frac{1}{51}, Y=1$

$$\begin{aligned}
 &= P_{Y|X}(1|0) \\
 &= \frac{P_{XY}(0,1)}{P_X(0)} \\
 &= \frac{P(0,1)}{P(0,1) + P(0,0)} \\
 &= \frac{1/100}{1/100 + 1/2} \\
 &= \frac{1/100}{51/100} = \frac{1}{51}
 \end{aligned}$$

product rule from above
 $P(X,Y) = P(X|Y)P(Y)$

Ex: Probability of being young given I have arthritis

$$P(X=0|Y=1) = P_{X|Y}(0|1)$$

Bayes' Rule

$$\begin{aligned}
 &= \frac{P_{Y|X}(1|0) P_X(0)}{P_Y(1)} \\
 &= \frac{\frac{1}{51} \frac{51}{100}}{40/100} = \frac{1}{40}
 \end{aligned}$$

Independence: Changing the value of one r.v. doesn't affect the probability of another r.v.