Robotics

Estimation and Learning
with Dan Lee

Basic Intro to Probability

Penn Engineering

Why Learn About Probability?

- The real world has huge aspects of randomness and uncertainty.
- Still, we hope to make useful predictions and inferences.
- Randomness often follows reliable laws.
- The language of these laws is the language of probability (and statistics).

Contents

1. Definition of Probability
2. Independence
3. Conditional Probability
4. Bayes Rule
5. Random Variables
6. Density and Distribution Functions

1. Definition of Probability

- Consider an (potentially abstract) experiment
  - A **sample space** Ω is the set of *all possible* outcomes of that experiment
  - An **elementary event** ω is a single outcome of the set.
- Example (Rolling two dice):
  - Sample space $\Omega = \{(1,1), (1,2), (1,3), \ldots, (6,5), (6,6)\}$
  - Each member $\omega \in \Omega$ is an elementary event
  - Example of non-elementary event : Rolling doubles

    $B = \{(1,1), (2,2), (3,3), \ldots, (6,6)\}$

1. Definition of Probability

- Consider a finite* set Ω
- A **probability space** (Ω, P) is a sample space Ω, together with a function P, satisfying the following:
  i. $0 \leq P(\omega) \leq 1$ for all $\omega \in \Omega$
  ii. $\sum_{\omega \in \Omega} P(\omega) = 1$
  iii. For any event $A \subseteq \Omega$, $P(A) = \sum_{\omega \in \mathrm{A}} P(\omega)$
- The function P is called **probability measure**.

*To deal with countably infinite or uncountable space, we need third element called sigma algebra, but here we are simplifying.

1. Definition of Probability

- Basic Consequences

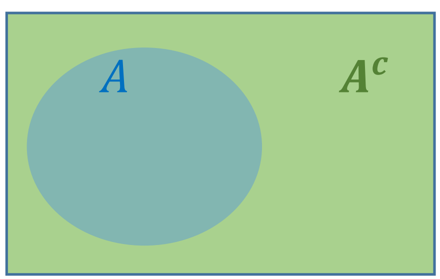
$$P(\emptyset) = 0$$

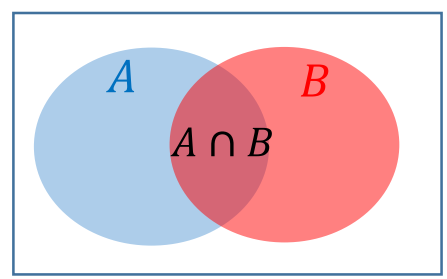
$$P(\Omega) = 1$$

$$P(A^C) = 1 - P(A)$$

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

$$P(A) = P(A \cap B) + P(A \cap B^c)$$





2. Independence

- Given a probability space, two events A and B are **independent** if and only if:

$$P(A \cap B) = P(A)P(B)$$

- Two events are dependent if they are not independent.

2. Independence

- Example (two coin flip): $\Omega = \{HH, HT, TH, TT\}$
  - Assume $\forall \omega \in \Omega, P(\omega) = \frac{1}{|\Omega|} = \frac{1}{4}$.
  - Define event A: "First flip is H." $A = \{HH, HT\}$
  - Define event B: "Second flip is H." $B = \{HH, TH\}$
  - Are A and B Independent?

    i) $P(A \cap B) = P(\{HH\}) = 1/4$

    ii) $P(A)P(B) = 1/2 * 1/2 = 1/4$ → Yes

2. Independence

- Example (two coin flip): $\Omega = \{HH, HT, TH, TT\}$
  - Assume $\forall \omega \in \Omega, P(\omega) = \frac{1}{|\Omega|} = \frac{1}{4}$.
  - Define event A: “First flip is H.” $A = \{HH, HT\}$
  - Define event B: “Contains a T.” $B = \{HT, TH, TT\}$
  - Are A and B Independent?

    i) $P(A \cap B) = P(\{HT\}) = 1/4$

    ii) $P(A)P(B) = 1/2 * 3/4 = 3/8$ → No

3. Conditional Probability

- Given some probability space (Ω, P), for any two events A and B, if $P(B) \neq 0$, then we define the **conditional probability** $P(A|B)$ that A occurs given that B occurs as,

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

- From this follows **Chain Rule**:
  $P(A \cap B) = P(A|B)\, P(B)$
  $$\begin{aligned} P(A \cap B \cap C) &= P(A|B \cap C)\, P(B \cap C) \\ &= P(A|B \cap C)\, P(B|C) P(C) \end{aligned}$$
  and so on..

3. Conditional Probability

- Example (two coin flip): What is the probability that both are head, GIVEN at least one is head?
- Probability Problem:

$\Omega = \{HH, HT, TH, TT\}$
$B = \{HH, HT, TH\}$ → "At least one is head."
$A = \{HH\}$ → "Both are heads."
$P(A \cap B) = P(\{HH\}) = 1/4$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/4}{3/4} = \frac{1}{3}$$

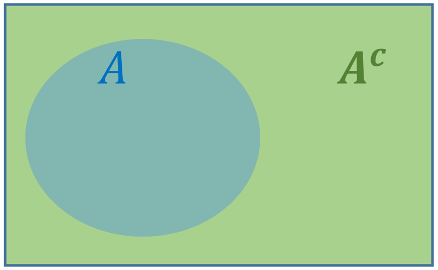
3. Conditional Probability

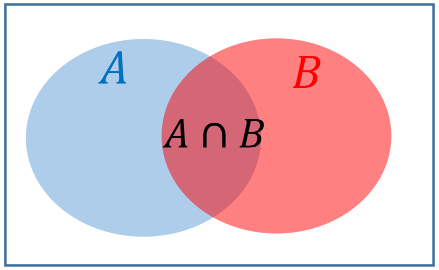
- Consequences

$$P(\emptyset|B) = 0$$

$$P(\mathrm{B}|\mathrm{B}) = 1$$

$$P(A|B) = 1 - P(A^C|B)$$





4. Bayes Rule

- From chain rule,
  $P(A \cap B) = P(A|B)\,P(B)$
  $P(B \cap A) = P(B|A)\,P(A)$
- Intersections are commutative, $A \cap B = B \cap A$.

$$\begin{array}{ccc} P(A \cap B) & = & P(A|B)\,P(B) \\ \| & & \| \\ P(B \cap A) & = & P(B|A)\,P(A) \end{array}$$

- As a result, we have **Bayes Rule**:

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

4. Bayes Rule

- Each term is often called:

$$\boxed{\overset{\text{Posterior}}{P(A|B)} = \frac{\overset{\text{Likelihood}}{P(B|A)}\,\overset{\text{Prior}}{P(A)}}{\underset{\text{Evidence}}{P(B)}}}$$

5. Random Variables

- Given some probability space (Ω, P), a **random variable** $X: \Omega \rightarrow R$ is a *function* that maps the sample space to the reals.

学了机器学习就懂了， 给事件标记叫做。
比如男、女是 1 和 0

- When we say $P(X = a)$, we actually mean the probability of the inverse image $X^{-1}(a)$. That is,

P(X=1)=P(男人)

$$P(X = a) = P\left(X^{-1}(a)\right) = P(\{\omega \in \Omega | X(\omega) = a\})$$

- Example (single coin flip): $\Omega = \{Head, Tail\}$

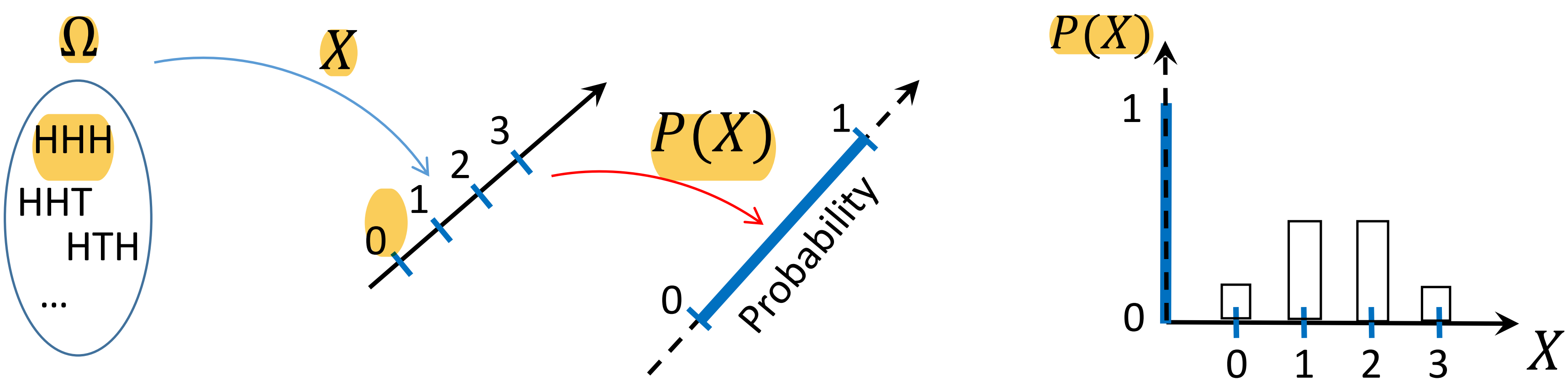
$$X(Head) = 1, X(Tail) = 0$$
$$P(X = 1) = P\left(X^{-1}(1)\right) = P(Head)$$

5. Random Variables

- Example: 3 coin flips
  - $\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$
  - Let us define $X(\omega)$ to be the number of Heads in a given flip. Then,

    $X(HHH) = 3, X(HHT) = X(HTH) = 2, \dots, X(TTT) = 0$

    $P(X = 3) = 1/8, P(X = 2) = P(X = 1) = 3/8, P(X = 0) = 1/8$

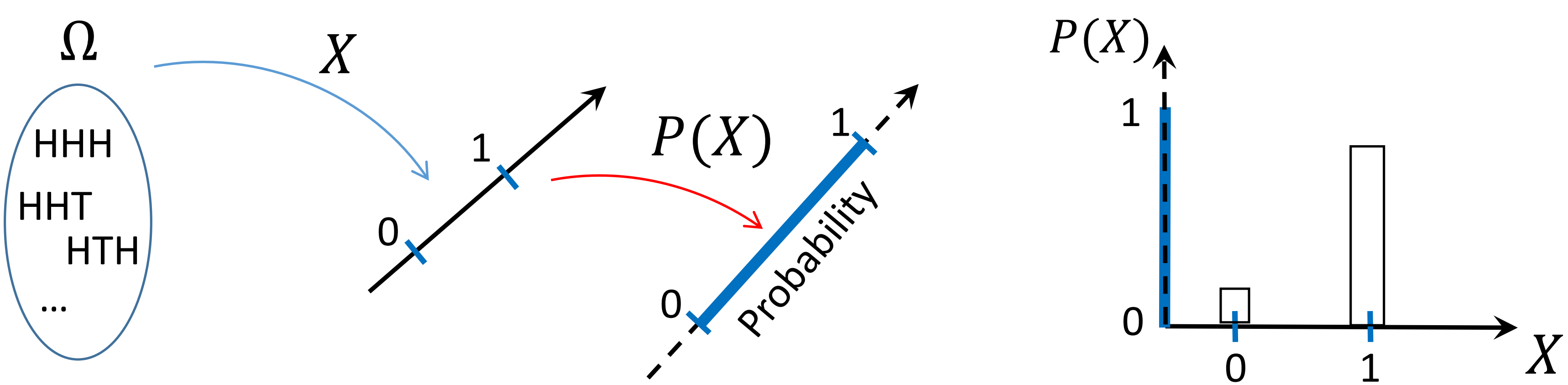


5. Random Variables

- Example: 3 coin flips
  - $\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$
  - This time, let us define $X(\omega)$ to be 1 if H appears in a given flip, otherwise is 0. Then,

    $X(HHH) = X(HHT) = \cdots = 1\,, X(TTT) = 0$

    $P(X = 1) = 7/8, P(X = 0) = 1/8$

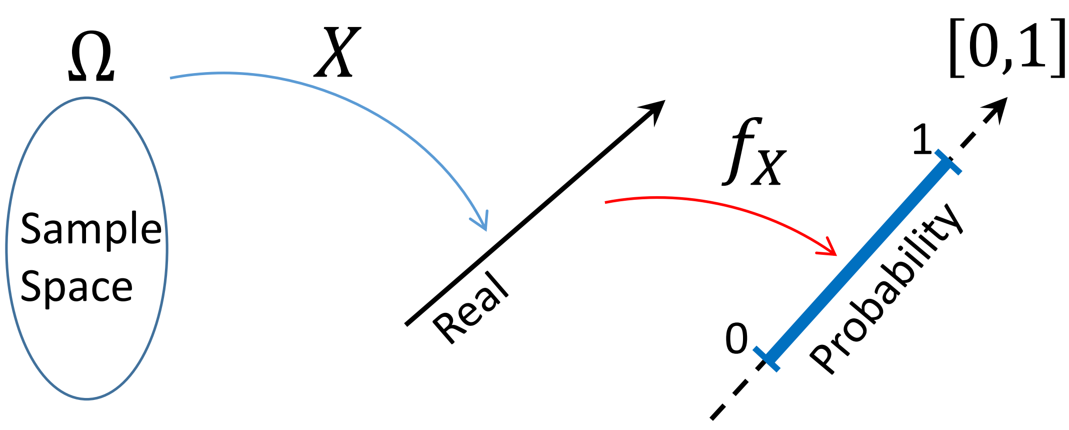


6. Density/Distribution Functions

- **Probability mass function** (pmf) of **discrete** RVs
- **Probability density function** (pdf) of **continuous** RVs

$f: R \to [0,1]$

$\forall a \in R, f_X(a) = P(X = a)$



6. Density/Distribution Functions

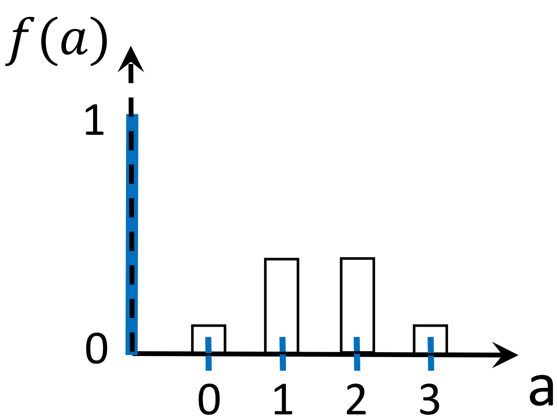
- **Cumulative distribution functions** (cdf)

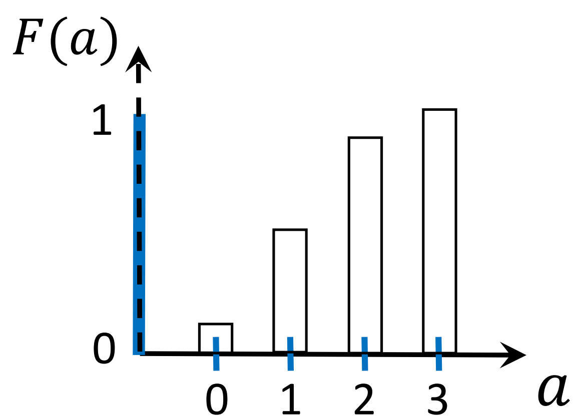
  $F: R \rightarrow [0,1]$

  $\forall a \in R, F_X(a) = P(X \leq a)$

- A cdf is a monotonic nondecreasing function, i.e.,

  $\forall x \leq y, F(x) \leq F(y)$





Acknowledgement

- Thanks to Daniel Moroz, Dan Lee’s master student at the University of Pennsylvania, for allowing us to take parts of contents from his slides.