Lecture 34: Random Variables

- We'll start with law of total prob., a useful technical tool.
- Outcomes tell us "what happened", RVs capture what we do with those observations - defns, basic properties & notation

Applications
- RVs are central to probabilistic algorithms. Let us define things like "the prob. that I get close to the right answer"
Claim: Law of total prob: if Bi partition S, then

\[ \Pr(A) = \sum_i \Pr(A|B_i) \Pr(B_i) \]

PP: \( A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_n) \) \hspace{1cm} \text{since Bi partitions } S \hspace{1cm} \text{since Bi partition } S.

\[ \text{So (by axiom 3):} \]

\[ \Pr(A) = \sum_i \Pr(A \cap B_i) \]

\[ = \sum_i \Pr(A|B_i) \Pr(B_i) \]

\[ = \sum_i \Pr(A|B_i) \frac{\Pr(A \cap B_i)}{\Pr(B_i)} \]

because \( \Pr(A|B_i) = \frac{\Pr(A \cap B_i)}{\Pr(B_i)} \)
Example:

A disease occurs in $\frac{1}{10,000}$ people.

Let $D$ be the event "I have disease".

$$\Pr(D) = \frac{1}{10,000}$$

Let $H$ be "I'm healthy".

Let $P$ be the event "the test is positive".

The test is right $99\%$ of the time:

$$\Pr(P|H) = \frac{99}{100} \quad \Pr(N|D) = \frac{1}{10,000}$$

$$\Pr(P|D) = \frac{99}{100}$$

I take the test, get positive, what's the probability that I have disease?

What is $\Pr(D|P)$?

$$\Pr(D|P) = \frac{\Pr(P|D)\Pr(D)}{\Pr(P)}$$

$$\Pr(D) = \Pr(P|H)\Pr(H) + \Pr(P|D)\Pr(D)$$

Law of total probability.

$$\Pr(D) = \frac{99}{100} \times \frac{1}{10,000} + \frac{99}{100} \times \frac{1}{10,000} \leq \frac{99}{10,000} + \frac{99}{10,000} \approx \frac{100}{10,000} = \frac{1}{100}$$

$$\Pr(D|P, \neg P_2) = ?$$
Random variables

**Definition:** A $T$-valued random variable $X$ is a function $X : S \rightarrow T$

- $IR$-valued RV: $X : S \rightarrow IR$

**Example:** Height: People $\rightarrow$ IR

$$
\begin{array}{ccc}
S & \rightarrow & IR \\
& & \\
\downarrow & & \downarrow \\
3 & \rightarrow & 3 \\
& & \\
\downarrow & & \downarrow \\
Ht : & a, d & \in \{a, d\} \\
\end{array}
$$

If $X$ is a RV, $x \in IR$

- $(X = x)$ is an event, the set of outcomes $S$ where $X(s) = x$
  
  $= \{ s \mid X(s) = x \}$

- $(X > x) = \{ s \mid X(s) > x \}$

- $(X = Y) = \{ s \mid X(s) = Y(s) \}$
Defn: If \( X \) and \( Y \) are RVs then \( X+Y \) is a RV given by

\[
(X+Y)(s) := X(s) + Y(s)
\]

\[
(XY)(s) := X(s) \cdot Y(s)
\]

\[
(3X)(s) := 3 \cdot X(s)
\]

Ex: Roll two dice, add results. \( N=\{1, \ldots, 6\} \)

\( S = \{(1,1), (1,2), \ldots, (1,6), (2,6), \ldots, (6,6)\} = N \times N \)

\[ X_1 : N \times N \to \mathbb{R} \quad X_1(1,6) = 1 \quad X_1(3,5) = 3 \]

\[ X_2 : N \times N \to \mathbb{R} \quad X_2(1,6) = 6 \quad X_2(3,5) = 5 \]

\[ X := X_1 + X_2 \quad X(1,6) = X_1(1,6) + X_2(1,6) = 1 + 6 = 7 \]
Probability mass function (PMF) of $X$

$\text{PMF}_X : T \rightarrow \mathbb{R}$

$\text{PMF}_X(x) := \Pr(X = x)$

$\text{PMF}_X(3) := \frac{1}{2}$

$\text{PMF}_X(2) := \frac{1}{4}$.