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 $\mathcal{A} \subseteq \Omega$

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

Logical operations:

• $\mathcal{A} \cap \mathcal{B}$ — logical AND
• $\mathcal{A} \cup \mathcal{B}$ — logical OR
• $\Omega \setminus \mathcal{A}$ — logical NOT

Union bounds:

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

$$\Pr\left[\bigcup_{i \in I} \mathcal{A}_i\right] \leq \sum_{i \in I} \Pr[\mathcal{A}_i]$$

and equality holds if the $\mathcal{A}_i$’s are pairwise disjoint
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** \( B \)

For any event \( A \):

\[
\Pr[A \mid B] = \sum_{\omega \in A} \Pr(\omega \mid B) = \frac{\Pr[A \cap B]}{\Pr[B]}.
\]
A and B are called independent if

- \( \Pr[A \cap B] = \Pr[A] \cdot \Pr[B] \),
- or equivalently, \( \Pr[A] = \Pr[A | B] \)

**Example:** roll a die and define events:

- \( A \): die odd,  \( B \): die > 2,  \( C \): die > 3

\[
\Pr[A] = \frac{1}{2}, \quad \Pr[B] = \frac{2}{3}, \quad \Pr[A \cap B] = \frac{1}{3}
\]

\[\Rightarrow A \text{ and } B \text{ independent}\]

also:

\[
\Pr[A | B] = \frac{\Pr[A \cap B]}{\Pr[B]} = \frac{1/3}{2/3} = \frac{1}{2} = \Pr[A]
\]

\[
\Pr[C] = \frac{1}{2}, \quad \Pr[A \cap C] = \frac{1}{6} \neq \frac{1}{4}
\]

\[\Rightarrow A \text{ and } C \text{ not independent}\]
Total probability and Bayes’ theorem

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

Let \( A \) be any event

**Law of total probability:**

\[
\Pr[A] = \sum_{i \in I} \Pr[A \cap B_i] = \sum_{i \in I} \Pr[A \mid B_i] \Pr[B_i]
\]

**Bayes’ theorem:** for \( j \in I \), we have

\[
\Pr[B_j \mid A] = \frac{\Pr[B_j \cap A]}{\Pr[A]} = \frac{\Pr[A \mid B_j] \Pr[B_j]}{\sum_{i \in I} \Pr[A \mid B_i] \Pr[B_i]}
\]
Example: *a rare disease.*

- Rate of incidence of disease $X$ is 1%
- Imperfect test: 5% false positive rate, 0% false negative rate
- A random person walks in to a doctor and takes the test
- The test comes out positive
- What is the probability they actually have $X$?
Define events:

\( P \): test positive, \( D \): has disease X

We know:

\[
\Pr[D] = 0.01, \quad \Pr[P | \overline{D}] = 0.05, \quad \Pr[P | D] = 1
\]

We want to calculate: \( \Pr[D | P] \)

Bayes says:

\[
\Pr[D | P] = \frac{\Pr[P | D] \Pr[D]}{\Pr[P | D] \Pr[D] + \Pr[P | \overline{D}] \Pr[\overline{D}]}
\]

\[
= \frac{1 \cdot 0.01}{1 \cdot 0.01 + 0.05 \cdot 0.99} \approx 0.17
\]
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

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

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**:

$$X_A(\omega) := \begin{cases} 
1 & \text{if } \omega \in \mathcal{A}, \\
0 & \text{otherwise}
\end{cases}$$
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 $t \in T$

Let $\{X_i\}_{i \in I}$ be a finite family of random variables, where each $X_i$ takes values in $S_i$

$\{X_i\}_{i \in I}$ is **pairwise independent** if $X_i$ and $X_j$ are independent for all $i \neq j$

$\{X_i\}_{i \in I}$ is **mutually independent** if

$$\Pr \left[ \bigcap_{j \in J} (X_j = s_j) \right] = \prod_{j \in J} \Pr[X_j = s_j]$$

for all $J \subseteq I$ and all possible assignments $\{s_j \in S_j\}_{j \in J}$
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 $k = 0..n$, we have

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

This is called the **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 $p$. 
Example: encryption with a one-time pad

Alice wants to send a message $M \in \{0, 1\}^t$ to Bob over an insecure network.

An adversary, Eve, may be eavesdropping on their communication.

Assume Alice and Bob have a shared secret key $K \in \{0, 1\}^t$.

Alice encrypts $M$ as $C \leftarrow M \oplus K$, and sends the "ciphertext" $C$ over the network.

Upon receiving $C$, Bob decrypts $C$ by computing $M \leftarrow C \oplus K$.

Observe: $C \oplus K = (M \oplus K) \oplus K = M \oplus (K \oplus K) = M \oplus 0^t = M$. 
How secure is this scheme? **Perfectly secure. Why?**

Model $K$ as a random variable uniformly distributed over \{0, 1\}^t

Model $M$ as a random variable with some distribution over \{0, 1\}^t

**Assumption:** $M$ and $K$ are independent

**Claim:** $M$ and $C$ are independent

**Interpretation of claim:**

- for all $m, c \in \{0, 1\}^t$: $\Pr[M = m | C = c] = \Pr[M = m]$
- this means: after seeing the ciphertext, Eve knows nothing more about the message than she did before seeing the ciphertext
- this means: the ciphertext leaks no information about the message
Proof that $M$ and $C$ are independent

Let $m, c \in \{0, 1\}^t$ be fixed. We have

$$\Pr[(M = m) \cap (C = c)]$$

$$= \Pr[(M = m) \cap (M \oplus K = c)] = \Pr[(M = m) \cap (m \oplus K = c)]$$

$$= \Pr[(M = m) \cap (K = m \oplus c)]$$

$$= \Pr[M = m] \cdot \Pr[K = m \oplus c] \text{ (independence of } M \text{ and } K)$$

$$= \Pr[M = m] \cdot 2^{-t} \text{ (distribution of } K)$$

We also have

$$\Pr[C = c] = \sum_m \Pr[(M = m) \cap (C = c)] \text{ (total probability)}$$

$$= \sum_m \Pr[M = m] \cdot 2^{-t} = 2^{-t} \sum_m \Pr[M = m]$$

$$= 2^{-t} \text{ (probabilities sum to 1)}$$

That proves

$$\Pr[(M = m) \cap (C = c)] = \Pr[M = m] \cdot \Pr[C = c]$$
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 $\{1, \ldots, n\}$:

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

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 $p_i = \Pr[X = i]$:

\[
\begin{array}{ccc}
  p_1 & & \\
  p_2 & p_2 & \\
  p_3 & p_3 & p_3 \\
  \vdots & \vdots & \vdots \\
\end{array}
\]

$i$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: Coupon collector’s problem

There are $n$ distinct coupons. You are trying to collect one of each. Each day you receive a random coupon in the mail.

What is the expected number of days until you collect one of each?
Solution: For $i = 1 \ldots n$, suppose you already have $i - 1$ distinct coupons, and let $X_i$ be the number of days until you receive a new one.

$X := \sum_{i=1}^{n} X_i$ is the total number of days.

$X_i$ has a geometric distribution with success probability $(n - i + 1)/n$:

$$E[X] = \sum_{i=1}^{n} E[X_i] = n \sum_{i=1}^{n} \frac{1}{i}$$

$$= n(\ln n + O(1)) = n \ln n + O(n)$$
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

$$\Pr[X \geq \alpha] \leq E[X]/\alpha$$

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

$$\Pr[X \geq \beta \mu] \leq 1/\beta$$