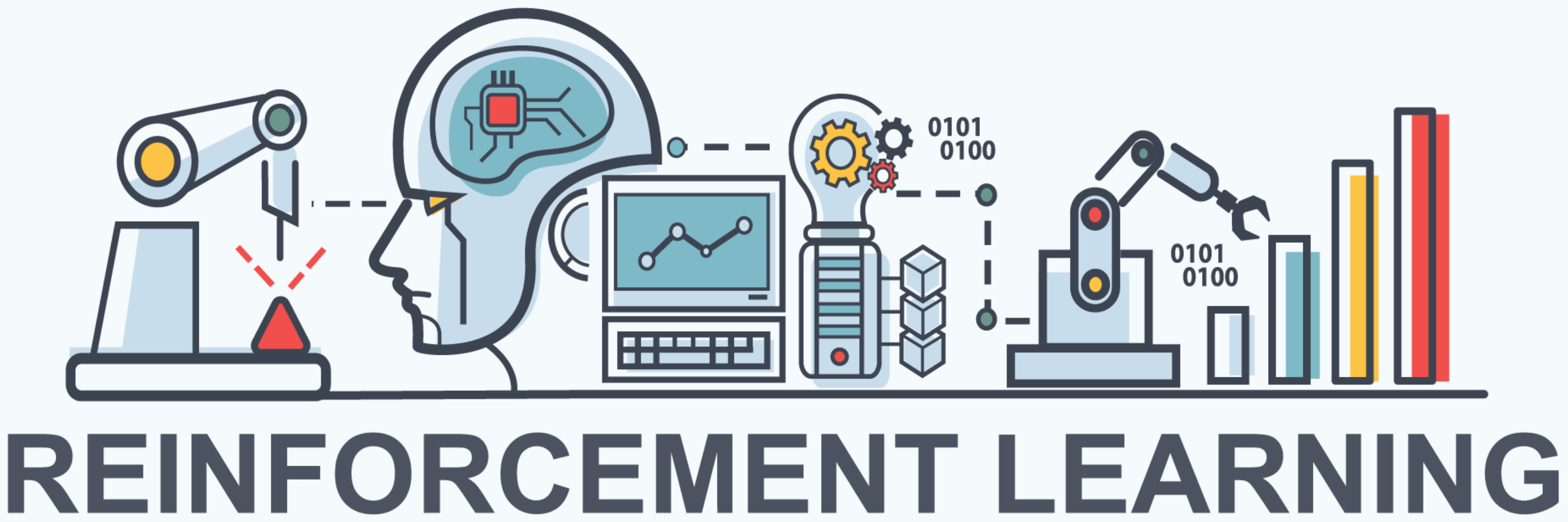


CP8319/CPS824
Lecture 2

Instructor: Nariman Farsad

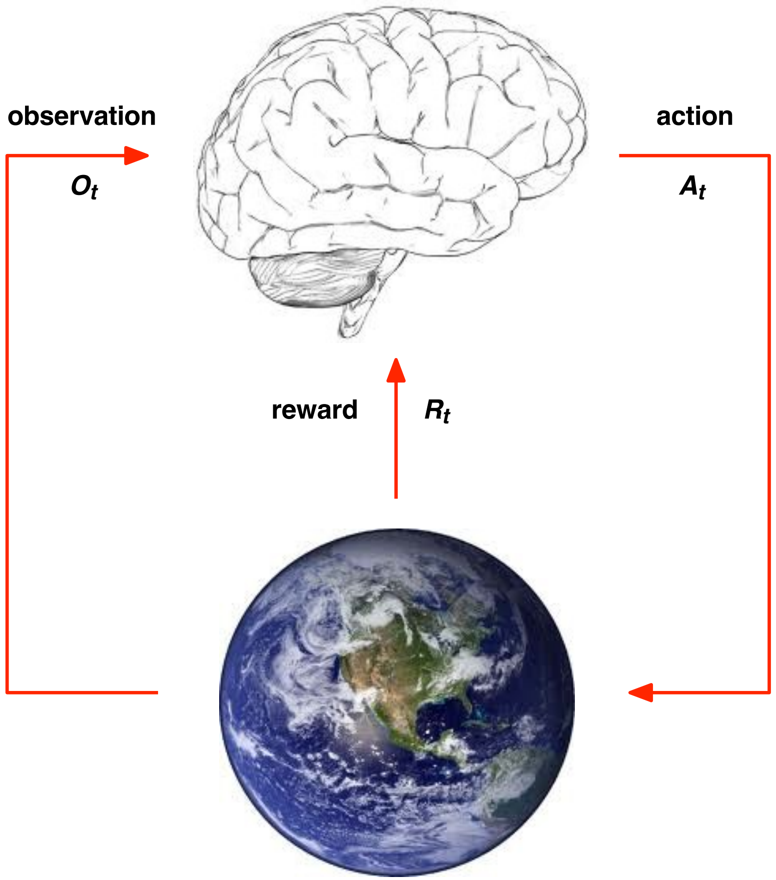
* Some of the slides in this deck are adopted from courses offered David Silver, Emma Brunskill, and Sergey Levine.

Today’s Agenda

1. Introduction to Reinforcement Learning Review

2. Quick Review of Probability

RL: The Agent and the Environment



- At each step t the agent:
  - Executes action A_t
  - Receives observation O_t
  - Receives scalar reward R_t
- The environment:
  - Receives action A_t
  - Emits observation O_{t+1}
  - Emits scalar reward R_{t+1}
- t increments at env. step

Characteristics of RL

What makes reinforcement learning different from other machine learning paradigms?

- There is no supervisor, only a *reward* signal
- Feedback is delayed, not instantaneous
- Time really matters (sequential, non i.i.d data)
- Agent’s actions affect the subsequent data it receives

History and State

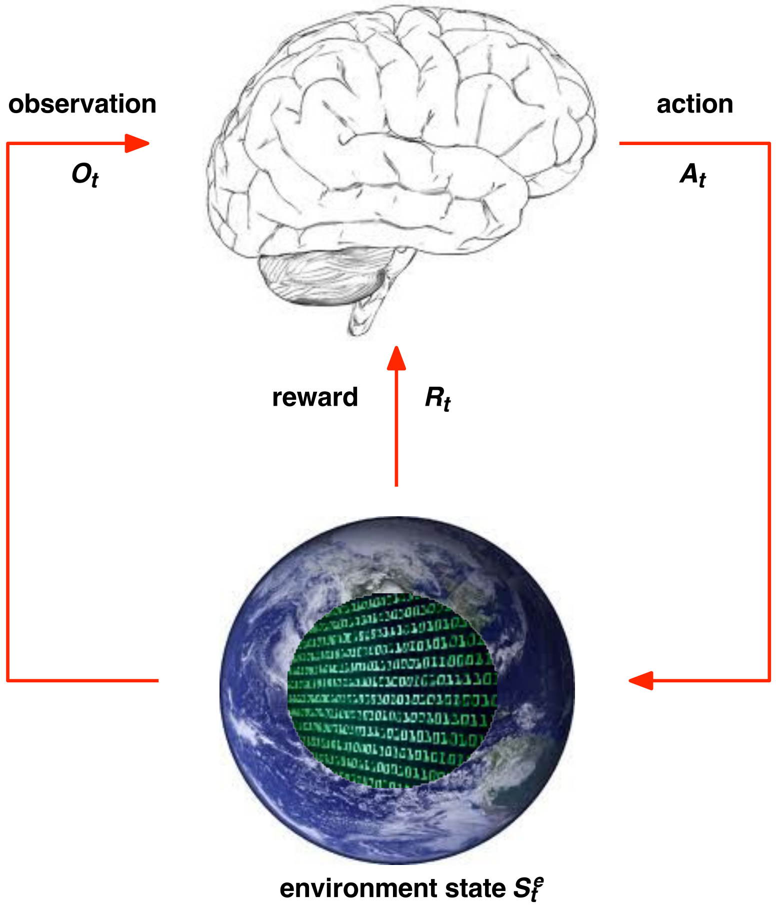
- The history is the sequence of observations, actions, rewards

$$H_t = O_1, R_1, A_1, ..., A_{t-1}, O_t, R_t$$

- i.e. all observable variables up to time t
- i.e. the sensorimotor stream of a robot or embodied agent What happens next depends on the history:
  - The agent selects actions
  - The environment selects observations/rewards
- The State is the information used to determine what happens next Formally, state is a function of the history:

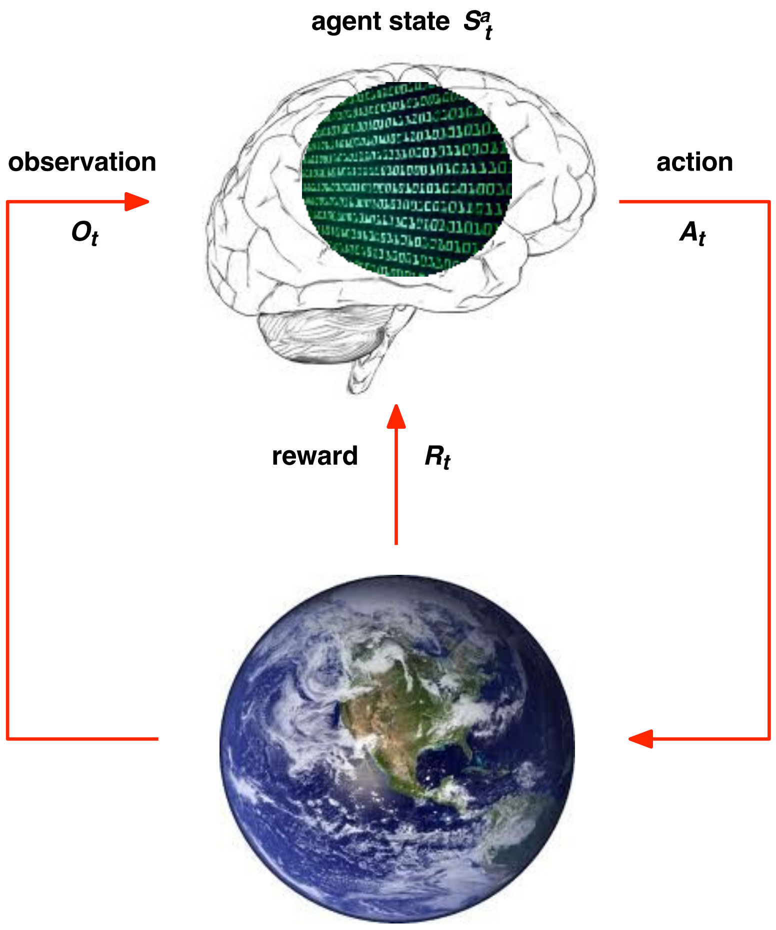
$$S_t = f(H_t)$$

Environment State



- The environment state S_t^e is the environment's private representation
- The environment uses the state to pick the next observation/reward
- The environment state is not usually visible to the agent directly
- Even when S_t^e is the visible it may contain irrelevant information

Agent State



- The agent state S^a_t is the agent's internal representation
- i.e. whatever information the agent uses to pick the next action
- i.e. it is the information used by reinforcement learning algorithms
- It can be any function of history:

$$S^a_t = f(H_t)$$

Today's Agenda

1. Introduction to Reinforcement Learning Review

2. **Quick Review of Probability**

Why Probability in RL?

- Often state of the environment and the agent are uncertain (e.g., due to noisy sensors)
  - Probability provides a framework to model and handle these uncertainties
  - Result: probability distribution over possible states of agent and environment

- Dynamics of environment and agent are often stochastic hence can’t optimize for a particular outcome, but only optimize to obtain a good distribution over outcomes
  - Probability provides a framework to reason in this setting
  - Result: ability to find good decision policies for stochastic dynamics and environments

Example: Flying Helicopter

- State: position, orientation, velocity, angular rate
- Sensors:
  - GPS : noisy estimate of position (sometimes also velocity)
  - Inertial sensing unit: noisy measurements from
    1. (i) 3-axis gyro [=angular rate sensor],
    2. (ii) 3-axis accelerometer [=measures acceleration + gravity; e.g., measures (0,0,0) in free-fall],
    3. (iii) 3-axis magnetometer
- Dynamics:
  - Noise from: wind, unmodeled dynamics in engine, servos, blades

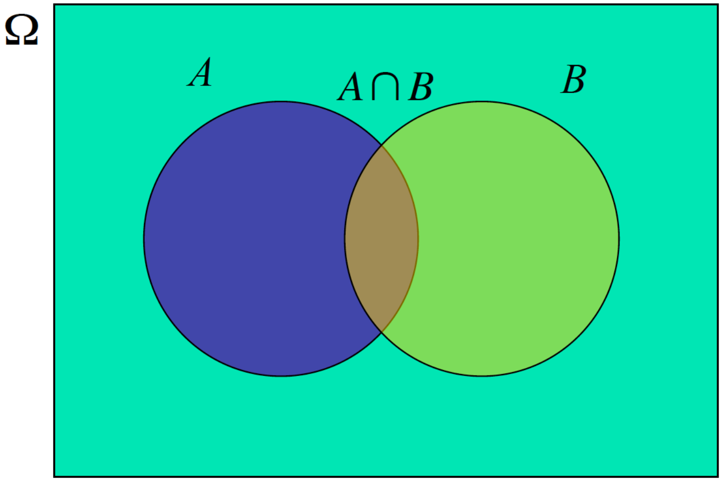
Sample space and Events

- Ω : Sample Space, result of an experiment
  - If you toss a coin twice $\Omega = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$
- Event: a subset of Ω
  - First toss is head = {HH,HT}
- $\mathcal{F}$: event space, a set of events:
  - Closed under finite union and complements
    - Entails other binary operation: union, diff, etc.
  - Contains the empty event and Ω

Probability Measure

- Defined over $(\Omega, \mathcal{F})$ s.t.
  - P(A) >= 0 for all *A* in $\mathcal{F}$
  - P(Ω) = 1
  - If *A*, *B* are disjoint, then
    - P(*A* U *B*) = p(*A*) + p(*B*)
- We can deduce other axioms from the above ones
  - Ex: P(*A* U *B*) for non-disjoint event

    P(*A* U *B*) = p(*A*) + p(*B*) – p(*A* ∩ *B*)



Conditional Probability and Independence

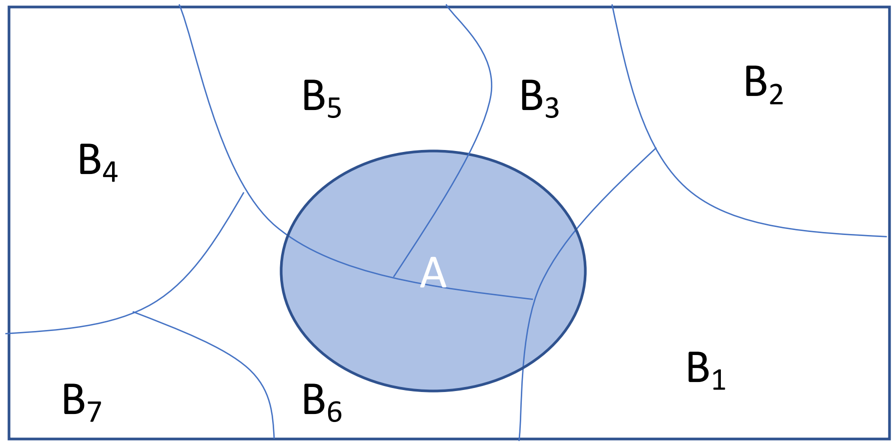
Let B be any event such that $P(B) \neq 0$.

$P(A|B) := \frac{P(A \cap B)}{P(B)}$

$A \perp B$ if and only if $P(A \cap B) = P(A)P(B)$

$A \perp B$ if and only if $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$

Rule of total probability

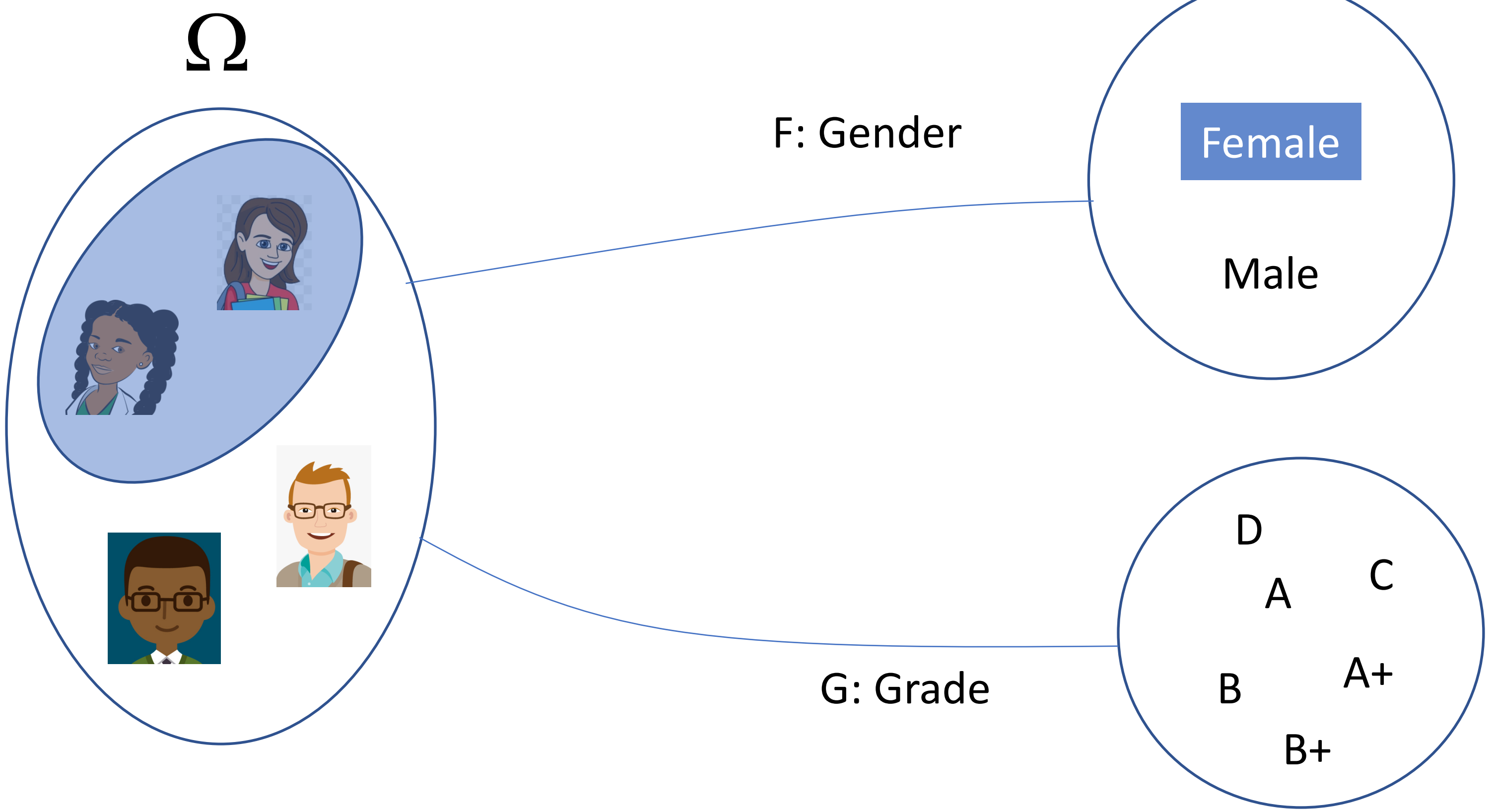


$$p(A) = \sum P(B_i)P(A \mid B_i)$$

From Events to Random Variable

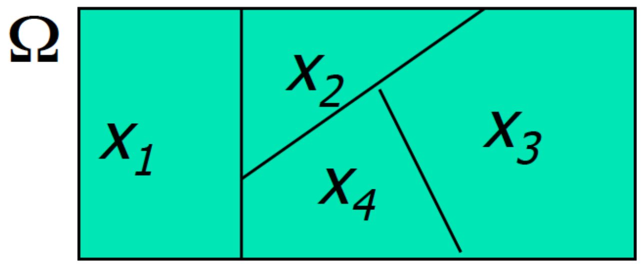
- Almost all the semester we will be dealing with random variables (RV)
- Concise way of specifying attributes of outcomes
- Modeling students (Grade and Gender):
  - $\Omega =$ all possible students
  - Example of some events
    - Grade_A = all students with grade A
    - Gender_F = all female students
  - Very cumbersome
  - We need “functions” that maps from Ω to an attribute space.
  - P(G = A) = P({student ϵ Ω : G(student) = A})

Random Variables



P(F = Female) = P({all students who identify as females})

Discrete Random Variable



- X denotes a random variable.
- X can take on a countable number of values in $\{x_1, x_2, \ldots, x_n\}$.
- $P(X=x_i)$, or $P(x_i)$, is the probability that the random variable X takes on value x_i.
- $P(\cdot)$ is called probability mass function.
- *E.g., X models the outcome of a coin flip*, x_1 = head, x_2 = tail, $P(x_1) = 0.5$, $P(x_2) = 0.5$

Probability of Discrete RV

- Probability mass function (pmf): $P(X = x_i)$
- Easy facts about pmf
  - $\Sigma_i \; P(X = x_i) = 1$
  - $P(X = x_i \cap X = x_j) = 0$ if $i \neq j$
  - $P(X = x_i \cup X = x_j) = P(X = x_i) + P(X = x_j)$ if $i \neq j$
  - $P(X = x_1 \cup X = x_2 \cup \ldots \cup X = x_k) = 1$

Common Distributions

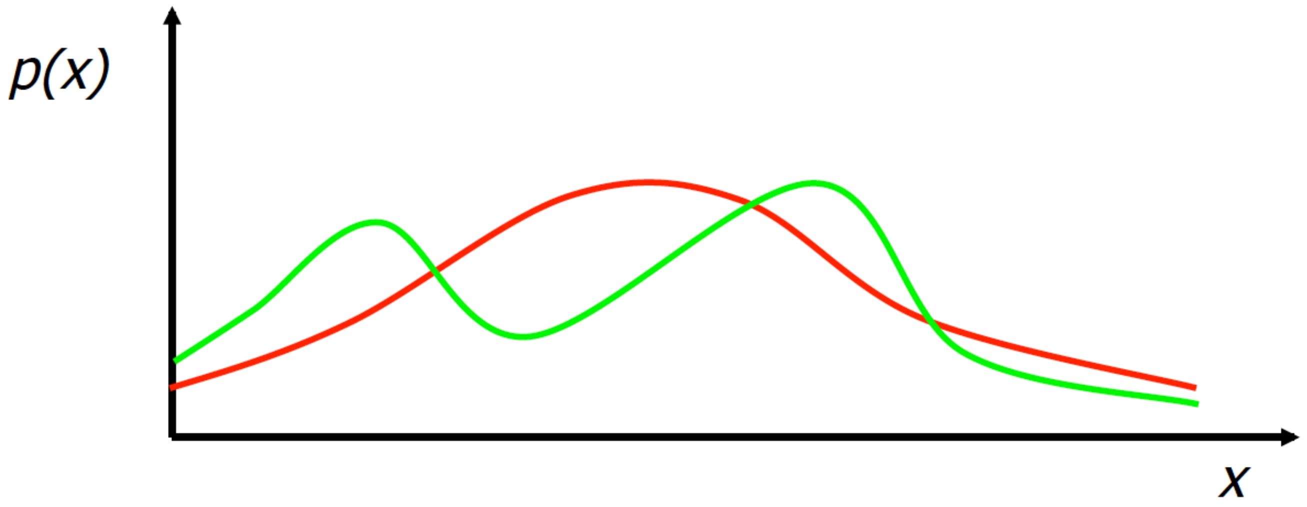
- Uniform X $U[1, \ldots, N]$
  - X takes values 1, 2, … N
  - P(X = i) = 1/N
  - E.g. picking balls of different colors from a box
- Binomial X $Bin(n, p)$
  - X takes values 0, 1, …, n
  - $$p(X = i) = \binom{n}{i} p^i (1 - p)^{n-i}$$
  - E.g. number of head in n coin flips

Continuous Random Variable

- X takes on values in the continuum.
- $p(X=x)$, or $p(x)$, is a probability density function.

$$\Pr(x \in (a,b)) = \int_a^b p(x)\,dx$$

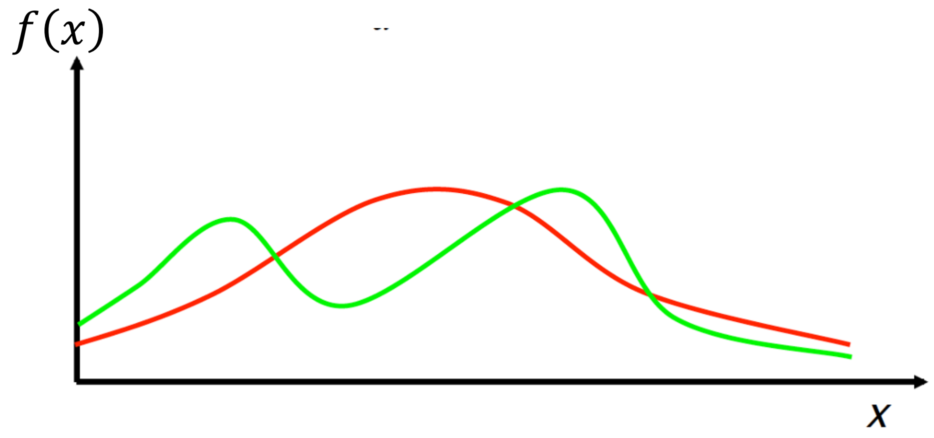
- E.g.



Probability of Continuous RV

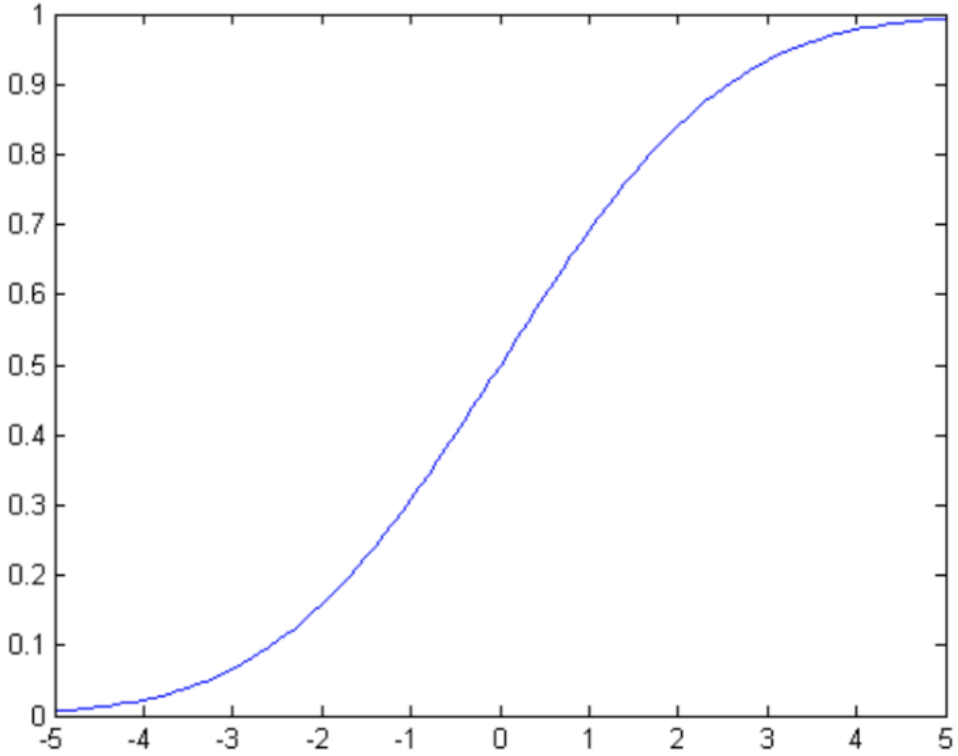
- The RV X takes values in the continuum
- Properties of probability density function (pdf)
  - $f(x) \geq 0, \forall x$
  - $\int_{-\infty}^{\infty} f(x)dx = 1$
- Actual probability can be obtained by taking the integral of pdf
  - E.g. the probability of X being between 0 and 1 is

$$P(0 \leq X \leq 1) = \int_0^1 f(x)dx$$



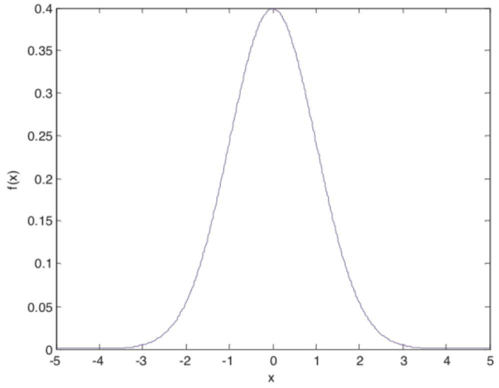
Cumulative Distribution Function (cdf)

- $F_X(v) = P(X \leq v)$
- Discrete RVs
  - $F_X(v) = \sum_{v_i \leq v} P(X = v_i)$
- Continuous RVs
  - $F_X(v) = \int_{-\infty}^{v} f(x)dx$
  - Derivative of cdf is pdf
    - $\frac{d}{dx} F_X(x) = f(x)$



Common Distributions

- Normal X $N(\mu, \sigma^2)$
  - $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right)$$
  - E.g. the height of the entire population



Multivariate Normal

- Generalization to higher dimensions of the one-dimensional normal

$x \in \mathrm{I\!R}^{\mathrm{n}}$. Model $p(x_1), p(x_2), \ldots.etc.$ at the same time. Parameters
$: \mu \in \mathrm{I\!R}^{\mathrm{n}}, \Sigma \in \mathrm{I\!R}^{\mathrm{n \times n}} (\mathrm{covariancematrix})$

$$p(x; \mu, \Sigma) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{\frac{1}{2}}} exp(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu))$$

Mean

Covariance
matrix

Joint Probability Distribution

- What if we have more than 1 RV?
- Joint probability distributions quantify this
- P(X= x, Y= y) = P(x, y)
  - E.g. P(Grade = A and Gender = Male)

- The joint probability distribution satisfies

$$\sum_x \sum_y P(X = x, Y = y) = 1$$

$$\int_x \int_y f_{X,Y}(x, y) dx dy = 1$$

- Generalizes to N-RVs

Chain Rule

- Always true
  - $P(x, y, z) = p(x)\, p(y|x)\, p(z|x, y)$
    $= p(z)\, p(y|z)\, p(x|y, z)$
    $= \ldots$

Marginalization

- We know p(X, Y), what is P(X)?
- We can use the low of total probability

$$p(x) = \sum_{y} P(x, y)$$
$$= \sum_{y} P(y) P(x \mid y)$$

- Another example

$$p(x) = \sum_{y,z} P(x, y, z)$$
$$= \sum_{z,y} P(y, z) P(x \mid y, z)$$

Conditional Probability

- Given that RV Y what is the probability of RV X
  - You read probability of X given Y

events

$$\mathrm{P}\left(\mathrm{X}=x \middle| \mathrm{Y}=y\right)=\frac{\mathrm{P}\left(\mathrm{X}=x \cap \mathrm{Y}=y\right)}{\mathrm{P}\left(\mathrm{Y}=y\right)}$$

But we will always write it this way:

$$P(x \mid y)=\frac{p(x, y)}{p(y)}$$

Bayes' Theorem

- Given the conditional probability of an event $P(x|y)$
- Want to find the "reverse" conditional probability, $P(y|x)$

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

where: $P(x) = \sum_{y' \in value\ y} P(x|y')P(y')$

X and Y are continuous

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

where: $f(x) = \int_{y' \in value\ y} f(x|y')f(y')dy'$

Example

- You randomly choose a treasure chest to open, and then randomly choose a coin from that treasure chest. If the coin you choose is gold, then what is the probability that you choose chest A?

$$a)\frac{1}{3} \qquad b)\frac{2}{3} \qquad c)1 \qquad d)None$$

Bayes Rule cont.

- You can condition on more variables

$$P\left(x \mid y, z\right) = \frac{P(x \mid z) P(y \mid x, z)}{P(y \mid z)}$$

Independence

- X is independent of Y means that knowing Y does not change our belief about X.
  - $P(X|Y=y) = P(X)$
  - $P(X=x, Y=y) = P(X=x)\, P(Y=y)$
  - The above should hold for all x, y
  - It is symmetric and written as $X \perp Y$

Independence

- $X_1, \ldots, X_n$ are independent if and only if

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{i=1}^{n} P\left(X_i \in A_i\right)$$

- If $X_1, \ldots, X_n$ are independent and identically distributed we say they are *iid* (or that they are a random sample) and we write

$$X_1, \ldots, X_n \sim P$$

Independence: Example

- Spin a spinner numbered 1 to 7, and toss a coin. What is the probability of getting an odd. number on the spinner and a tail on the coin?

$$p_{XY}(x, y) = p_X(x)p_Y(y) = \frac{1}{2} \times \frac{4}{7} = \frac{2}{7}$$

CI: Conditional Independence

- RV are rarely independent but we can still leverage local structural properties like Conditional Independence.
- $X \perp Y \mid Z$ if once Z is observed, knowing the value of Y does not change our belief about X
  - P(rain $\perp$ sprinkler’s on | cloudy)
  - P(rain $\perp$ sprinkler’s on | wet grass)

Conditional Independence

- $P(X{=}x \mid Z{=}z, Y{=}y) = P(X{=}x \mid Z{=}z)$
- $P(Y{=}y \mid Z{=}z, X{=}x) = P(Y{=}y \mid Z{=}z)$
- $P(X{=}x, Y{=}y \mid Z{=}z) = P(X{=}x \mid Z{=}z)\, P(Y{=}y \mid Z{=}z)$

We call these factors : very useful concept !!

Mean or Expectation

- Mean (Expectation): $\mu = E(X) = \mathbb{E}[X]$
  - Discrete RVs:

$$E(\mathrm{X}) = \sum_{v_i} v_i \mathrm{P}(\mathrm{X} = v_i)$$

$$E(g(X)) = \sum_{v_i} g(v_i) P(X = v_i)$$

  - Continuous RVs:

$$E(\mathrm{X}) = \int_{-\infty}^{+\infty} x f(x)\, dx$$

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f(x)\, dx$$

Variance and Covariance

- Variance: $\sigma^2 = Var(X) = V(X) = E((X-\mu)^2)$

$$= E(X^2) - \mu^2$$

  - Discrete RVs:

$$V(\mathrm{X}) = \sum_{v_i} (v_i - \mu)^2 \, \mathrm{P}(\mathrm{X} = v_i)$$

  - Continuous RVs:

$$V(\mathrm{X}) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx$$

- Covariance:

$$Cov(X,Y) = E((X-\mu_x)(Y-\mu_y)) = E(XY) - \mu_x \mu_y$$

Properties

- Mean
  - $E(\mathrm{X}+\mathrm{Y}) = E(\mathrm{X}) + E(\mathrm{Y})$
  - $E(a\mathrm{X}) = aE(\mathrm{X})$
  - If X and Y are independent, $E(\mathrm{XY}) = E(\mathrm{X}) \cdot E(\mathrm{Y})$
- Variance
  - $V(a\mathrm{X}+b) = a^2 V(\mathrm{X})$
  - If X and Y are independent, $V(\mathrm{X}+\mathrm{Y}) = V(\mathrm{X}) + V(\mathrm{Y})$

Some more properties

- The conditional expectation of Y given X when the value of X = x is:

$$E(Y|X = x) = \int y \cdot p(y|x)dy$$

- The Law of Total Expectation or Law of Iterated Expectation:

$$E(Y) = E\left[E(Y \mid X)\right] = \int E(Y \mid X = x)\, p_X(x)dx$$

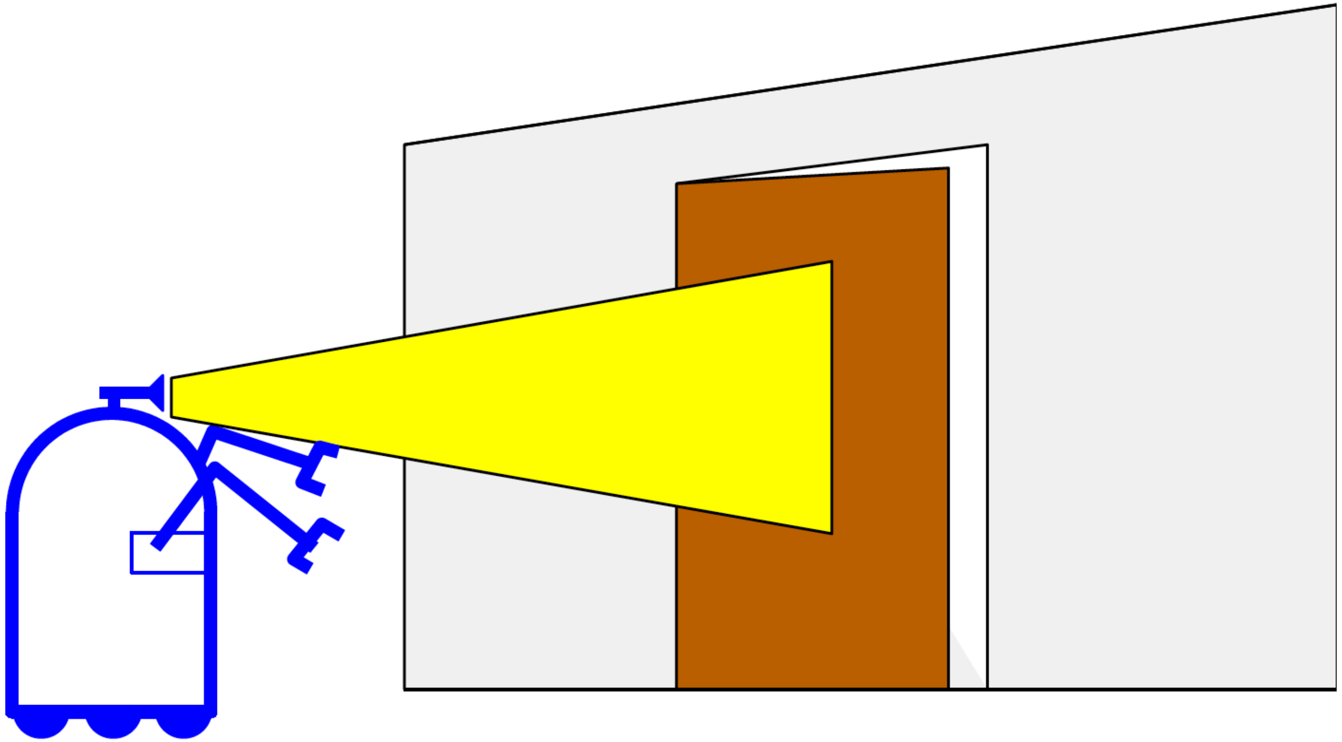
Some more properties

- The law of Total Variance:

$$Var(Y) = Var\left[E(Y \mid X)\right] + E\left[Var(Y \mid X)\right]$$

Simple Example of State Estimation

- Suppose a robot obtains measurement *z*
- What is *P(open|z)*?



Simple Example of State Estimation

- $P(z|open) = 0.6 \qquad P(z|\neg open) = 0.3$
- $P(open) = P(\neg open) = 0.5$

$$P(open \mid z) = \frac{P(z \mid open)P(open)}{P(z)}$$

$$P(open \mid z) = \frac{P(z \mid open)P(open)}{P(z \mid open)\,p(open) + P(z \mid \neg open)\,p(\neg open)}$$

$$P(open \mid z) = \frac{0.6 \cdot 0.5}{0.6 \cdot 0.5 + 0.3 \cdot 0.5} = \frac{2}{3} = 0.67$$

- z raises the probability that the door is open.

What is we have multiple measurements?

- Suppose our robot obtains another observation z_2.
- How can we integrate this new information?
- More generally, how can we estimate $P(x| z_1 \ldots z_n)$?

Recursive Bayesian Updating

$$P(x \mid z_1, \ldots, z_n) = \frac{P(z_n \mid x, z_1, \ldots, z_{n-1})\, P(x \mid z_1, \ldots, z_{n-1})}{P(z_n \mid z_1, \ldots, z_{n-1})}$$

Markov assumption: z_n is independent of $z_1, \ldots, z_{n-1}$ if we know x.

$$P(x \mid z_1, \ldots, z_n) = \frac{P(z_n \mid x)\, P(x \mid z_1, \ldots, z_{n-1})}{P(z_n \mid z_1, \ldots, z_{n-1})}$$

Example: Second Measurement

- $P(z_2|open) = 0.5$ $\qquad$ $P(z_2|\neg open) = 0.6$
- $P(open|z_1)=2/3$

$$P(open \mid z_2, z_1) = \frac{P(z_2 \mid open)\, P(open \mid z_1)}{P(z_2 \mid open)\, P(open \mid z_1) + P(z_2 \mid \neg open)\, P(\neg open \mid z_1)}$$

$$= \frac{\frac{1}{2} \cdot \frac{2}{3}}{\frac{1}{2} \cdot \frac{2}{3} + \frac{3}{5} \cdot \frac{1}{3}} = \frac{5}{8} = 0.625$$

- z_2 lowers the probability that the door is open.