Dynamic Programming Lecture #3

Outline:

- Probability Review
  - Probability space
  - Conditional probability
  - Total probability
  - Bayes rule
  - Independent events
  - Conditional independence
  - Mutual independence

Probability Space

- Sample Space: A set Ω
- Event: A subset of Ω
- To "any " subset of $A \subset \Omega$, we denote

$$\mathrm{P}\left[A\right] = \text{the probabilty of } A$$

- Axioms:
  1. $\mathrm{P}\left[A\right] \geq 0$
  2. $\mathrm{P}\left[\Omega\right] = 1$
  3. For disjoint sets A_1, A_2, A_3, $\ldots$

$$\mathrm{P}\left[A_1 \bigcup A_2 \bigcup A_3 \bigcup \ldots\right] = \mathrm{P}\left[A_1\right] + \mathrm{P}\left[A_2\right] + \mathrm{P}\left[A_3\right] + \ldots$$

- We will restrict our attention to "countable" probability spaces:

$$\Omega = \{\omega_1, \omega_2, \omega_3, ...\}$$

$$\mathrm{P}\left[\omega_i\right] = p_i$$

$$\mathrm{P}\left[A\right] = \sum_{\omega_i \in A} p_i$$

Example: Pair of Dice

$$\Omega = \{(i,j) : 1 \le i \le 6 \quad \& \quad 1 \le j \le 6\}$$

- Visualization:

$i \backslash j$	1	2	3	4	5	6
1	·	·	·	·	·	·
2	·	·	·	·	·	·
3	·	·	·	·	·	·
4	·	·	·	·	·	·
5	·	·	·	·	·	·
6	·	·	·	·	·	·

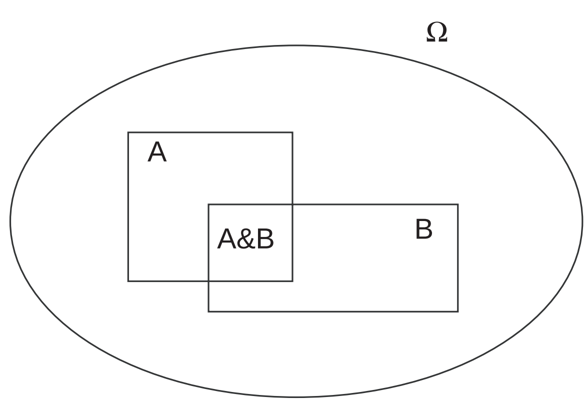
- Set $\mathrm{P}[(i,j)] = p_{ij}$
- For "fair" dice, $p_{ij} = 1/36$ for all "rolls", i.e., (i,j) pairs.
- To compute probabilities, must translate statements into events (i.e., subsets):
  - Doubles: $p_{11} + \ldots + p_{66}$
  - Larger die = 3: $p_{13} + p_{23} + p_{33} + p_{32} + p_{31}$
  - Sum of dice = 4: $p_{13} + p_{22} + p_{31}$
- Same events *regardless* of p_{ij} values—only resulting probabilities differ.

Conditional Probability

- Motivation: Compare probability of an event versus probability of the same event GIVEN additional information.
- Example:
  - (Probability sum of dice ≥ 7)
  - (Probability sum of dice ≥ 7) given (one of dice ≥ 5)
- Example: Probability car needs repair given engine light is on?
- CONDITIONAL PROBABILITY: Let A and B be events. Define "probability of A given B":

$$\mathrm{P}\left[A|B\right] = \frac{\mathrm{P}\left[A \cap B\right]}{\mathrm{P}\left[B\right]}$$

- Note that $A \cap B$ is another event, interpreted as BOTH A AND B.



- "given B" effectively redefines the sample space of events.
- Extreme examples: $A \subset B$? $B \subset A$? $A \cap B = \emptyset$?

Example: Dice

- What is probability (sum of dice ≤ 4) given (larger die = 3)?
- Translate to events:
  - A : (sum of dice ≤ 4) $= \{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)\}$

$i\backslash j$	1	2	3	4	5	6
1	X	X	X	·	·	·
2	X	X	·	·	·	·
3	X	·	·	·	·	·
4	·	·	·	·	·	·
5	·	·	·	·	·	·
6	·	·	·	·	·	·

  - B : (larger die $= 3$) $= \{(1,3),(2,3),(3,3),(3,2),(3,1)\}$

$i\backslash j$	1	2	3	4	5	6
1	·	·	X	·	·	·
2	·	·	X	·	·	·
3	X	X	X	·	·	·
4	·	·	·	·	·	·
5	·	·	·	·	·	·
6	·	·	·	·	·	·

  - $A \cap B = \{(1,3),(3,1)\}$
- Result:

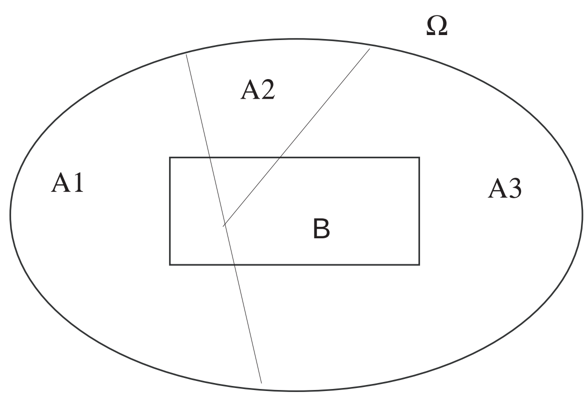
$$\mathrm{P}[A|B] = \frac{p_{13} + p_{31}}{p_{13} + p_{23} + p_{33} + p_{32} + p_{31}}$$

- Neither the definition nor computation relies on "natural" probabilities of dice.

Total Probability

- Suppose A_1 and A_2 satisfy:
  - $A_1 \cap A_2 = \emptyset$
  - $A_1 \cup A_2 = \Omega$

  i.e., A_1 and A_2 form a partition of Ω.
- For any event B: $B = (B \cap A_1) \cup (B \cap A_2)$
- From axioms:
$$\mathrm{P}[B] = \mathrm{P}\left[B \bigcap A_1\right] + \mathrm{P}\left[B \bigcap A_2\right]$$
- Using conditional probability:
$$\mathrm{P}[B] = \mathrm{P}[B|A_1]\,\mathrm{P}[A_1] + \mathrm{P}[B|A_2]\,\mathrm{P}[A_2]$$
- More generally, given a mutually exclusive $A_1, \ldots, A_N$ partition of Ω:
$$\mathrm{P}[B] = \mathrm{P}[B|A_1]\,\mathrm{P}[A_1] + \ldots + \mathrm{P}[B|A_N]\,\mathrm{P}[A_N]$$



Bayes' Rule

- Recall Total Probability: Given a mutually exclusive $A_1, \ldots, A_N$ partition of Ω:

$$\mathrm{P}[B] = \mathrm{P}[B|A_1]\,\mathrm{P}[A_1] + \ldots + \mathrm{P}[B|A_N]\,\mathrm{P}[A_N]$$

- For any A_i:

$$\mathrm{P}[A_i|B] = \frac{\mathrm{P}[A_i \ \&\ B]}{\mathrm{P}[B]}$$

$$\mathrm{P}[B|A_i] = \frac{\mathrm{P}[A_i \ \&\ B]}{\mathrm{P}[A_i]}$$

  same RHS numerator

$$\Rightarrow$$

$$\mathrm{P}[A_i|B]\,\mathrm{P}[B] = \mathrm{P}[B|A_i]\,\mathrm{P}[A_i]$$

- Now rewrite, using total probability:

$$\mathrm{P}[A_i|B] = \frac{\mathrm{P}[B|A_i]\,\mathrm{P}[A_i]}{\mathrm{P}[B|A_1]\,\mathrm{P}[A_1] + \ldots + \mathrm{P}[B|A_N]\,\mathrm{P}[A_N]}$$

- Known as Bayes' rule
- Utility: Hypothesis revision
  - A_i: hypotheses
  - B: new data
  - $\mathrm{P}[B|A_i]$: anticipated data given hypothesis
  - $\mathrm{P}[A_i]$: before-data (prior) "belief/confidence"
  - $\mathrm{P}[A_i|B]$: after-data (posterior) belief/confidence

Example: Cheater Detection

- Two coins:
  - Coin 1: $\mathrm{P}[H] = p_1$, $\mathrm{P}[T] = (1 - p_1)$
  - Coin 2: $\mathrm{P}[H] = p_2$, $\mathrm{P}[T] = (1 - p_2)$
- After seeing data, which coin is being used? Never know for sure, but can compute probabilities.
- Associate:
  - A_1: using coin 1
  - A_2: using coin 2
  - B: observed data
- Suppose we see 2 flips of coin: HH
  - $\mathrm{P}[HH|A_1] = p_1^2$
  - $\mathrm{P}[HH|A_2] = p_2^2$
- Similar with other flips, HT, TH, TT.
- Now apply Bayes' rule:

$$\mathrm{P}[A_i|B] = \frac{\mathrm{P}[B|A_i]\,\mathrm{P}[A_i]}{\mathrm{P}[B|A_1]\,\mathrm{P}[A_1] + \mathrm{P}[B|A_2]\,\mathrm{P}[A_2]}$$

- Numerical example: $\mathrm{P}[A_1] = 0.3$, $\mathrm{P}[A_2] = 0.7$, $p_1 = 0.9$, $p_2 = .5$, $B = HH$:

$$\mathrm{P}[A_1|HH] = \frac{(0.9)^2(0.3)}{(0.9^2)(0.3) + (0.5)^2(0.7)} = 0.58$$

$$\mathrm{P}[A_2|HH] = \frac{(0.5)^2(0.7)}{(0.9^2)(0.3) + (0.5)^2(0.7)} = 0.42$$

  i.e., increased suspicion of cheating!

Independent Events

- DEFINE: Events A and B are INDEPENDENT if

$$\mathrm{P}\left[A \bigcap B\right] = \mathrm{P}\left[A\right]\mathrm{P}\left[B\right]$$

- In terms of conditional probability:

$$\mathrm{P}\left[A|B\right] = \frac{\mathrm{P}\left[A \cap B\right]}{\mathrm{P}\left[B\right]} = \frac{\mathrm{P}\left[A\right]\mathrm{P}\left[B\right]}{\mathrm{P}\left[B\right]} = \mathrm{P}\left[A\right] \quad \text{(if independent!)}$$

- Information of B does not affect probability of A.
- Example: Fair dice
  - $A =$ doubles
  - $B =$ sum of roll ≤ 3
  - $\mathrm{P}\left[A \cap B\right] = 1/36 = \mathrm{P}\left[A\right]\mathrm{P}\left[B\right] = (1/6)(3/36)$? Not independent.
  - Changing B to event (first die $= 3$) results in independence.
- Independence depends on events A & B AND underlying probabilities.

Conditional Independence

- Conditional independence: Given an event C, the events A and B are conditionally independent given C if
$$\mathrm{P}\left[A\bigcap B|C\right] = \mathrm{P}\left[A|C\right]\mathrm{P}\left[B|C\right]$$
- Rewrite:
$$\mathrm{P}\left[A\bigcap B|C\right] = \frac{\mathrm{P}\left[A\cap B\cap C\right]}{\mathrm{P}\left[C\right]} = \frac{\mathrm{P}\left[A|(B\cap C)\right]\mathrm{P}\left[B\cap C\right]}{\mathrm{P}\left[C\right]} = \mathrm{P}\left[A|(B\bigcap C)\right]\mathrm{P}\left[B|C\right]$$
$$\Rightarrow \text{(under conditional independence)}$$
$$\mathrm{P}\left[A|C\right] = \mathrm{P}\left[A|(B\bigcap C)\right]$$
i.e., additional knowledge of B does not affect probability of A
- Conditioning may change independence.
- Example: Two coins. Probability of heads for each coin p_A and p_B
  - Randomly choose coin (A, B) with probability $(q, 1-q)$
  - Toss selected coin 2 times
  - Events:
$$E_1 = \text{heads on first toss}$$
$$E_2 = \text{heads on second toss}$$
$$E_0 = \text{choose coin } A$$
  - Question: Are E_1 and E_2 independent? No.
$$\mathrm{P}\left[E_1 \& E_2\right] = \mathrm{P}\left[E_1\right]\mathrm{P}\left[E_2\right]?$$
$$\begin{aligned} & \mathrm{P}\left[E_1 \& E_2|E_0\right]\mathrm{P}\left[E_0\right] + \mathrm{P}\left[E_1 \& E_2|E_0^c\right]\mathrm{P}\left[E_0^c\right] \\ = \;& p_A^2 q + p_B^2(1-q) \\ = \;& (\,p_A q + p_B(1-q)\,)\,(\,p_A q + p_B(1-q)\,)? \end{aligned}$$
  - Are E_1 and E_2 conditionally independent on E_0? Yes.
$$\mathrm{P}\left[E_1 \& E_2|E_0\right] = p_A^2 = \mathrm{P}\left[E_1|E_0\right]\mathrm{P}\left[E_2|E_0\right]$$

Mutual Independence

- Mutual independence: A set of events $A_1, \ldots, A_N$ are mutually independent if for *any* sub-collection, S

$$\mathrm{P}\left[\bigcap_{i \in S} A_i\right] = \Pi_{i \in S} \mathrm{P}\left[A_i\right]$$

- Pairwise independence does not imply mutual independence.
- Example: 2 fair dice
  - A : doubles; B : first die = 6; C : second die = 1.

$i \backslash j$	1	2	3	4	5	6
1	(A, C)	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
2	C	A	$\cdot$	$\cdot$	$\cdot$	$\cdot$
3	C	$\cdot$	A	$\cdot$	$\cdot$	$\cdot$
4	C	$\cdot$	$\cdot$	A	$\cdot$	$\cdot$
5	C	$\cdot$	$\cdot$	$\cdot$	A	$\cdot$
6	(B, C)	B	B	B	B	(A, B)

  - $A \& B$ are independent, $A \& C$ are independent, $B \& C$ are independent.
  - $\mathrm{P}[A \ \& \ B \ \& \ C] = 0 \neq \mathrm{P}[A] \, \mathrm{P}[B] \, \mathrm{P}[C]$