Basics on Probability

Coin Flips

- You flip a coin
  - Head with probability 0.5

- You flip 100 coins
  - How many heads would you expect

Coin Flips cont.

- You flip a coin
  - Head with probability *p*
  - Binary random variable
  - Bernoulli trial with success probability *p*
- You flip *k* coins
  - How many heads would you expect
  - Number of heads X: discrete random variable
  - Binomial distribution with parameters *k* and *p*

Discrete Random Variables

- Random variables (RVs) which may take on only a **countable** number of **distinct** values
  - E.g., the total number of heads X you get if you flip 100 coins

- X is a RV with arity *k* if it can take on exactly one value out of $\{x_1, \ldots, x_k\}$
  - E.g., the possible values that X can take are 0, 1, 2, …, 100

Probability of Discrete RV

- Probability mass function (pmf): $\mathrm{P}\left(\mathrm{X}=x_i\right)$
- Easy facts about pmf
  - $\sum_i \mathrm{P}\left(\mathrm{X}=x_i\right)=1$
  - $\mathrm{P}\left(\mathrm{X}=x_i \cap \mathrm{X}=x_j\right)=0$ if $i \neq j$
  - $\mathrm{P}\left(\mathrm{X}=x_i \cup \mathrm{X}=x_j\right)=\mathrm{P}\left(\mathrm{X}=x_i\right)+\mathrm{P}\left(\mathrm{X}=x_j\right)$ if $i \neq j$
  - $\mathrm{P}\left(\mathrm{X}=x_1 \cup \mathrm{X}=x_2 \cup \ldots \cup \mathrm{X}=x_k\right)=1$

Common Distributions

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

Coin Flips of Two Persons

- Your friend and you both flip coins
  - Head with probability 0.5
  - You flip 50 times; your friend flip 100 times
  - How many heads will both of you get

Joint Distribution

- Given two discrete RVs X and Y, their **joint distribution** is the distribution of X and Y together
  - E.g., P(You get 21 heads AND you friend get 70 heads)
- $$\sum_x \sum_y \mathrm{P}\left(\mathrm{X} = x \cap \mathrm{Y} = y\right) = 1$$
  - E.g.,

$$\sum_{i=0}^{50} \sum_{j=0}^{100} \mathrm{P}\left(\text{You get } i \text{ heads AND your friend get } j \text{ heads}\right) = 1$$

Conditional Probability

- $\mathrm{P}\left(\mathrm{X}=x \middle| \mathrm{Y}=y\right)$ is the probability of $\mathrm{X}=x$, given the occurrence of $\mathrm{Y}=y$
  - E.g., you get 0 heads, given that your friend gets 61 heads
- $\mathrm{P}\left(\mathrm{X}=x \middle| \mathrm{Y}=y\right)=\dfrac{\mathrm{P}\left(\mathrm{X}=x \cap \mathrm{Y}=y\right)}{\mathrm{P}\left(\mathrm{Y}=y\right)}$

Law of Total Probability

- Given two discrete RVs X and Y, which take values in $\{x_1, \ldots, x_m\}$ and $\{y_1, \ldots, y_n\}$, we have

$$
\begin{aligned}
\mathrm{P}(\mathrm{X} = x_i) &= \sum_j \mathrm{P}(\mathrm{X} = x_i \cap \mathrm{Y} = y_j) \\
&= \sum_j \mathrm{P}(\mathrm{X} = x_i \mid \mathrm{Y} = y_j)\, \mathrm{P}(\mathrm{Y} = y_j)
\end{aligned}
$$

Marginalization

Marginal Probability

Joint Probability

$$P\left(X = x_i\right) = \sum_j P\left(X = x_i \cap Y = y_j\right)$$

$$= \sum_j P\left(X = x_i \middle| Y = y_j\right) P\left(Y = y_j\right)$$

Conditional Probability

Marginal Probability

Bayes Rule

- X and Y are discrete RVs…

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

$$\mathrm{P}\left(\mathrm{X}=x_i \middle| \mathrm{Y}=y_j\right)=\frac{\mathrm{P}\left(\mathrm{Y}=y_j \middle| \mathrm{X}=x_i\right)\mathrm{P}\left(\mathrm{X}=x_i\right)}{\sum_k \mathrm{P}\left(\mathrm{Y}=y_j \middle| \mathrm{X}=x_k\right)\mathrm{P}\left(\mathrm{X}=x_k\right)}$$

Independent RVs

- Intuition: X and Y are independent means that $\mathrm{X} = x$ **neither** makes it **more or less** probable that $\mathrm{Y} = y$
- Definition: X and Y are independent iff

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

More on Independence

- $$P\left(X = x \cap Y = y\right) = P\left(X = x\right)P\left(Y = y\right)$$

$$P\left(X = x \middle| Y = y\right) = P\left(X = x\right) \quad P\left(Y = y \middle| X = x\right) = P\left(Y = y\right)$$

- E.g., no matter how many heads you get, your friend will not be affected, and vice versa

Conditionally Independent RVs

- Intuition: X and Y are conditionally independent given Z means that once Z is **known**, the value of X does not add any **additional** information about Y
- Definition: X and Y are conditionally independent given Z iff

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

More on Conditional Independence

$$P\left(X = x \cap Y = y \middle| Z = z\right) = P\left(X = x \middle| Z = z\right) P\left(Y = y \middle| Z = z\right)$$

$$P\left(X = x \middle| Y = y, Z = z\right) = P\left(X = x \middle| Z = z\right)$$

$$P\left(Y = y \middle| X = x, Z = z\right) = P\left(Y = y \middle| Z = z\right)$$

Probability of Events

- X denotes an event that could possibly happen
  - E.g., X="you will fail in this course"
- P(X) denotes the **likelihood** that X happens, or X=true
  - E.g., what's the probability that you will fail in this course?
- Ω denotes the entire event set
  - $\Omega = \left\{X, \bar{X}\right\}$

The Axioms of Probabilities

- 0 <= P(X) <= 1
- $\mathrm{P}(\Omega) = 1$
- $\mathrm{P}(\mathrm{X}_1 \cup \mathrm{X}_2 \cup \ldots) = \sum_i \mathrm{P}(\mathrm{X}_i)$, where X_i are disjoint events
- Useful rules
  - $\mathrm{P}(\mathrm{X}_1 \cup \mathrm{X}_2) = \mathrm{P}(\mathrm{X}_1) + \mathrm{P}(\mathrm{X}_2) - \mathrm{P}(\mathrm{X}_1 \cap \mathrm{X}_2)$
  - $\mathrm{P}(\overline{\mathrm{X}}) = 1 - \mathrm{P}(\mathrm{X})$

Interpreting the Axioms