CS145: Probability & Computing
Lecture 6: Multiple Discrete Variables, Joint & Conditional Distributions, Independence

Instructor: Cyrus Cousins
Brown University Computer Science

Figure credits:
Bertsekas & Tsitsiklis, Introduction to Probability, 2008
Pitman, Probability, 1999
Collections of discrete random variables
Independent random variables
Expectations of multiple discrete variables
The range of a random variable is the set of values with positive probability. For a discrete random variable, the range is finite or countably infinite.

The probability mass function or probability distribution of random variable:

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

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

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

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

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

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

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

The joint distribution is uniquely specified by $NM-1$ numbers.
Reminder: Total & Conditional Probability

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

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

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

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

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

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

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

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

The marginal distributions are defined by \((N-1) + (M-1)\) numbers. Many joint distributions may have the same marginals.
## Marginal Probability Distributions

The joint probability mass function (PMF) $P_{X,Y}(x,y)$ in tabular form is shown below:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1/20</th>
<th>1/20</th>
<th>1/20</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3/20</td>
<td>7/20</td>
<td>7/20</td>
<td>3/20</td>
</tr>
<tr>
<td>3</td>
<td>1/20</td>
<td>2/20</td>
<td>3/20</td>
<td>1/20</td>
</tr>
<tr>
<td>2</td>
<td>1/20</td>
<td>2/20</td>
<td>3/20</td>
<td>1/20</td>
</tr>
<tr>
<td>1</td>
<td>1/20</td>
<td>1/20</td>
<td>1/20</td>
<td>0</td>
</tr>
</tbody>
</table>

**Row Sums:**
- Marginal PMF $P_Y(y)$

**Column Sums:**
- Marginal PMF $P_X(x)$
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

"SLICE VIEW" of Conditional PMF $P_{X|Y}(x|y)$

PMF $P_{X,Y}(x,y)$

Conditional PMF $P_{X|Y}(x|3)$

Conditional PMF $P_{X|Y}(x|2)$

Conditional PMF $P_{X|Y}(x|1)$
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).
Several Random Variables

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

\[ p_{XY}(x, y) = \sum_{z \in Z} p_{XYZ}(x, y, z) \quad p_X(x) = \sum_{y \in Y} p_{XY}(x, y) \quad p_{XY|Z}(x, y | z) = \frac{p_{XYZ}(x, y, z)}{p_Z(z)} \]

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

Equivalent conditions on conditional probabilities:

\[ p_{X|Y}(x \mid y) = p_X(x) \quad \text{for all } p_Y(y) > 0 \]

\[ p_{Y|X}(y \mid x) = p_Y(y) \quad \text{for all } p_X(x) > 0 \]
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

\[ P(x, y, z) = P_X(x)P_Y(y)P_Z(z) \]

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

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

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

\[
p_X(x) = \]

\[
p_Y(y) = \]
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.
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

\[ Z = 1_{\{X \leq 2, Y \geq 3\}} \]
\[ p_{X|Z}(x \mid 1) = \]
\[ p_{Y|Z}(y \mid 1) = \]
Collections of discrete random variables
Independent random variables
Expectations of multiple discrete variables
The expectation or expected value of a discrete random variable is:

\[ E[X] = \sum_{x \in \mathcal{X}} x p_X(x) \]

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

\[ x_{\min} \leq E[x] \leq x_{\max} \]

\[ x_{\min} = \min\{x \mid x \in \mathcal{X}\} \]
\[ x_{\max} = \max\{x \mid x \in \mathcal{X}\} \]

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

\[ p_X(E[x]) = 0 \text{ for some random variables } X. \]
The expectation or expected value of a function of two discrete variables:

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

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

Expectations of sums of functions are sums of expectations:

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

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

Specializing to linear functions, this implies that:

\[ E[aX + bY + c] = aE[X] + bE[Y] + c \]
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:

\[
E[X] = np
\]

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**

\[
E[X] = \sum_{j=0}^{n} j \binom{n}{j} p^j (1 - p)^{n-j}
\]

\[
= \sum_{j=0}^{n} j \frac{n!}{j!(n-j)!} p^j (1 - p)^{n-j}
\]

\[
= \sum_{j=1}^{n} \frac{n!}{(j-1)!(n-j)!} p^j (1 - p)^{n-j}
\]

\[
= np \sum_{j=1}^{n} \frac{(n-1)!}{(j-1)!(n-j-(j-1))!} p^{j-1} (1 - p)^{(n-1)-(j-1)}
\]

\[
= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!((n-1)-k)!} p^k (1 - p)^{(n-1)-k}
\]

\[
= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1 - p)^{(n-1)-k} = np.
\]