Let’s start with a single random variable…

\[ X = \begin{cases} 
0 \\
1 
\end{cases} \]
Random Experiment

A random experiment has three elements

1. sample space $\Omega$: set of all possible outcomes
   e.g., $\Omega=\{1,2,3,4,5,6\}$

2. event space $E$: collection of subsets of $\Omega$
   e.g., $E=\{\{1\},\ldots,\{1,2\},\ldots,\{3,4,6\},\ldots,\{1,2,3,4,5,6\}\}=2^\Omega$

3. probability measure $P$: a special function $P: E \to \mathbb{R}$
   e.g., $P(\{1\})=1/6$, $P(\{3,4,6\})=3/6$
Axioms of Probability

Probability measure $P$ needs to satisfy

1. $P(e) \geq 0$, for any event $e \in E$
   
   e.g., $P(\{3,4,6\}) = \frac{3}{6} \geq 0$

2. $P(\Omega) = 1$
   
   e.g., $P(\{1,2,3,4,5,6\}) = 1$

3. If $e_1, e_2 \in E$ are disjoint, then $P(e_1 \cup e_2) = P(e_1) + P(e_2)$
   
   e.g., $P(\{1\} \cup \{2,4,6\}) = P(\{1\}) + P(\{2,4,6\})$
<table>
<thead>
<tr>
<th>Tomorrow’s Temperature</th>
<th>Tomorrow’s Weather</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Space $\Omega = ?$</td>
<td>Sample Space $\Omega = ?$</td>
</tr>
<tr>
<td>Event Space $E = ?$</td>
<td>Event Space $E = ?$</td>
</tr>
<tr>
<td>Probability Measure $P$</td>
<td>Probability Measure $P$</td>
</tr>
</tbody>
</table>

**Exercise**
Exercise

A random experiment is flipping a coin \textit{twice} (getting H or T in each trial)

1. sample space $\Omega = ?$

2. event space $E = ?$

3. probability measure $P \ ?$

A random experiment can be a single trial or a sequence of trails.
A random variable $X$ is a function $X : \Omega \rightarrow \mathbb{R}$.

Q: are the following random variables?

1. $X(\{1\}) = 1, \ X(\{3\}) = 3, \ldots$
2. $X(\{1\}) = 19, \ X(\{3\}) = 45.7, \ldots$
3. $X(\{1\}) = 1, \ X(\{3\}) = 1, \ldots$
A random variable $X$ is a function $X : E \rightarrow \mathbb{R}$.

Q: are the following random variables?

1. $X(\{1\}) = 1, \ X(\{3\}) = 3, \ldots$
2. $X(\{1\}) = 19, \ X(\{3\}) = 45.7, \ldots$
3. $X(\{1\}) = 1, \ X(\{3\}) = 1, \ldots$

Q: why bother?
Random Variable

Random variable makes presentation easier!

Example 1
If $X(\{1\})=19$ and $X(\{3\})=45.7$, instead of writing $P(\{1\})$, $P(\{3\})$, we can write $P(X=19)$, $P(X=45.7)$.

Example 2
If $X$ counts odds over three trials, e.g., $X(\{1, 2, 6\})=1$, $X(\{1, 2, 3\})=2$, instead of writing $P(1 \text{ odd})$ or $P(2 \text{ odds})$, we can write $P(X=1)$ or $P(X=2)$. 
Discrete Variable vs Continuous Variable

X is discrete if it maps E to a discrete domain (our focus)
- e.g., \( X(\{3\})=1, \ X(\{1,2,3\})=2 \)
- specify \( P(X=1), P(X=2), \) etc

X is continuous if it maps E to a continuous domain
- e.g., \( X(\text{angle}) = 19.32 \)
- specify \( P(0 \leq X \leq 30), \) etc

For continuous random variable \( X, \) \( P(X=a)=0. \)
Properties of Probability

1. if $e_1 \subseteq e_2$, then $P(e_1) \leq P(e_2)$
   
e.g., since $\{2\} \subseteq \{2,4,6\}$, we have $P(\{2\}) \leq P(\{2,4,6\})$

2. $P(e_1 \cap e_2) \leq \min\{P(e_1), P(e_2)\}$
   
e.g., $P(\{2,3\} \cap \{2,4,6\}) \leq \min\{P(\{2,3\}), P(\{2,4,6\})\}$

3. $P(e_1 \cup e_2) \leq P(e_1) + P(e_2)$ [union bound]
   
e.g., $P(\{2,3\} \cup \{2,4,6\}) \leq P(\{2,3\}) + P(\{2,4,6\})$
Properties of Probability

4. \( P(\Omega \setminus e) = 1 - P(e) \)
   e.g., \( P(\{1,2,3,4,5,6\}\setminus\{1,2,3\}) = 1 - P(\{1,2,3\}) \)

5. if \( e_1, e_2 \) are disjoint, \( e_1 \cup e_2 = \Omega \), then \( P(e_1) + P(e_2) = 1 \)
   [law of total probability]
   e.g., \( P(\{1,2,3\}) + P(\{4,5,6\}) = 1 \)
What is the distribution of a random variable?
Cumulative Distribution Function (CDF)

CDF of $X$ is a function $CDF_X : \mathbb{R} \rightarrow [0,1]$ specifying a special probability

$$CDF_X(a) = P(X \leq a)$$

Example

If $X$ counts odds, $P(X \leq 3)$ is the chance of getting three trials with no more than 3 odds,

i.e., $P(X \leq 3) = P(\{e:X(e)\leq3\})$

If $X$ is temperature, $P(X \leq 27)$ is the chance of getting tomorrow’s temperature of no higher than 27F

i.e., $P(X \leq 27) = P(\{e:X(e)\leq27\})$
Properties of CDF

CDF of $X$ is a function $CDF_X : \mathbb{R} \rightarrow [0,1]$ specifying a special probability

$$CDF_X(a) = P(X \leq a)$$

1. $0 \leq CDF(a) \leq 1$
2. $\lim_{a \rightarrow -\infty} CDF(a) = 0$
3. $\lim_{a \rightarrow +\infty} CDF(a) = 1$
4. if $a \leq b$, then $CDF(a) \leq CDF(b)$
Probability Density Function (PDF)

PDF of a continuous random variable $X$ is the derivative of its CDF

$$p(x) = \frac{\partial \text{CDF}(x)}{\partial x}$$
Example: Gaussian Distribution

If $X$ follows a Gaussian distribution (or, normal distribution), then

$$
\frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{x - \mu}{\sigma \sqrt{2}} \right) \right]
$$

$$
\frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
$$

$$
\operatorname{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{-t^2} \, dt
$$
Different PDFs

- Normal
- Triangular
- Poisson
- Binomial
- Lognormal
- Uniform
- Exponential
- Geometric
- Weibull
- Beta
- Hypergeometric
- Custom
- Gamma
- Logistic
- Pareto
- Extreme Value
Properties of PDF

PDF of a continuous random variable $X$ is the derivative of its CDF

$$p(x) = \frac{\partial \text{CDF}(x)}{\partial x}$$

1. $p(x) \geq 0$

2. $\int_{x \in (-\infty, +\infty)} p(x) \, dx = 1$

3. $\int_{x \in S} p(x) \, dx = P(X \in S)$
Probability Mass Function (PMF)

PMF of a discrete random variable X is a table of probabilities for all events.

\[
P(T) \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad P(W)
\]

\[
\begin{array}{|c|c|}
\hline
T & P \\
\hline
\text{hot} & 0.5 \\
\hline
\text{cold} & 0.5 \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|}
\hline
W & P \\
\hline
\text{sun} & 0.6 \\
\hline
\text{rain} & 0.1 \\
\hline
\text{fog} & 0.3 \\
\hline
\text{meteor} & 0.0 \\
\hline
\end{array}
\]
Different PMFs
Properties of PMF

PMF of a discrete random variable X is a table of probabilities for all events.

1. $1 \geq p(x) \geq 0$

2. $\Sigma_{x \in (-\infty, +\infty)} p(x) = 1$

3. $\Sigma_{x \in S} p(x) = P(X \in S)$
Based on distribution, we can define expectation and variance of a variable...
Expectation of a random variable is its “weighted average” value.

\[ E[X] = \sum_{x \in D} x \cdot p(x) \quad \text{or} \quad E[X] = \int_{x \in D} x \cdot p(x) \, dx \]

<table>
<thead>
<tr>
<th>X</th>
<th>p(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
</tr>
</tbody>
</table>

\[
E[X] = \sum_{x \in D} x \cdot p(x) \\
= 1 \cdot 0.5 + 2 \cdot 0.25 + 3 \cdot 0.25 \\
= 1.75
\]
Expectation of Standard Normal Distribution

Expectation of a normally distributed random variable $\sim N(0, \sigma^2)$ is zero.

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

$$= (2\pi)^{-1/2} \int_{-\infty}^{\infty} x \exp\left(-\frac{1}{2}x^2\right)dx$$

$$= (2\pi)^{-1/2} \int_{-\infty}^{0} x \exp\left(-\frac{1}{2}x^2\right)dx + (2\pi)^{-1/2} \int_{0}^{\infty} x \exp\left(-\frac{1}{2}x^2\right)dx$$

$$= (2\pi)^{-1/2} \left[ -\exp\left(-\frac{1}{2}x^2\right) \right]_0^{\infty} + (2\pi)^{-1/2} \left[ -\exp\left(-\frac{1}{2}x^2\right) \right]_0^{\infty}$$

$$= (2\pi)^{-1/2} [-1 + 0] + (2\pi)^{-1/2} [0 + 1]$$

$$= (2\pi)^{-1/2} - (2\pi)^{-1/2}$$

$$= 0$$

One can similarly prove expectation of variable from $N(m, \sigma^2)$ is $m$. 
Properties of Expectation

Expectation of a random variable is its “weighted average” value.

\[ E[X] = \sum_{x \in D} x \cdot p(x) \quad \text{or} \quad E[X] = \int_{x \in D} x \cdot p(x) \, dx \]

1. \( E[a] = a \), for any constant \( a \)

2. \( E[a \cdot X] = a \cdot E[X] \)

3. \( E[X+Y] = E[X] + E[Y] \)

4. \( E[1_{X=k}] = P(X=k) \)
Variance

Variance of a random variable is its average deviation from its expectation.

\[ \text{Var}[X] = E \left[ (X-E[X])^2 \right] \]

<table>
<thead>
<tr>
<th>X</th>
<th>p(x)</th>
<th>E[X]</th>
<th>Var[X]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5</td>
<td>(1 \times 0.5 + 2 \times 0.25 + 3 \times 0.25)</td>
<td>(0.6875)</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>(1.75)</td>
<td>((1-1.75)^2 \times 0.5 + (2-1.75)^2 \times 0.25 + (3-1.75)^2 \times 0.25)</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
<td>(1.75)</td>
<td>(0.6875)</td>
</tr>
</tbody>
</table>

Note expectation \(E[X]\) is a constant.
Variance of Normally Distributed Variable

Figure 6: Normal Distribution Curves for Mean of 0 and Variances of 25, 100, and 400
Variance of Other Distribution

Poisson Distribution

Rayleigh Distribution
Properties of Variance

Variance of a random variable is its average deviation from its expectation.

\[ \text{Var}[X] = E \left[ (X - E[X])^2 \right] \]

1. \( \text{Var}[X] = E[X^2] - E[X]^2 \)

2. \( \text{Var}[a] = 0 \) for constant \( a \)

3. \( \text{Var}[aX] = a^2 \text{Var}[X] \)

\[
= E[X^2] - 2E[X]E[X] + E[X]^2 \\
= E[X^2] - E[X]^2,
\]
## Common Distributions

<table>
<thead>
<tr>
<th>Distribution</th>
<th>PDF or PMF</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
</table>
| Bernoulli($p$)        | \[
\begin{cases}
    p, & \text{if } x = 1 \\
    1 - p, & \text{if } x = 0.
\end{cases}
\] | $p$     | $p(1 - p)$ |
| Binomial($n$, $p$)    | \[
\binom{n}{k} p^k (1 - p)^{n-k} \quad \text{for } 0 \leq k \leq n
\] | $np$    | $npq$      |
| Geometric($p$)        | $p(1 - p)^{k-1}$ for $k = 1, 2, \ldots$                                  | $\frac{1}{p}$ | $\frac{1-p}{p^2}$ |
| Poisson($\lambda$)    | $e^{-\lambda} \frac{\lambda^x}{x!} \quad \text{for } k = 1, 2, \ldots$ | $\lambda$ | $\lambda$   |
| Uniform($a$, $b$)     | $\frac{1}{b-a}$ \quad \forall x \in (a, b)                              | $\frac{a+b}{2}$ | $\frac{(b-a)^2}{12}$ |
| Gaussian($\mu$, $\sigma^2$) | $\frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ | $\mu$   | $\sigma^2$  |
| Exponential($\lambda$)| $\lambda e^{-\lambda x} \quad x \geq 0, \lambda > 0$                    | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ |
Now we look at two (discrete) random variables...

The following slides are largely built on the lecture slides of “probability” by Klein and Abbeel.
Recap: Distribution of Single Variable

Q: what is the probability that tomorrow is hot and sun, i.e. \( P(T=\text{hot}, W=\text{sun}) \)?
Joint Distribution of Two Variables

Joint distribution is a table of probabilities of all joint events.

\[ P(T, W) \]

<table>
<thead>
<tr>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>sun</td>
<td>0.4</td>
</tr>
<tr>
<td>hot</td>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>cold</td>
<td>sun</td>
<td>0.2</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>
Marginal Distribution

Sometimes we are interested in the distribution of one variable. We can obtain it from the joint distribution by marginalizing it.

\[
P(T, W)
\]

\[
\begin{array}{ccc}
T & W & P \\
\text{hot} & \text{sun} & 0.4 \\
\text{hot} & \text{rain} & 0.1 \\
\text{cold} & \text{sun} & 0.2 \\
\text{cold} & \text{rain} & 0.3 \\
\end{array}
\]

\[
P(T) = \sum_{s} P(t, s)
\]

\[
P(W) = \sum_{t} P(t, s)
\]

\[
\begin{array}{cc}
P(T) & \\
\text{T} & P \\
\text{hot} & 0.5 \\
\text{cold} & 0.5 \\
\end{array}
\]

\[
\begin{array}{cc}
P(W) & \\
\text{W} & P \\
\text{sun} & 0.6 \\
\text{rain} & 0.4 \\
\end{array}
\]
Exercise

\[ P(X, Y) \]

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>+x</td>
<td>+y</td>
<td>0.2</td>
</tr>
<tr>
<td>+x</td>
<td>-y</td>
<td>0.3</td>
</tr>
<tr>
<td>-x</td>
<td>+y</td>
<td>0.4</td>
</tr>
<tr>
<td>-x</td>
<td>-y</td>
<td>0.1</td>
</tr>
</tbody>
</table>

\[ P(x) = \sum_y P(x, y) \]

\[ P(y) = \sum_x P(x, y) \]
Conditional Distribution

Sometimes we are interested in the distribution of one variable $X$ while fixing the other $Y$. This is conditional distribution and defined as

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

<table>
<thead>
<tr>
<th>$P(T, W)$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>W</td>
<td>P</td>
</tr>
<tr>
<td>hot</td>
<td>sun</td>
<td>0.4</td>
</tr>
<tr>
<td>hot</td>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>cold</td>
<td>sun</td>
<td>0.2</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>

$P(W = s|T = c) = \frac{P(W = s, T = c)}{P(T = c)} = \frac{0.2}{0.5} = 0.4$
Sometimes we are interested in the distribution of one variable $X$ while fixing the other $Y$. This is conditional distribution and defined as

$$P(X|Y) = \frac{P(X,Y)}{P(Y)}.$$
Exercise

\[ P(X, Y) \]

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>+x</td>
<td>+y</td>
<td>0.2</td>
</tr>
<tr>
<td>+x</td>
<td>-y</td>
<td>0.3</td>
</tr>
<tr>
<td>-x</td>
<td>+y</td>
<td>0.4</td>
</tr>
<tr>
<td>-x</td>
<td>-y</td>
<td>0.1</td>
</tr>
</tbody>
</table>

- \( P(+x \mid +y) \) ?
- \( P(-x \mid +y) \) ?
- \( P(-y \mid +x) \) ?
Remark

Variable X can have different distributions when conditioned on different Y.

\[
P(W | T = \text{hot})
\]

\[
\begin{array}{|c|c|}
\hline
W & P \\
\hline
\text{sun} & 0.8 \\
\text{rain} & 0.2 \\
\hline
\end{array}
\]

\[
P(W | T = \text{cold})
\]

\[
\begin{array}{|c|c|}
\hline
W & P \\
\hline
\text{sun} & 0.4 \\
\text{rain} & 0.6 \\
\hline
\end{array}
\]

\[
P(T, W)
\]

\[
\begin{array}{|c|c|c|}
\hline
T & W & P \\
\hline
\text{hot} & \text{sun} & 0.4 \\
\text{hot} & \text{rain} & 0.1 \\
\text{cold} & \text{sun} & 0.2 \\
\text{cold} & \text{rain} & 0.3 \\
\hline
\end{array}
\]
Product Rule

Infer joint probability from marginal probability and conditional probability.

\[ P(y)P(x|y) = P(x, y) \quad \iff \quad P(x|y) = \frac{P(x, y)}{P(y)} \]
Exercise

\[ P(y)P(x|y) = P(x, y) \]

- Example:

| \( P(W) \) | \( P(D|W) \) | \( P(D, W) \) |
|---|---|---|
| R | P |  |
| sun | 0.8 |  |
| rain | 0.2 |  |

- | D | W | P |
- | wet | sun | 0.1 |
- | dry | sun | 0.9 |
- | wet | rain | 0.7 |
- | dry | rain | 0.3 |
- | wet | sun |  |
- | dry | sun |  |
- | wet | rain |  |
- | dry | rain |  |
Chain Rule

Split a joint probability of multiple variables into conditional probabilities of each.

\[ P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2) \]
Chain Rule

Split a joint probability of multiple variables into conditional probabilities of each.

\[ P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2) \]

Q: how to derive it?

\[ P(y)P(x|y) = P(x, y) \]
Bayes’s Rule

Bayes’s rule connects two conditional probabilities.

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

- \(P(A)\) is prior of \(A\), \(P(A\mid B)\) is posterior of \(A\)

- \(p(A=a\mid B=b) = \frac{p(B=b\mid A=a) \cdot p(A=a)}{p(B=b)} \) [discrete]

- \(p(a\mid b) = \frac{p(b\mid a) \cdot p(a)}{p(b)} \) [continuous]

Q: how to derive Bayes’s Rule? Why Bayes’s rule?
Example

We randomly pick up a coin and flip it to get head (H) or tail (T).

We know a priori that

- 40% coins are fair (F), and 60% are unfair (U)
- 80% chance we get a head (from an arbitrary coin)

If we get head, what is the chance the flipped coin is fair? 

\[ P(F|H) \]
Exercise

- Given:

\[ P(W) \]

<table>
<thead>
<tr>
<th>R</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.8</td>
</tr>
<tr>
<td>rain</td>
<td>0.2</td>
</tr>
</tbody>
</table>

\[ P(D|W) \]

<table>
<thead>
<tr>
<th>D</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>wet</td>
<td>sun</td>
<td>0.1</td>
</tr>
<tr>
<td>dry</td>
<td>sun</td>
<td>0.9</td>
</tr>
<tr>
<td>wet</td>
<td>rain</td>
<td>0.7</td>
</tr>
<tr>
<td>dry</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>

- What is \( P(W \mid \text{dry}) \)?
Independence

Two variables $X$ and $Y$ are independent if

$$P(X, Y) = P(X) P(Y)$$

- or, equivalently $P(X \mid Y) = P(X)$
- $p(X=x, Y=y) = p(X=x) p(Y=y)$ [discrete]
- $p(x,y) = p(x) p(y)$ [continuous]
Expectation

Expectation of two random variables $X,Y$ is

$$E[g(X,Y)] = \sum_{(x,y) \in D} g(x,y) * p(x,y)$$

or

$$E[g(X,Y)] = \int_{(x,y) \in D} g(x,y) * p(x) \, dx \, dy$$
Covariance

Covariance of two variables X, Y is

$$\text{Cov}(X,Y) = E[(X-E[X])(Y - E[Y])]$$

- if $X=Y$, then $\text{Cov}(X,Y) = \text{Var}(X) = \text{Var}(Y)$
- $\text{Cov}(X,Y) = E[XY] - E[X] E[Y]$
- $\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y] + \text{Cov}[X,Y]$
- if $X$, $Y$ are independent, then $\text{Cov}[X,Y] = 0$ (uncorrelated)
Property of Covariance

\[
Cov[X, Y] = E[(X - E[X])(Y - E[Y])]
= E[XY] - E[X]E[Y].
\]