Probability Review

Basic definitions

**Discrete probability distribution:** a function $Pr : \Omega \rightarrow [0, 1]$ such that $\sum_{\omega \in \Omega} Pr(\omega) = 1$

- $\Omega$ called *sample space*
- a point $\omega \in \Omega$ represents the *outcome* of some experiment
- $Pr(\omega)$ represents the probability of outcome $\omega$
- $\Omega$ may be *finite* or *countably infinite*
Example: rolling a die. $\Omega = \{1, \ldots, 6\}$, $\Pr(\omega) = 1/6$ for all $\omega \in \Omega$

Example: uniform distribution. $|\Omega| = n$, $\Pr(\omega) = 1/n$ for all $\omega \in \Omega$

Example: Bernoulli trial. An experiment with two outcomes. Probability of “success” is $p$, probability of “failure” is $q := 1 - p$. 
An event is a subset $A \subseteq \Omega$

The probability of $A$ is $\Pr[A] := \sum_{\omega \in A} \Pr(\omega)$

Logical operations:

- $A \cap B$ — logical AND
- $A \cup B$ — logical OR
- $\Omega \setminus A$ — logical NOT

Union bounds:

- $\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B]$
- For any family of events $\{A_i\}_{i \in I}$:

\[
\Pr\left[ \bigcup_{i \in I} A_i \right] \leq \sum_{i \in I} \Pr[A_i]
\]

and equality holds if the $A_i$’s are pairwise disjoint
Example (Alice and Bob)

Alice rolls two dice, and asks Bob to guess a value that appears on either of the two dice (without looking)

What is the probability that Bob guesses correctly?

Model: uniform distribution on \( \Omega := \{1, \ldots, 6\} \times \{1, \ldots, 6\} \)

For \((s, t) \in \Omega\): \(s = \) first die, \(t = \) second die

For \(k = 1, \ldots, 6\), define

- event \(A_k\) : first die = \(k\)
- event \(B_k\) : second die = \(k\)
- \(C_k := A_k \cup B_k\) (\(k\) appears on either die)

Pr[\(A_k\)] = 6/36 = 1/6, Pr[\(B_k\)] = 6/36 = 1/6, Pr[\(A_k \cap B_k\)] = 1/36

Therefore:

\[
\Pr[C_k] = \Pr[A_k \cup B_k] = \Pr[A_k] + \Pr[B_k] - \Pr[A_k \cap B_k] \\
= 1/6 + 1/6 - 1/36 = 11/36
\]

So no matter Bob’s guess, he is correct with probability \(11/36 < 1/3\)
Conditional probability and independence

Suppose \( \Pr[B] \neq 0 \)

Define \( \Pr(\omega \mid B) := \begin{cases} \frac{\Pr(\omega)}{\Pr[B]} & \text{if } \omega \in B, \\ 0 & \text{otherwise.} \end{cases} \)

\( \Pr(\cdot \mid B) \) is a new probability distribution on \( \Omega \): the conditional distribution given \( B \)

**Intuition:**

- we run an experiment
- we learn that \( B \) occurs
- then \( \Pr(\cdot \mid B) \) assigns new probabilities to all outcomes, reflecting this partial knowledge
For any event $\mathcal{A}$:

$$\Pr[\mathcal{A} \mid \mathcal{B}] = \sum_{\omega \in \mathcal{A}} \Pr(\omega \mid \mathcal{B}) = \frac{\Pr[\mathcal{A} \cap \mathcal{B}]}{\Pr[\mathcal{B}]}.$$

$\mathcal{A}$ and $\mathcal{B}$ are called **independent** if

- $\Pr[\mathcal{A} \cap \mathcal{B}] = \Pr[\mathcal{A}] \cdot \Pr[\mathcal{B}]$,
- or equivalently, $\Pr[\mathcal{A}] = \Pr[\mathcal{A} \mid \mathcal{B}]$

**Intuition:**

- we run an experiment
- we learn that $\mathcal{B}$ occurs
- then $\Pr[\mathcal{A} \mid \mathcal{B}]$ tells us how likely it is for $\mathcal{A}$ to occur, given this partial knowledge
- independence means: learning that $\mathcal{B}$ occurs tells us nothing about $\mathcal{A}$
Back to Alice and Bob . . .

Suppose Alice tells Bob the sum of the two dice before he guesses. For example, suppose sum = 4. What is Bob’s best strategy?

For $\ell = 2, \ldots, 12$, define event $D_\ell$: sum = $\ell$

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

$Pr[C_1 | D_4] = (2/36)/(3/36) = 2/3$

$Pr[C_2 | D_4] = (1/36)/(3/36) = 1/3$

$Pr[C_3 | D_4] = (2/36)/(3/36) = 2/3$

$Pr[C_4 | D_4] = Pr[C_5 | D_4] = Pr[C_6 | D_4] = 0$

Bob’s best choice: 1 or 3
Total probability

Suppose \( \{B_i\}_{i \in I} \) is a partition of \( \Omega \)

Let \( A \) be any event

**Law of total probability:**

\[
\Pr[A] = \sum_{i \in I} \Pr[A \cap B_i] = \sum_{i \in I} \Pr[A | B_i] \Pr[B_i]
\]
Back to Alice and Bob . . .

Let us compute Bob’s overall winning probability

<table>
<thead>
<tr>
<th></th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

If the sum = 2 or = 12, Bob wins for sure

Suppose sum = \( \ell \), with \( 1 < \ell < 12 \), and \( N_\ell \) is the number of pairs with sum = \( \ell \)

Bob can always choose a value that appears twice among these \( N_\ell \) pairs (for example, Bob can choose 1 if \( \ell \leq 7 \) and 6 if \( \ell > 7 \))

Let \( C \) be the event that Bob wins

Total probability: \( \Pr[C] = \sum_{\ell=2}^{12} \Pr[C | D_\ell] \Pr[D_\ell] \)
Alice and Bob (cont’d)

We have

\[ \Pr[C | D_2] \Pr[D_2] = 1 \cdot \frac{1}{36} = \frac{1}{36} \]

\[ \Pr[C | D_{12}] \Pr[D_{12}] = 1 \cdot \frac{1}{36} = \frac{1}{36} \]

For \( \ell = 3, \ldots, 11 \), we have

\[ \Pr[C | D_\ell] \Pr[D_\ell] = \frac{2}{N_\ell} \cdot \frac{N_\ell}{36} = \frac{1}{18} \]

Therefore,

\[ \Pr[C] = \frac{1}{36} + \frac{1}{36} + \frac{9}{18} = \frac{10}{18}. \]
Random variables

A random variable taking values in a set $S$:

$$X : \Omega \rightarrow S$$

For $s \in S$, the event “$X = s$” is \{\(\omega \in \Omega : X(\omega) = s\}\}, and

$$Pr[X = s] = \sum_{\omega \in \Omega, X(\omega) = s} Pr(\omega)$$

Building new random variables:

- $Y = f(X)$ means $Y(\omega) = f(X(\omega))$ for all $\omega \in \Omega$
- $Z = X + Y$ means $Z(\omega) = X(\omega) + Y(\omega)$ for all $\omega \in \Omega$
A random variable $X$ taking values in $S$ defines a probability distribution on $S$:

$$\Pr_X(s) = \Pr[X = s]$$

For an event $\mathcal{A}$, we can define the **indicator variable**:

$$X_{\mathcal{A}}(\omega) := \begin{cases} 
1 & \text{if } \omega \in \mathcal{A}, \\
0 & \text{otherwise}
\end{cases}$$
Alice and Bob again . . .

$X$ is the value of the first die

- $X$ is uniformly distributed over $\{1, \ldots, 6\}$

$Y$ is the value of the second die

- $Y$ is uniformly distributed over $\{1, \ldots, 6\}$

Define $Z := X + Y$


Define $W$ to be the indicator for the event that $X = 1$ or $Y = 1$

- $\Pr[W = 1] = 11/36$, $\Pr[W = 0] = 1 - 11/36 = 25/36$
Independent random variables

$X$ takes values in $S$, $Y$ takes values in $T$

$X$ and $Y$ are called **independent** if

$$\Pr[(X = s) \cap (Y = t)] = \Pr[X = s] \cdot \Pr[Y = t]$$

for all $s \in S$ and $t \in T$

Equivalently,

$$\Pr[X = s \mid Y = t] = \Pr[X = s]$$

for all $s \in S$ and $t \in T$

**Intuition:** learning the value of $Y$ gives us no information about the value of $X$
Alice and Bob again . . .

\( X \) is the value of the first die

\( Y \) is the value of the second die

\[ Z := X + Y \]

\( X \) and \( Y \) are independent

\( X \) and \( Z \) are not independent

\( Y \) and \( Z \) are not independent
Example: sum mod m.

Suppose $X$ and $Y$ are independent random variables, with each uniformly distributed over $\mathbb{Z}_m$.

This means that $(X, Y)$ is uniformly dist’d over $\mathbb{Z}_m \times \mathbb{Z}_m$.

Set $Z := X + Y$.

Claim: $Z$ is uniformly distributed over $\mathbb{Z}_m$.

1. Why? For each $\alpha \in \mathbb{Z}_m$, there are $m$ solutions $(s, t) \in \mathbb{Z}_m \times \mathbb{Z}_m$ to the equation $s + t = \alpha$.

Claim: $X$ and $Z$ are independent.

Let $\alpha, \beta \in \mathbb{Z}_m$ be fixed.

Want to show $\Pr[(X = \alpha) \cap (Z = \beta)] = 1/m^2$.

\[
\Pr[(X = \alpha) \cap (Z = \beta)] = \Pr[(X = \alpha) \cap (X + Y = \beta)] \\
= \Pr[(X = \alpha) \cap (\alpha + Y = \beta)] \\
= \Pr[(X = \alpha) \cap (Y = \beta - \alpha)] \\
= \Pr[X = \alpha] \cdot \Pr[Y = \beta - \alpha] \quad (X, Y \text{ indep.}) \\
= (1/m) \cdot (1/m) = 1/m^2
\]
**Example:** *one-time pad.*

Suppose $X$ and $Y$ are independent random variables, where $Y$ is uniformly distributed over $\mathbb{Z}_m$

$X$ may have an arbitrary distribution over $\mathbb{Z}_m$

Set $Z := X + Y$

*Fact:* $X$ and $Z$ are independent

**Application to cryptography**

Suppose $Y$ represents an encryption key shared between Alice and Bob

Alice encrypts a message $X$ by computing the ciphertext $Z = X + Y$ and sends $Z$ over an insecure network

Bob can decrypt the ciphertext by computing $X = Z - Y$

Independence of $Z$ and $X$ ensures that an eavesdropper who only learns the value of the ciphertext $Z$ learns nothing about the message $X$
Mutual and $k$-wise independence

Let $\{X_i\}_{i \in I}$ be a finite family of random variables

Let us call a corresponding family of values $\{s_i\}_{i \in I}$ an **assignment** to $\{X_i\}_{i \in I}$ if $s_i$ is in the image of $X_i$ for each $i \in I$

$\{X_i\}_{i \in I}$ is called **mutually independent** if for every assignment $\{s_i\}_{i \in I}$ to $\{X_i\}_{i \in I}$, we have

$$\Pr\left[\bigcap_{i \in I} (X_i = s_i)\right] = \prod_{i \in I} \Pr[X_i = s_i].$$

For $k \leq |I|$, we say that $\{X_i\}_{i \in I}$ is **$k$-wise independent** if $\{X_j\}_{j \in J}$ is mutually independent for every subset $J \subseteq I$ of size $k$

We say $\{X_i\}_{i \in I}$ is **pairwise independent** if it is 2-wise independent.
Example: \textit{sum mod m.}

Suppose $X$ and $Y$ are independent random variables, with each uniformly distributed over $\mathbb{Z}_m$

Set $Z := X + Y$

We saw that $Z$ is uniformly distributed over $\mathbb{Z}_m$ and that $X$ and $Z$ are independent

Same argument shows $Y$ and $Z$ are independent

It follows that $X, Y, Z$ are pairwise independent

However, they are not mutually independent:

\[
\Pr[(X = 0) \cap (Y = 0) \cap (Z = 1)] = 0 \neq 1/m^3
\]
**Fact:** If \( \{X_i\}_{i \in I} \) is \( k \)-wise independent, then it is also \( \ell \)-wise independent for any \( \ell < k \)

**Fact:** Let \( \{X_i\}_{i=1}^n \) be a family of random variables, where each \( X_i \) takes values in a finite set \( S_i \)

Then the following are equivalent:

(i) \( (X_1, \ldots, X_n) \) is uniformly distributed over \( S_1 \times \cdots \times S_n \)

(ii) \( \{X_i\}_{i=1}^n \) is mutually independent and each \( X_i \) is uniformly distributed over \( S_i \)

**Fact:** Suppose \( \{X_i\}_{i=1}^n \) is a mutually independent family of random variables

Further, suppose that for \( i = 1, \ldots, n \), we have \( Y_i = g_i(X_i) \) for some function \( g_i \)

Then \( \{Y_i\}_{i=1}^n \) is mutually independent
Example: \( k \)-wise independence from polynomial evaluation.

Let \( p \) be a prime

Choose a random polynomial \( G \in \mathbb{Z}_p[\mathbf{x}] \) of degree less than \( k \)

For each \( \gamma \in \mathbb{Z}_p \), \( G(\gamma) \) is the value of \( G \) at \( \gamma \)

Claim: \( \{G(\gamma)\}_{\gamma \in \mathbb{Z}_p} \) is a \( k \)-wise independent family of random variables, with each \( G(\gamma) \) uniformly distributed over \( \mathbb{Z}_p \)

This follows from Lagrange interpolation:

Let \( \gamma_1, \ldots, \gamma_k \in \mathbb{Z}_p \) be fixed, distinct evaluation points

Lagrange interpolation says the map

\[
(a_0, \ldots, a_{k-1}) \mapsto (g(\gamma_1), \ldots, g(\gamma_k)), \text{ where } g := \sum_j a_j \mathbf{x}^j \in \mathbb{Z}_p[\mathbf{x}]
\]

is bijective

Therefore, a random coefficient vector maps to a random evaluation vector

Note: \( \{G(\gamma)\}_{\gamma \in \mathbb{Z}_p} \) is not \((k + 1)\)-wise independent

Again, Lagrange interpolation: the values of \( G \) at \( k \) distinct evaluation points completely determine \( G \), and hence the value of \( G \) at any other evaluation point.
Example (cont’d): Threshold secret sharing.

Alice has a secret $\sigma \in \mathbb{Z}_p$

She computes a random polynomial $G \in \mathbb{Z}_p[x]$ of degree less than $k$

She sets $H := G + \sigma x^k \in \mathbb{Z}_p[x]$

She computes “secret shares” $S_i = H(\gamma_i)$ for $i = 1, \ldots, n$, where $\gamma_1, \ldots, \gamma_n \in \mathbb{Z}_p$ are distinct, fixed evaluation points

Fact: the $S_i$’s are $k$-wise independent, and each $S_i$ is uniformly distributed over $\mathbb{Z}_p$, but any $k + 1$ shares determine $H$ (and hence $\sigma$)

Alice backs up her secret by storing the $S_i$’s “in the cloud” on $n$ different servers

Any coalition of $k$ or fewer servers learn nothing about her secret

Alice can reconstruct her secret from any $k + 1$ shares

Other applications: nuclear launch codes (used by Russia in the 1990’s)
Example: Binomial distribution.

Suppose we perform \( n \) independent experiments, where each experiment succeeds with probability \( p \) and fails with probability \( q := 1 - p \)

Let \( X_i = 1 \) if \( i \)th experiment succeeds, and 0 otherwise

The family \( \{X_i\}_{i=1}^n \) is mutually independent

Define \( X := \sum_{i=1}^n X_i \)

For \( k = 0 \ldots n \), we have

\[
\Pr[X = k] = \binom{n}{k} p^k q^{n-k}
\]

This is called the \textbf{binomial distribution}, and is parameterized by \( p \) and \( n \)
Example: *Geometric distribution.*

Suppose we repeatedly perform independent experiments, where each experiment succeeds with probability $p$ and fails with probability $q := 1 - p$

Let $X$ be the number of experiments we perform until one succeeds

For $k = 1, 2, \ldots$

$$\Pr[X = k] = q^{k-1}p$$

This is called the **geometric distribution**, and is parameterized by $p$
Expectation

If $X$ is a real-valued random variable:

$$E[X] := \sum_{\omega \in \Omega} X(\omega) \cdot Pr(\omega)$$

If $X$ has image $S$:

$$E[X] = \sum_{s \in S} s \cdot Pr[X = s]$$

More generally, if $X$ takes values in $S$ and $f : S \to \mathbb{R}$:

$$E[f(X)] = \sum_{s \in S} f(s) \cdot Pr[X = s]$$

*Note: $E[X]$ well-defined even for infinite $\Omega$, assuming absolute convergence*
**Linearity of expectation**

**Theorem:** if $X$ and $Y$ are real-valued random variables and $a \in \mathbb{R}$, then

$$E[X + Y] = E[X] + E[Y] \quad \text{and} \quad E[aX] = a E[X]$$

More generally, if $\{X_i\}_{i \in I}$ is a family of real-valued random variables:

$$E\left[ \sum_{i \in I} X_i \right] = \sum_{i \in I} E[X_i]$$

*Note: holds even for infinite families, assuming each $X_i \geq 0$ and $\sum_i X_i(\omega)$ converges for each $\omega \in \Omega$*
Example: uniform distribution.

$X$ is uniformly distributed over $\{1, \ldots, n\}$:

$$E[X] = \sum_{i=1}^{n} \frac{i}{n} = \frac{1}{n} \sum_{i=1}^{n} i = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2}$$

Example: Bernoulli distribution.

$X = 1$ with probability $p$, $X = 0$ with probability $q := 1 - p$:

$$E[X] = 1 \cdot p + 0 \cdot q = p$$

Example: Indicator variable.

$X_A = 1$ with probability $\Pr[A]$, $X_A = 0$ with probability $1 - \Pr[A]$:

$$E[X_A] = \Pr[A]$$
**Example:** Binomial distribution.

Recall: $X = \sum_{i=1}^{n} X_i$

For $k = 0 \ldots n$, we have

$$\Pr[X = k] = \binom{n}{k} p^k q^{n-k}$$

So, $E[X] = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} \ldots !!$?

**Linearity!!**

$$E[X] = \sum_{i=1}^{n} E[X_i] = np$$
The tail sum formula

**Theorem:** If $X$ is a random variable that takes non-negative integer values, then

$$E[X] = \sum_{i \geq 1} \Pr[X \geq i]$$

**Proof by picture.** Let $p_i = \Pr[X = i]$:

$$
\begin{array}{c}
p_1 \\
p_2 & p_2 \\
p_3 & p_3 & p_3 \\
\vdots & \vdots & \vdots & \ddots & \\
\end{array}
$$

$i$th row sums to $i \Pr[X = i]$

$i$th column sums to $\Pr[X \geq i]$
Example: Geometric distribution.

For \( k = 1, 2, \ldots \)

\[
\Pr[X = k] = q^{k-1} p
\]

Compute: \( E[X] = \sum_{k \geq 1} kq^{k-1} p \ldots \) !#$%&^@

Use the tail sum formula — observe

\[
\Pr[X \geq i] = q^{i-1}
\]

Therefore,

\[
E[X] = \sum_{i \geq 1} \Pr[X \geq i] = \sum_{i \geq 1} q^{i-1} = \frac{1}{1-q} = \frac{1}{p}
\]
**Example:** expected minimum.

We roll four dice. For \( i = 1, \ldots, 4 \), let \( X_i \) be the value of the \( i \)th die.

So \( X_1, \ldots, X_4 \) is a mutually independent family of random variables, where each \( X_i \) is uniformly distributed over \( \{1, \ldots, 6\} \).

Let \( M := \min(X_1, \ldots, X_4) \).

Tail sum formula:

\[
E[M] = \sum_{j=1}^{6} \Pr[M \geq j].
\]

\( M \geq j \) occurs \( \iff \) \( X_i \geq j \) for all \( i = 1, \ldots, 4 \).

By independence, we have

\[
\Pr[M \geq j] = \Pr[X_1 \geq j] \cdots \Pr[X_4 \geq j] = \left( \frac{7-j}{6} \right)^4
\]

So we have

\[
E[M] = \sum_{j=1}^{6} \Pr[M \geq j] = \frac{6^4 + 5^4 + 4^4 + 3^4 + 2^4 + 1^4}{6^4} \approx 1.75
\]
Conditional expectation

Let $B$ be an event with $\Pr[B] \neq 0$

Let $X$ be a real-valued random variable

We can calculate the expectation of $X$ with respect to the conditional distribution given $B$:

$$E[X \mid B] = \sum_{\omega \in \Omega} X(\omega) \Pr(\omega \mid B)$$

**Law of total expectation:** if $\{B_i\}_{i \in I}$ be a partition of $\Omega$, then

$$E[X] = \sum_{i \in I} E[X \mid B_i] \Pr[B_i]$$
Example: We roll a die
Let $X$ denote the value of the die
Let $\mathcal{A}$ be the event that the value is even
The distribution of $X$ given $\mathcal{A}$ is the uniform distribution on $\{2, 4, 6\}$, so
$$E[X | \mathcal{A}] = \frac{2 + 4 + 6}{3} = 4$$

The distribution of $X$ given $\overline{\mathcal{A}}$ is the uniform distribution on $\{1, 3, 5\}$, so
$$E[X | \overline{\mathcal{A}}] = \frac{1 + 3 + 5}{3} = 3$$

So we have
$$E[X] = E[X | \mathcal{A}] \Pr[\mathcal{A}] + E[X | \overline{\mathcal{A}}] \Pr[\overline{\mathcal{A}}]
= 4 \cdot \frac{1}{2} + 3 \cdot \frac{1}{2} = \frac{7}{2}$$
Expectation of products

**Theorem:** If $X$ and $Y$ are independent real-valued random variables, then

$$
E[X \cdot Y] = E[X] \cdot E[Y]
$$

**Example:** Let $X_1$ and $X_2$ be independent random variables, each uniformly distributed over $\{0, 1\}$. Set $X := X_1 + X_2$

$$
E[X] = E[X_1] + E[X_2] = 1/2 + 1/2 = 1
$$

$$
E[X^2] = E[(X_1 + X_2)(X_1 + X_2)] = E[X_1^2] + 2 E[X_1] E[X_2] + E[X_2^2]
$$

$$
= 1/2 + 2 \cdot (1/4) + 1/2 = 3/2
$$

Observe: $3/2 = E[X^2] > E[X]^2 = 1$
Some basic inequalities

**Jensen’s inequality (special case):** If $X$ is a real-valued random variable, then

$$E[X^2] \geq E[X]^2$$

**Markov’s inequality:** If $X$ takes only non-negative real values, then for every $\alpha > 0$, we have

$$\Pr[X \geq \alpha] \leq \frac{E[X]}{\alpha}$$

Setting $\mu := E[X]$ and plugging in $\alpha := \beta \mu$, we obtain

$$\Pr[X \geq \beta \mu] \leq \frac{1}{\beta}$$