CS145: Lecture 6 Outline

- Collections of discrete random variables
- Independent random variables
- Expectations of multiple discrete variables
Discrete Probability Distributions

The **probability mass function** or **probability distribution** of random variable:

\[ p_X(x) = P(X = x) = P(\{\omega \in \Omega \mid X(\omega) = x\}) \]

\[ p_X(x) \geq 0, \quad \sum_{x \in \mathcal{X}} p_X(x) = 1. \]

The **range** of a random variable is the set of values with positive probability

\[ \mathcal{X} = \{x \in \mathbb{R} \mid X(\omega) = x \text{ for some } \omega \in \Omega, P(\omega) > 0\} \]

For a **discrete random variable**, the range is finite or countably infinite.

Suppose the range of \( X \) is of size \( N \), the PMF is a vector of length \( N \).
Joint Probability Distributions

Consider two random variables $X, Y$. Suppose range of $X$ is size $N$, range of $Y$ is size $M$.

The joint probability mass function or joint distribution of two variables:

$$p_{XY}(x, y) = P(X = x \text{ and } Y = y)$$

$$p_{XY}(x, y) \geq 0, \quad \sum_x \sum_y p_{XY}(x, y) = 1.$$ 

The joint distribution is uniquely specified by $NM-1$ numbers.

In this example, $N=2$ and $M=8$, and the joint PMF is a $2x8$ matrix.
Reminder: Total & Conditional Probability

- Partition of sample space into $A_1, A_2, A_3$
- Have $P(B \mid A_i)$, for every $i$

\[
\text{Shaded region is event } B
\]

- One way of computing $P(B)$:

\[
P(B) = P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + P(A_3)P(B \mid A_3)
\]

- “Prior” probabilities $P(A_i)$
  - initial “beliefs”
- Wish to compute $P(A_i \mid B)$
  - revise “beliefs”, given that $B$ occurred

\[
P(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)}
\]

\[
= \frac{P(A_i)P(B \mid A_i)}{P(B)}
\]

\[
= \frac{P(A_i)P(B \mid A_i)}{\sum_j P(A_j)P(B \mid A_j)}
\]
The joint probability mass function or joint distribution of two variables:

\[ p_{XY}(x, y) = P(X = x \text{ and } Y = y) \]

The range of each variable defines a partition of the sample space, so the marginal distributions can be computed from the joint distribution:

\[ p_X(x) = P(X = x) = \sum_y p_{XY}(x, y) \]
\[ p_Y(y) = P(Y = y) = \sum_x p_{XY}(x, y) \]
Marginal Probability Distributions

Figure 2.11: Illustration of the tabular method for calculating marginal PMFs from joint PMFs. The joint PMF is represented by a table, where the number in each square \((x, y)\) gives the value of \(p_{X,Y}(x, y)\). To calculate the marginal PMF \(p_X(x)\) for a given value of \(x\), we add the numbers in the column corresponding to \(x\). For example \(p_X(2) = \frac{8}{20}\). Similarly, to calculate the marginal PMF \(p_Y(y)\) for a given value of \(y\), we add the numbers in the row corresponding to \(y\). For example \(p_Y(2) = \frac{5}{20}\).

More than Two Random Variables

The joint PMF of three random variables \(X\), \(Y\), and \(Z\) is defined in analogy with the above as \(p_{X,Y,Z}(x, y, z) = P(X = x, Y = y, Z = z)\), for all possible triplets of numerical values \((x, y, z)\). Corresponding marginal PMFs are analogously obtained by equations such as

\[
    p_{X,Y}(x, y) = \sum_z p_{X,Y,Z}(x, y, z),
\]

and

\[
    p_X(x) = \sum_y \sum_z p_{X,Y,Z}(x, y, z).
\]

The expected value rule for functions takes the form

\[
    E[g(X, Y, Z)] = \sum_{x,y,z} g(x, y, z) p_{X,Y,Z}(x, y, z),
\]

and if \(g\) is linear and of the form \(aX + bY + cZ + d\), then

\[
\]
By the definition of conditional probability:

\[
P(X = x \mid Y = y) = \frac{P(X = x \text{ and } Y = y)}{P(Y = y)}
\]

The conditional probability mass function is then:

\[
p_{X \mid Y}(x \mid y) = P(X = x \mid Y = y) = \frac{p_{XY}(x, y)}{p_Y(y)} = \frac{p_{XY}(x, y)}{\sum_{x'} p_{XY}(x', y)}
\]
Conditional Probability Distributions

Conditioning one Random Variable on Another

Let $X$ and $Y$ be two random variables associated with the same experiment. If we know that the experimental value of $Y$ is some particular $y$ (with $p_Y(y) > 0$), this provides partial knowledge about the value of $X$. This knowledge is captured by the conditional PMF $p_{X|Y}$ of $X$ given $Y$, which is defined by specializing the definition of $p_{X|A}$ to events $A$ of the form $\{Y = y\}$:

$$p_{X|Y}(x|y) = \frac{P(X = x, Y = y)}{P(Y = y)}.$$

Using the definition of conditional probabilities, we have

$$p_{X|Y}(x|y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p_{X,Y}(x,y)}{p_Y(y)}.$$

Let us fix some $y$, with $p_Y(y) > 0$ and consider $p_{X|Y}(x|y)$ as a function of $x$. This function is a valid PMF for $X$: it assigns nonnegative values to each possible $x$, and these values add to 1. Furthermore, this function of $x$, has the same shape as $p_{X,Y}(x,y)$ except that it is normalized by dividing with $p_Y(y)$, which enforces the normalization property $\sum_x p_{X|Y}(x|y) = 1$.

Figure 2.13 provides a visualization of the conditional PMF.
Example: The Absent-Minded Prof

- At office hours, a Professor gets 0, 1, or 2 questions with equal probability.
- Each question is answered correctly with probability $\frac{3}{4}$ (independently).

\[
\begin{align*}
\text{Joint PMF } P_{X,Y}(x,y) \\
\text{in tabular form}
\end{align*}
\]

\[
\begin{array}{ccc}
0 & 1 & 2 \\
0 & 16/48 & 12/48 & 9/48 \\
1 & 4/48 & 6/48 \\
2 & 0 & 1/48 \\
\end{array}
\]

\[
P(\text{at least one wrong answer}) = p_{X,Y}(1,1) + p_{X,Y}(2,1) + p_{X,Y}(2,2)
\]
\[
= \frac{4}{48} + \frac{6}{48} + \frac{1}{48}.
\]
Several Random Variables

\[ p_{XYZ}(x, y, z) = P(X = x \text{ and } Y = y \text{ and } Z = z) \]

May compute marginal of any subset of variables, possibly conditioned on values of any other variables.
CS145: Lecture 6 Outline

- Collections of discrete random variables
- Independent random variables
- Expectations of multiple discrete variables
Reminder: Independence of Events

Independence of Two Events: \[ P(A \cap B) = P(A)P(B) \]

This implies that \[ P(A \mid B) = P(A), \ P(B \mid A) = P(B). \]

- Observing \( B \) provides no information about whether \( A \) occurred
- Observing \( A \) provides no information about whether \( B \) occurred

Definition of Conditional Probabilities:

- **Definition:** Assuming \( P(B) \neq 0 \),
  \[ P(A \mid B) = \frac{P(A \cap B)}{P(B)} \]
  \( P(A \mid B) \) undefined if \( P(B) = 0 \)
Independent Random Variables

\[ X \perp Y \]

\[ p_{XY}(x, y) = p_X(x)p_Y(y) \]

for all \( x \in \mathcal{X}, y \in \mathcal{Y} \)

- Equivalent conditions on conditional probabilities:
  \[ p_{X|Y}(x \mid y) = p_X(x) \text{ for all } p_Y(y) > 0 \]
  \[ p_{Y|X}(y \mid x) = p_Y(y) \text{ for all } p_X(x) > 0 \]
## Independent Random Variables

For a given set of marginal distributions, there exists a **unique** joint distribution under which those variables are independent.

Three random variables are independent if and only if

\[
  p_{XY}(x, y) = p_X(x)p_Y(y)
\]

for all \( x \in X, y \in Y \)

\[
  p_{XYZ}(x, y, z) = p_X(x)p_Y(y)p_Z(z)
\]
Example: Independence

\[ p_{XY}(x, y) = P(X = x \text{ and } Y = y) \]

Verify that \( X \) and \( Y \) are not independent:

\[ p_X(x) = \]
\[ p_Y(y) = \]
Conditional Independence

Apply the same definition of independence for $X$ and $Y$, but condition all probability distributions on some other variable $Z$.

Independence does not always imply conditional independence, and conditional independence does not always imply independence.

$$X \perp Y \mid Z = z$$

$$p_{XY|Z}(x, y \mid z) = p_{X|Z}(x \mid z)p_{Y}(y \mid z)$$

for all $x \in \mathcal{X}, y \in \mathcal{Y}$
Example: (Conditional) Independence

\[ p_{XY}(x, y) = P(X = x \text{ and } Y = y) \]

Verify that \( X \) and \( Y \) are not independent:

\[ p_X(x) = \]

\[ p_Y(y) = \]

But \( X \) and \( Y \) are conditionally independent given

\[ Z = 1_{\{X \leq 2, Y \geq 3\}} \]

\[ p_{X|Z}(x \mid 1) = \]

\[ p_{Y|Z}(y \mid 1) = \]
CS145: Lecture 6 Outline

- Collections of discrete random variables
- Independent random variables
- Expectations of multiple discrete variables
The **expectation** or **expected value** of a discrete random variable is:

$$E[X] = \sum_{x \in X} xp_X(x)$$

The expectation is a single number, not a random variable. It encodes the “center of mass” of the probability distribution:

$$x_{\text{min}} \leq E[x] \leq x_{\text{max}} \quad\quad x_{\text{min}} = \min\{x \mid x \in X\} \quad\quad x_{\text{max}} = \max\{x \mid x \in X\}$$

The expectation is an average or interpolation. It is possible that

$$p_X(E[x]) = 0$$

for some random variables $X$. 
Expectation of Multiple Variables

- The expectation or expected value of a function of two discrete variables:

\[ E[g(X, Y)] = \sum_{x} \sum_{y} g(x, y)p_{XY}(x, y) \]

- A similar formula applies to functions of 3 or more variables

- Expectations of sums of functions are sums of expectations:

\[ E[g(X) + h(Y)] = E[g(X)] + E[h(Y)] = \left[ \sum_{x} g(x)p_X(x) \right] + \left[ \sum_{y} h(y)p_Y(y) \right] \]

- This is always true, whether or not \( X \) and \( Y \) are independent

- Specializing to linear functions, this implies that:

\[ E[aX + bY + c] = aE[X] + bE[Y] + c \]
Mean of Binomial Probability Distribution

- Suppose you flip $n$ coins with bias $p$, count number of heads
- A binomial random variable $X$ has parameters $n$, $p$:
  
  $$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

- For binomial, expected values are expected counts of events:
  
  $$E[X] = pn$$

- Simple proof uses indicator variables $X_i$ for whether each of $n$ tosses is heads:
  
  $$E[X_i] = p \cdot 1 + (1-p) \cdot 0 = p = \Pr(X_i = 1)$$.

  $$E[X] = E\left[ \sum_{i=1}^{n} X_i \right] = \sum_{i=1}^{n} E[X_i] = np.$$
Binomial Mean: The Hard Way

\[ \begin{align*}
E[X] &= \sum_{j=0}^{n} j \binom{n}{j} p^j (1-p)^{n-j} \\
     &= \sum_{j=0}^{n} \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j} \\
     &= \sum_{j=1}^{n} \frac{n!}{(j-1)!(n-j)!} p^j (1-p)^{n-j} \\
     &= np \sum_{j=1}^{n} \frac{(n-1)!}{(j-1)!((n-1)-(j-1))!} p^{j-1} (1-p)^{(n-1)-(j-1)} \\
     &= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!((n-1)-k)!} p^k (1-p)^{(n-1)-k} \\
     &= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k} = np.
\end{align*} \]