Outline

1. Review

2. Covariance and Correlation

3. Coupon Collecting

4. Loose Ends: Random Facts about Random Things
Expectation and Variance Review

- The expected value $E[X]$ of a random variable $X$ is a probability-weighted average of the possible values of $X$:

$$E[X] = \sum_k k P(X = k)$$

- If $X$ is a random variable and $f : \mathbb{R} \to \mathbb{R}$ then $Y = f(X)$ is also a random variable with expectation

$$E(Y) = \sum_k f(k)P(X = k)$$

- The variance is quantifies how close to $\mu = E[X]$ we expect $X$ to be:

$$\text{var}(X) = \sum_k (k - \mu)^2 P(X = k) = E[X^2] - \mu^2.$$  

and the standard deviation of $X$ is $\sigma_X = \sqrt{\text{var}(X)}$
Given two random variables, $X$ and $Y$ mapping from $\Omega$ to $\mathbb{R}$, we can define events of the form

$$\{X = i, Y = j\} = \{X = i\} \cap \{Y = j\}$$
Multiple Random Variables

- Given two random variables, $X$ and $Y$ mapping from $\Omega$ to $\mathbb{R}$, we can define events of the form

$$\{X = i, Y = j\} = \{X = i\} \cap \{Y = j\} = \{o \in \Omega \mid X(o) = i \text{ and } Y(o) = j\}$$

- The probabilities of these events give the joint PMF of $X$ and $Y$:

$$P(X = i, Y = j) = P(\{X = i, Y = j\})$$

- Given the joint PMF, we can compute the marginal probabilities:
Multiple Random Variables

- Given two random variables, $X$ and $Y$ mapping from $\Omega$ to $\mathbb{R}$, we can define events of the form

$$\{X = i, Y = j\} = \{X = i\} \cap \{Y = j\} = \{o \in \Omega \mid X(o) = i \text{ and } Y(o) = j\}$$

- The probabilities of these events give the joint PMF of $X$ and $Y$:

$$P(X = i, Y = j) = P(\{X = i, Y = j\})$$

- Given the joint PMF, we can compute the marginal probabilities:

$$P(X = i) = \sum_j P(X = i, Y = j)$$

$$P(Y = j) = \sum_i P(X = i, Y = j)$$
Functions of Multiple Random Variables

Given random variables $X_1, X_2, \ldots, X_N$ and $f : \mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R} \to \mathbb{R}$, 

$$Z = f(X_1, X_2, \ldots, X_N)$$ 

is a new random variable with expectation 

$$E(Z) = \sum_{a_1, a_2, \ldots, a_N} f(a_1, a_2, \ldots, a_N)P(X_1 = a_1, X_2 = a_2, \ldots, X_N = a_N)$$
Functions of Multiple Random Variables

- Given random variables $X_1, X_2, \ldots, X_N$ and $f : \mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R} \to \mathbb{R}$,

$$Z = f(X_1, X_2, \ldots, X_N)$$

is a new random variable with expectation

$$E(Z) = \sum_{a_1, a_2, \ldots, a_N} f(a_1, a_2, \ldots, a_N)P(X_1 = a_1, X_2 = a_2, \ldots, X_N = a_N)$$

- Linearity of Expectation: If $Z = \sum_{i=1}^{N} c_i X_i$,

$$E(Z) = E(\sum_{i=1}^{N} c_i X_i) = \sum_{i=1}^{N} c_i E(X_i)$$
Functions of Multiple Random Variables

- Given random variables $X_1, X_2, \ldots, X_N$ and $f : \mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R} \to \mathbb{R}$,
  
  $$Z = f(X_1, X_2, \ldots, X_N)$$

  is a new random variable with expectation

  $$E(Z) = \sum_{a_1, a_2, \ldots, a_N} f(a_1, a_2, \ldots, a_N) P(X_1 = a_1, X_2 = a_2, \ldots, X_N = a_N)$$

- Linearity of Expectation: If $Z = \sum_{i=1}^N c_i X_i$,
  
  $$E(Z) = E\left(\sum_{i=1}^N c_i X_i\right) = \sum_{i=1}^N c_i E(X_i)$$

- Linearity of Variance: If $Z = \sum_{i=1}^N c_i X_i$,
  
  $$\text{var}(Z) = \text{var}\left(\sum_{i=1}^N c_i X_i\right) = \sum_{i=1}^N c_i^2 \text{var}(X_i)$$

  if $X_1, \ldots, X_N$ are pairwise independent, i.e., for all $i, j, a, b$
  
  $$P(X_i = a, X_j = b) = P(X_i = a)P(X_j = b).$$
Outline

1. Review
2. Covariance and Correlation
3. Coupon Collecting
4. Loose Ends: Random Facts about Random Things
Independence

Two discrete random variables $X$ and $Y$ are independent if and only if $P(X = a, Y = b) = P(X = a)P(Y = b)$ for all $a$ and $b$. 
Independence

- Two discrete random variables $X$ and $Y$ are independent if and only if $P(X = a, Y = b) = P(X = a)P(Y = b)$ for all $a$ and $b$.
- When two random variables are not independent, it’s natural to want to measure how dependent they are.
The covariance between $X$ and $Y$ is one measure of dependence that quantifies the degree to which there is a linear relationship between $X$ and $Y$.

$$cov(X, Y) = E[(X - E[X])(Y - E[Y])]$$

$$= E[XY] - E[X]E[Y]$$
Quantifying Dependence: Covariance

- The **covariance** between $X$ and $Y$ is one measure of dependence that quantifies the degree to which there is a *linear relationship* between $X$ and $Y$.

$$
$$

- The covariance of $X$ and $Y$ is positive if when $X$ is large, $Y$ is also large. It’s negative if when $X$ is large, $Y$ is small.
Quantifying Dependence: Covariance

- The **covariance** between $X$ and $Y$ is one measure of dependence that quantifies the degree to which there is a **linear relationship** between $X$ and $Y$.

$$\text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$
$$\quad = E[XY] - E[X]E[Y]$$

- The covariance of $X$ and $Y$ is positive if when $X$ is large, $Y$ is also large. It’s negative if when $X$ is large, $Y$ is small.

- If $X$ and $Y$ are independent then $\text{cov}(X, Y) = 0$ but $\text{cov}(X, Y) = 0$ does not necessarily imply that $X$ and $Y$ are independent.
The **covariance** between $X$ and $Y$ is one measure of dependence that quantifies the degree to which there is a **linear relationship** between $X$ and $Y$.

\[
\text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])] \\
= E[XY] - E[X]E[Y]
\]

The covariance of $X$ and $Y$ is positive if when $X$ is large, $Y$ is also large. It’s negative if when $X$ is large, $Y$ is small.

If $X$ and $Y$ are independent then $\text{cov}(X, Y) = 0$ but $\text{cov}(X, Y) = 0$ does not necessarily imply that $X$ and $Y$ are independent.

We can write $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y)$. 
Example

<table>
<thead>
<tr>
<th>P(X,Y)</th>
<th>Y = 0</th>
<th>Y = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>X = 0</td>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>X = 1</td>
<td>0.2</td>
<td>0.3</td>
</tr>
</tbody>
</table>

- $P(X = 0) = 0.5$, $P(X = 1) = 0.5$ and so $E[X] = 0.5$
Example

<table>
<thead>
<tr>
<th>P(X,Y)</th>
<th>Y = 0</th>
<th>Y = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>X = 0</td>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>X = 1</td>
<td>0.2</td>
<td>0.3</td>
</tr>
</tbody>
</table>

- $P(X = 0) = 0.5$, $P(X = 1) = 0.5$ and so $E[X] = 0.5$
- $P(Y = 0) = 0.6$, $P(Y = 1) = 0.4$ and so $E[Y] = 0.4$
Example

<table>
<thead>
<tr>
<th>P(X,Y)</th>
<th>Y = 0</th>
<th>Y = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>X = 0</td>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>X = 1</td>
<td>0.2</td>
<td>0.3</td>
</tr>
</tbody>
</table>

- \( P(X = 0) = 0.5, P(X = 1) = 0.5 \) and so \( E[X] = 0.5 \)
- \( P(Y = 0) = 0.6, P(Y = 1) = 0.4 \) and so \( E[Y] = 0.4 \)
- \( E[XY] \) can be computed as follows

\[
E[XY] = 0 \times 0 \times P(X = 0, Y = 0) + 0 \times 1 \times P(X = 0, Y = 1) + 1 \times 0 \times P(X = 1, Y = 0) + 1 \times 1 \times P(X = 1, Y = 1)
\]
\[
= 0.3
\]
Example

<table>
<thead>
<tr>
<th>P(X,Y)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X\Y</td>
<td>Y = 0</td>
<td>Y = 1</td>
</tr>
<tr>
<td>X = 0</td>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>X = 1</td>
<td>0.2</td>
<td>0.3</td>
</tr>
</tbody>
</table>

- $P(X = 0) = 0.5$, $P(X = 1) = 0.5$ and so $E[X] = 0.5$
- $P(Y = 0) = 0.6$, $P(Y = 1) = 0.4$ and so $E[Y] = 0.4$
- $E[XY]$ can be computed as follows

\[
E[XY] = 0 \times 0 \times P(X = 0, Y = 0) + 0 \times 1 \times P(X = 0, Y = 1) + 1 \times 0 \times P(X = 1, Y = 0) + 1 \times 1 \times P(X = 1, Y = 1)
\]
\[
= 0.3
\]

- $\text{cov}(X, Y) = E(XY) - E(X)E(Y) = 0.3 - 0.5 \times 0.4 = 0.1$
Quantifying Dependence: Correlation

- The maximum magnitude of the covariance depends on the variance of $X$ and the variance of $Y$. 
The maximum magnitude of the covariance depends on the variance of $X$ and the variance of $Y$.

The **correlation** between $X$ and $Y$ is closely related to the covariance, but is normalized to the range $[-1, 1]$. 1 indicates maximum positive covariance and $-1$ indicates maximum negative covariance:

$$\rho(X, Y) = corr(X, Y) = \frac{cov(X, Y)}{\sqrt{var(X)var(Y)}}$$
Visualizing Correlations: Height vs Weight ($\rho = 0.56$)
Visualizing Correlations: Linear vs Non-Linear
Causation

**Question:** When two random variables are correlated does this mean one random variable causes the other?
Causation

- **Question:** When two random variables are correlated does this mean one random variable causes the other?

- **Example:** There are more fireman at the scene of larger fires? Do fireman cause an increase in the size of a fire.
Causation

- **Question:** When two random variables are correlated does this mean one random variable causes the other?

- **Example:** There are more fireman at the scene of larger fires? Do fireman cause an increase in the size of a fire.

- **Example:** More people drown on days where a lot of ice cream is sold. Does ice cream cause drowning?
Causation

- **Question**: When two random variables are correlated does this mean one random variable causes the other?

- **Example**: There are more firemen at the scene of larger fires? Do firemen cause an increase in the size of a fire.

- **Example**: More people drown on days where a lot of ice cream is sold. Does ice cream cause drowning?

- **Example**: In the height/weight example, height and weight were positively correlated. Does increasing your weight make you taller?
Causation

■ **Question:** When two random variables are correlated does this mean one random variable causes the other?

■ **Example:** There are more fireman at the scene of larger fires? Do fireman cause an increase in the size of a fire.

■ **Example:** More people drown on days where a lot of ice cream is sold. Does ice cream cause drowning?

■ **Example:** In the height/weight example, height and weight were positively correlated. Does increasing your weight make you taller?

■ **Example:** When you see a wind turbine turning it is usually windy. Do wind turbines create wind?
Causation

Given two correlated random variables \( X \) and \( Y \):

- \( X \) might cause \( Y \) (i.e., causation)
Causation

Given two correlated random variables $X$ and $Y$:

- $X$ might cause $Y$ (i.e., causation)
- $Y$ might cause $X$ (i.e., reverse causation)
Causation

Given two correlated random variables $X$ and $Y$:

- $X$ might cause $Y$ (i.e., causation)
- $Y$ might cause $X$ (i.e., reverse causation)
- A third random variable $Z$ might cause $X$ and $Y$ (i.e., common cause)
Causation

Given two correlated random variables $X$ and $Y$:

- $X$ might cause $Y$ (i.e., causation)
- $Y$ might cause $X$ (i.e., reverse causation)
- A third random variable $Z$ might cause $X$ and $Y$ (i.e., common cause)
- A combination of all of these (e.g., self-reinforcement)
Causation

Given two correlated random variables $X$ and $Y$:

- $X$ might cause $Y$ (i.e., causation)
- $Y$ might cause $X$ (i.e., reverse causation)
- A third random variable $Z$ might cause $X$ and $Y$ (i.e., common cause)
- A combination of all of these (e.g., self-reinforcement)
- The correlation might be spurious due to small sample size
Outline

1. Review
2. Covariance and Correlation
3. Coupon Collecting
4. Loose Ends: Random Facts about Random Things
Coupon Collecting/Shuffle Mode

- You have $n$ songs on your phone.
Coupon Collecting/Shuffle Mode

- You have $n$ songs on your phone.
- In **shuffle mode**, the player picks songs uniformly at random.
Coupon Collecting/Shuffle Mode

- You have $n$ songs on your phone.
- In shuffle mode, the player picks songs uniformly at random.
- Let $T$ be the total number of songs played until every song is played.
Coupon Collecting/Shuffle Mode

- You have $n$ songs on your phone.
- In **shuffle mode**, the player picks songs uniformly at random.
- Let $T$ be the total number of songs played until every song is played.
- $T$ could be infinite or as small as $n$. 

You have $n$ songs on your phone.

In shuffle mode, the player picks songs uniformly at random.

Let $T$ be the total number of songs played until every song is played.

$T$ could be infinite or as small as $n$.

For this section, recall that if $X$ is a geometric random variable with parameter $p$ then $P(X = k) = (1 - p)^{k-1} p$ and has expectation $1/p$. 
What’s the probability that $T = n$?
What’s the probability that $T = n$?

- What’s the probability that $T = n$?
- Number of possible sequences of $n$ songs:
What’s the probability that $T = n$?

Number of possible sequences of $n$ songs: $n^n$
What’s the probability that $T = n$?

- What’s the probability that $T = n$?
- Number of possible sequences of $n$ songs: $n^n$
- Number of possible sequences of $n$ songs including every song:
What’s the probability that \( T = n \)?

- What’s the probability that \( T = n \)?
- Number of possible sequences of \( n \) songs: \( n^n \)
- Number of possible sequences of \( n \) songs including every song: \( n! \)
What’s the probability that $T = n$?

- Number of possible sequences of $n$ songs: $n^n$
- Number of possible sequences of $n$ songs including every song: $n!$
- Therefore, probability is:
What’s the probability that $T = n$?

- Number of possible sequences of $n$ songs: $n^n$
- Number of possible sequences of $n$ songs including every song: $n!$
- Therefore, probability is:

$$\frac{n!}{n^n} = \frac{n}{n} \times \frac{n-1}{n} \times \ldots \times \frac{1}{n}$$
What’s the probability that $T = n$?

- Number of possible sequences of $n$ songs: $n^n$
- Number of possible sequences of $n$ songs including every song: $n!$
- Therefore, probability is:

$$\frac{n!}{n^n} = \frac{n}{n} \times \frac{n-1}{n} \times \ldots \times \frac{1}{n} \leq 2^{-n/2}$$
Expected Value of $T$

To analyze $E[T]$, we define $C_1, C_2, \ldots, C_n$ where $C_i = \text{songs played after } (i - 1)\text{-th new song until } i\text{-th new song is played}$ and note that $T = \sum_{i=1}^{n} C_i$

By linearity of expectation:

$E[T] = \sum_{i=1}^{n} E[C_i]$

$C_i$ is a geometric random variable with $P(C_i = j) = p_i (1 - p_i)^{j-1}$ for $j = 1, 2, \ldots$

where $p_i = \frac{n - i + 1}{n}$

$E[C_i] = \frac{1}{p_i} = \frac{n}{n - i + 1}$

So $E[T] = \sum_{i=1}^{n} \frac{n}{n - i + 1} = n H_n \approx n \ln n$
Expected Value of $T$

To analyze $E [T]$, we define $C_1, C_2, \ldots, C_n$ where

$$C_i = \text{songs played after} \ (i - 1)-\text{th new song until} \ i-\text{th new song is played}$$

and note that $T = \sum_{i=1}^{n} C_i$
Expected Value of $T$

- To analyze $E[T]$ we define $C_1, C_2, \ldots, C_n$ where
  
  $$C_i = \text{songs played after } (i - 1)-\text{th new song until } i\text{-th new song is played}$$
  
  and note that $T = \sum_{i=1}^{n} C_i$

- By linearity of expectation:

  $$E[T] = \sum_{i=1}^{n} E[C_i]$$
Expected Value of $T$

- To analyze $E[T]$ we define $C_1, C_2, \ldots, C_n$ where
  \[ C_i = \text{songs played after $(i-1)$-th new song until $i$-th new song is played} \]
  and note that $T = \sum_{i=1}^{n} C_i$
- By linearity of expectation:
  \[ E[T] = \sum_{i=1}^{n} E[C_i] \]
- $C_i$ is a geometric random variable with
  \[ P(C_i = j) = p_i(1 - p_i)^{j-1} \quad \text{for } j = 1, 2, \ldots \]
  where $p_i = \frac{n-i+1}{n}$
Expected Value of $T$

- To analyze $E [T]$ we define $C_1, C_2, \ldots, C_n$ where
  
  \[ C_i = \text{songs played after (i − 1)-th new song until i-th new song is played} \]

  and note that $T = \sum_{i=1}^{n} C_i$

- By linearity of expectation:
  
  \[ E [T] = \sum_{i=1}^{n} E [C_i] \]

- $C_i$ is a geometric random variable with
  
  \[ P (C_i = j) = p_i (1 - p_i)^{j-1} \quad \text{for } j = 1, 2, \ldots \]

  where $p_i = \frac{n-i+1}{n}$

- $E [C_i] = \frac{1}{p_i} = \frac{n}{n-i+1}$
Expected Value of $T$

- To analyze $E[T]$ we define $C_1, C_2, \ldots, C_n$ where

$$C_i = \text{songs played after } (i - 1)-\text{th new song until } i\text{-th new song is played}$$

and note that $T = \sum_{i=1}^{n} C_i$

- By linearity of expectation:

$$E[T] = \sum_{i=1}^{n} E[C_i]$$

- $C_i$ is a geometric random variable with

$$P(C_i = j) = p_i(1 - p_i)^{j-1} \quad \text{for } j = 1, 2, \ldots$$

where $p_i = \frac{n-i+1}{n}$

- $E[C_i] = \frac{1}{p_i} = \frac{n}{n-i+1}$

- So

$$E[T] = \frac{n}{n} + \frac{n}{n-1} + \ldots + \frac{n}{1} = nH_n \approx n \ln n$$
Outline

1. Review
2. Covariance and Correlation
3. Coupon Collecting
4. Loose Ends: Random Facts about Random Things
Secrets of the Chebyshev Bound

- **Chebyshev Bound:**

\[
P(X \leq E(X) - c) + P(X \geq E(X) + c) = P(|X - E(X)| \geq c) \leq \frac{\text{Var}(X)}{c^2}
\]
Secrets of the Chebyshev Bound

- Chebyshev Bound:

\[ P(X \leq E(X) - c) + P(X \geq E(X) + c) = P(|X - E(X)| \geq c) \leq \frac{Var(X)}{c^2} \]

- The bound is useful when we are trying to bound the probability that \( X \) is much smaller or larger than its expectation.
Secrets of the Chebyshev Bound

- Chebyshev Bound:

\[ P(X \leq E(X) - c) + P(X \geq E(X) + c) = P(|X - E(X)| \geq c) \leq \frac{\text{Var}(X)}{c^2} \]

- The bound is useful when we are trying to bound the probability that \( X \) is much smaller or larger than its expectation.

- However, it also implies bounds on just one tail.
Secrets of the Chebyshev Bound

- Chebyshev Bound:

\[ P(X \leq E(X) - c) + P(X \geq E(X) + c) = P(|X - E(X)| \geq c) \leq \frac{\text{Var}(X)}{c^2} \]

- The bound is useful when we are trying to bound the probability that \( X \) is much smaller or larger than its expectation.

- However, it also implies bounds on just one tail.

- For example, if \( E(X) = 10 \) and \( \text{var}(X) = 2 \) then

\[ P(X \geq 15) = P(X \geq E(X) + 5) \leq P(|X - E(X)| \geq 5) \leq \frac{2}{25} \]
Poisson Expectation

For a Poisson random variable, \( P(X = k) = \frac{e^{-\lambda}}{k!} \lambda^k \). Hence,

\[
E[X] = \sum_{k=0}^{\infty} k \cdot \frac{e^{-\lambda}}{k!} \lambda^k
\]

\[
= \lambda \sum_{k=1}^{\infty} k \cdot \frac{e^{-\lambda}}{k!} \lambda^{k-1}
\]

\[
= \lambda \sum_{k=1}^{\infty} \frac{e^{-\lambda}}{(k-1)!} \lambda^{k-1}
\]

\[
= \lambda(P(X = 0) + P(X = 1) + P(X = 2) + \ldots)
\]

\[
= \lambda
\]

The last line follows because the events \( \{X = 0\}, \{X = 1\}, \{X = 2\}, \ldots \) partition the sample space and hence the probabilities sum up to 1.
Poisson Expectation

For a Poisson random variable, \( P(X = k) = \frac{e^{-\lambda}}{k!} \lambda^k \). Hence,

\[
E[X] = \sum_{k=0}^{\infty} k \cdot \frac{e^{-\lambda}}{k!} \lambda^k
\]

\[
= \lambda \sum_{k=1}^{\infty} k \cdot \frac{e^{-\lambda}}{k!} \lambda^{k-1}
\]

\[
= \lambda \sum_{k=1}^{\infty} \frac{e^{-\lambda}}{(k-1)!} \lambda^{k-1}
\]

\[
= \lambda (P(X = 0) + P(X = 1) + P(X = 2) + \ldots)
\]

\[
= \lambda
\]

The last line follows because the events \( \{X = 0\}, \{X = 1\}, \{X = 2\}, \ldots \) partition the sample space and hence the probabilities sum up to 1.
For a Geometric random variable, $P(X = k) = (1 - p)^{k-1} p$. You’ll prove in the homework that:

- $E[X] = P(X \geq 1) + P(X \geq 2) + P(X \geq 3) \ldots$
- $P(X \geq k) = (1 - p)^{k-1}$

Using these,

\[
E[X] = P(X \geq 1) + P(X \geq 2) + P(X \geq 3) \ldots \\
= 1 + (1 - p) + (1 - p)^2 + \ldots \\
= \frac{1}{p}
\]
Alternative Expression for Expectation

If $Y = f(X)$, we can write $E[Y] = \sum_k f(k)P(X = k)$. 
Alternative Expression for Expectation

- If \( Y = f(X) \), we can write \( E[Y] = \sum_k f(k)P(X = k) \).
- Use the fact that \( P(Y = r) = \sum_{k:f(k)=r} P(X = k) \) and then,

\[
E[Y]
\]
Alternative Expression for Expectation

- If $Y = f(X)$, we can write $E[Y] = \sum_k f(k)P(X = k)$.
- Use the fact that $P(Y = r) = \sum_{k:f(k)=r} P(X = k)$ and then,

$$E[Y] = \sum_r rP(Y = r)$$
Alternative Expression for Expectation

- If $Y = f(X)$, we can write $E[Y] = \sum_k f(k)P(X = k)$.
- Use the fact that $P(Y = r) = \sum_{k:f(k)=r} P(X = k)$ and then,

$$E[Y] = \sum_r rP(Y = r) = \sum_r r \sum_{k:f(k)=r} P(X = k)$$
Alternative Expression for Expectation

- If $Y = f(X)$, we can write $E[Y] = \sum_k f(k)P(X = k)$.
- Use the fact that $P(Y = r) = \sum_{k:f(k)=r} P(X = k)$ and then,

\[
E[Y] = \sum_r rP(Y = r)
\]

\[
= \sum_r r \sum_{k:f(k)=r} P(X = k)
\]

\[
= \sum_r \sum_{k:f(k)=r} rP(X = k)
\]
Alternative Expression for Expectation

- If $Y = f(X)$, we can write $E[Y] = \sum_k f(k)P(X = k)$.
- Use the fact that $P(Y = r) = \sum_{k:f(k)=r} P(X = k)$ and then,

$$E[Y] = \sum_r rP(Y = r)$$

$$= \sum_r r \sum_{k:f(k)=r} P(X = k)$$

$$= \sum_r \sum_{k:f(k)=r} rP(X = k)$$

$$= \sum_r \sum_{k:f(k)=r} f(k)P(X = k)$$
If $Y = f(X)$, we can write $E[Y] = \sum_k f(k)P(X = k)$.

Use the fact that $P(Y = r) = \sum_{k:f(k)=r} P(X = k)$ and then,

$$E[Y] = \sum_r rP(Y = r)$$

$$= \sum_r r \sum_{k:f(k)=r} P(X = k)$$

$$= \sum_r \sum_{k:f(k)=r} rP(X = k)$$

$$= \sum_r \sum_{k:f(k)=r} f(k)P(X = k)$$

$$= \sum_k f(k)P(X = k)$$
Suppose we have some Bernoulli random variables $X_1, X_2, \ldots, X_n$ where for all $i < j$ the joint probabilities are given in the following table:

<table>
<thead>
<tr>
<th>$X_i \setminus X_j$</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>1</td>
<td>0.25</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Are the variables pairwise independent? I.e., for all $i < j$ and $a, b \in \{0, 1\}$, $P(X_i = a, X_j = b) = P(X_i = a)P(X_j = b)$. Yes.

Are they also three-wise independent? I.e., for all $i < j < k$ and $a, b, c \in \{0, 1\}$, $P(X_i = a, X_j = b, X_k = c) = P(X_i = a)P(X_j = b)P(X_k = c)$. Not necessarily, e.g., let $X_1$ and $X_2$ be the result of tossing two independent coins and $X_3 = X_1 + X_2 \mod 2$. 
Suppose we have some bernoulli random variables $X_1, X_2, \ldots, X_n$ where for all $i < j$ the joint probabilities are given in the following table:

<table>
<thead>
<tr>
<th>$X_i \backslash X_j$</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>1</td>
<td>0.25</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Are the variables pairwise independent? I.e., for all $i < j$ and $a, b \in \{0, 1\}$, $P(X_i = a, X_j = b) = P(X_i = a)P(X_j = b)$. 

Not necessarily, e.g., let $X_1$ and $X_2$ be the result of tossing two independent coins and $X_3 = X_1 + X_2 \text{(mod 2)}$. 

Secrets of Pairwise Independence
Secrets of Pairwise Independence

- Suppose we have some Bernoulli random variables $X_1, X_2, \ldots, X_n$ where for all $i < j$ the joint probabilities are given in the following table:

<table>
<thead>
<tr>
<th>$X_i \setminus X_j$</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>1</td>
<td>0.25</td>
<td>0.25</td>
</tr>
</tbody>
</table>

- Are the variables pairwise independent? I.e., for all $i < j$ and $a, b \in \{0, 1\}$, $P(X_i = a, X_j = b) = P(X_i = a)P(X_j = b)$. Yes.
Suppose we have some Bernoulli random variables $X_1, X_2, \ldots, X_n$ where for all $i < j$ the joint probabilities are given in the following table:

<table>
<thead>
<tr>
<th>$X_i \setminus X_j$</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>1</td>
<td>0.25</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Are the variables pairwise independent? I.e., for all $i < j$ and $a, b \in \{0, 1\}$, $P(X_i = a, X_j = b) = P(X_i = a)P(X_j = b)$. Yes.

Are they also three-wise independent? I.e., for all $i < j < k$ and $a, b, c \in \{0, 1\}$

$$P(X_i = a, X_j = b, X_k = c) = P(X_i = a)P(X_j = b)P(X_k = c)$$
Suppose we have some Bernoulli random variables $X_1, X_2, \ldots, X_n$ where for all $i < j$ the joint probabilities are given in the following table:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>1</td>
<td>0.25</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Are the variables pairwise independent? I.e., for all $i < j$ and $a, b \in \{0, 1\}$, $P(X_i = a, X_j = b) = P(X_i = a)P(X_j = b)$. Yes.

Are they also three-wise independent? I.e., for all $i < j < k$ and $a, b, c \in \{0, 1\}$

$$P(X_i = a, X_j = b, X_k = c) = P(X_i = a)P(X_j = b)P(X_k = c)$$

Not necessarily, e.g., let $X_1$ and $X_2$ be the result of tossing two independent coins and $X_3 = X_1 + X_2 \pmod{2}$. 