Overview

- Random Variables
- Indicator Random Variable
- Probability Mass Function
- Cumulative Distribution Function
- Joint Probability Mass Function
- Conditional Probability
- Marginal Distributions
- Chain Rule, Bayes and Independence
Random Variables

Often we are interested in numbers associated with the outcome of a random experiment. E.g.

- Out of 2 coin tosses, how many came up heads
- When I throw two dice, what is the sum
- How many MB’s of youtube video I watched today
- How long did my bus journey take this morning
- My ST3009 exam score
Out of 2 coin tosses, how many came up heads. Let’s call this random variable $X$ (usual convention is to use upper case for RVs).

- $Y$ takes values in \{0, 1, 2\}
- Sample space $S = \{(H, H), (H, T), (T, H), (T, T)\}$
- We can associate a value of $Y$ with outcomes of the experiment e.g. $X = 0$ when outcome is $(T, T)$, $X = 1$ when outcome is $(H, T)$ or $(T, H)$, $X = 2$) when outcome is $(H, H)$. 
Random Variables

When I throw two dice, what is the sum.

- $X$ takes values in $\{2, \cdots, 12\}$ (value 1 isn’t possible)
- Sample space $S = \{(1,1), (1,2), \cdots, (6,6)\}$
- We can associate a value of $X$ with outcomes of the experiment e.g. $X = 2$ when outcome is $(1,1)$, $X = 3$ when outcome is $(1,2)$ or $(2,1)$ etc.
Random Variables

A **random variable** is a function that maps from the sample space $S$ to a real number.

- Write $X(\omega)$, where $\omega \in S$ is an outcome.
- But often $\omega$ is dropped and just write $X$. This is just a convenience though.
- When $X$ can take only a finite number of values e.g. $\{1, 2\}$ then it is called a **discrete** random variable.
- Otherwise it's a **continuous** random variable.
Indicator Random Variable: takes value 1 if event $E$ occurs and 0 if event $E$ does not occur.

\[ I = \begin{cases} 
1 & \text{if } E \text{ occurs} \\
0 & \text{if } E \text{ doesn’t occur} 
\end{cases} \]

$I$ is a random variable, a function of outcomes in sample space $S$ that takes values in \{0, 1\}.

- Sometimes we might see e.g. for lunch today random variable $X = Sandwich$.
- $X$ here is not real-valued, so not a random variable.
- But its rough shorthand for the indicator random variable $I$ taking value 1 when I eat a sandwich for lunch today i.e. $X = Sandwich$ is the same as $I = 1$. 
Probability Mass Function

A probability is associated with each value that a discrete random variable can take.

- We write $P(X = x)$ for the probability that random variable $X$ takes value $x$.
- This is often abbreviated to $P(x)$ or $p(x)$, where the random variable $X$ is understood.
- It’s called the probability mass function (PMF) of $X$.
- The set of outcomes for which $X = x$ is $E_x = \{\omega | X(\omega) = x, \omega \in S\}$
- So $P(X = x)$ is the probability that event $E_x$ occurs i.e. $P(X = x) = P(E_x)$.

Example: The number of heads from two coin flips.

- Sample space $S = \{(H, H), (H, T), (T, H), (T, T)\}$
- $P(X = 0) = \frac{1}{4}$ (event \{(T, T)\})
- $P(X = 1) = \frac{1}{2}$ (event \{(H, T), (T, H)\})
- $P(X = 2) = \frac{1}{4}$ (event \{(H, H)\})
- See that $P(X = 0) + P(X = 1) + P(X = 2) = 1$
Probability Mass Function

Suppose discrete random variable $X(\omega)$ takes values in $D = \{x_1, x_2, \cdots, x_n\}$ as $\omega$ ranges over sample space $S$.

- $P(X = x_i) \geq 0$ for all $i = 1, 2, \cdots, n$ since its a probability
- $\sum_{i=1}^{n} P(X = x_i) = 1$ since $P(X = x_i)$ is the probability of event $E_i = \{\omega | X(\omega) = x_i, \omega \in S\}$ and $\bigcup_{i=1}^{n} E_i = S$ so $P(\bigcup_{i=1}^{n} E_i) = 1.$

![Histogram of sum of two dice](chart.png)
Cumulative Distribution Function

- For a random variable $X$ the **cumulative distribution function** (CDF) is defined as: $F(a) = P(X \leq a)$ where $a$ is real-valued.
- For a discrete random variable taking values in $D = \{x_1, x_2, \cdots, x_n\}$, the CDF is $F(a) = P(X \leq a) = \sum_{x_i \leq a} P(X = x_i)$.
- If $a \leq b$ then $F(a) \leq F(y)$

CDF for sum of two dice
Cumulative Distribution Function

Example. Suppose a discrete random variable $X$ takes values in \{0, 1, 2, 3, 4\} and its probability mass function is $P(X = x) = \frac{x}{10}$. What is its CDF?

- For any $x < 1$, $F(x) = \sum_{x_i \leq 0} P(X = x_i) = P(X = 0) = 0$
- For $1 \leq x < 2$, 
  \[
  F(x) = \sum_{x_i \leq 1} P(X = x_i) = P(X = 0) + P(X = 1) = \frac{1}{10}
  \]
- For $2 \leq x < 3$, 
  \[
  F(x) = \sum_{x_i \leq 2} P(X = x_i) = P(X = 0) + P(X = 1) + P(X = 2)
  = \frac{1}{10} + \frac{2}{10} = \frac{3}{10}
  \]
- Continuing ...

\[
F(x) = \begin{cases} 
  0 & x < 1 \\
  \frac{1}{10} & 1 \leq x < 2 \\
  \frac{3}{10} & 2 \leq x < 3 \\
  \frac{6}{10} & 3 \leq x < 4 \\
  1 & 4 \leq x 
\end{cases}
\]
Cumulative Distribution Function

A discrete random variable $X$ has CDF

$$F(x) = \begin{cases} 
0 & x < 1 \\
\frac{1}{10} & 1 \leq x < 2 \\
\frac{3}{10} & 2 \leq x < 3 \\
\frac{6}{10} & 3 \leq x < 4 \\
1 & 4 \leq x 
\end{cases}$$

(1)

What is its probability mass function?
CDF only changes value at 0, 1, 2, 3, 4 so $X$ takes values in \{0, 1, 2, 3, 4\}

- $F(0) = 0$ so $P(X = 0) = 0$
- $F(1) = \frac{1}{10} = P(X = 0) + P(X = 1)$ so $P(X = 1) = \frac{1}{10}$
- $F(2) = \frac{3}{10} = P(X = 0) + P(X = 1) + P(X = 2)$ so $P(X = 2) = \frac{2}{10}$
- $F(3) = \frac{6}{10} = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$ so $P(X = 3) = \frac{3}{10}$
- $F(4) = 1$ so $P(X = 4) = \frac{4}{10}$
Why are these important?

- Random variables: convenient way to represent events in the real world
- PMF and CDF: concise way to represent the probability of events

Note on notation:
- Convention is to use uppercase $X$ for random variables and lowercase $x$ for values e.g. $P(X = x)$.
- We’ll use $P(X = x)$, but alternatives are $P_X(x)$ or just $P(x)$ where RV is clear, or $p_X(x)$ or $p(x)$.
- We’ll use $P(X = x \text{ and } Y = y)$, but could use $P_{XY}(x, y)$ or just $P(x, y)$
Joint Probability Mass Function

Suppose we have two discrete random variables $X$ and $Y$ on same sample space $S$.

- $P(X = x \text{ and } Y = y)$ is called their joint probability mass function
- Let’s go back to sample space $S$. Remember RV $X$ is really a function mapping from $S$ to a real value i.e. should really be written $X(\omega)$. Ditto $Y$.
- Let $E_x = \{\omega \in S : X(\omega) = x\}$ be set of outcomes for which $X = x$
- Let $E_y = \{\omega \in S : Y(\omega) = y\}$ be set of outcomes for which $Y = y$
- $P(X = x) = P(E_x)$, $P(Y = y) = P(E_y)$
- Probability of both is $P(E_x \cap E_y)$ and $P(X = x \text{ and } Y = y) = P(E_x \cap E_y)$. 
Joint Probability Mass Functions

Example: operating system loyalty. Person buys one computer, then another. \( X = 1 \) if first computer runs windows, else 0. \( Y = 1 \) is second computer runs windows, else 0.

- Joint probability mass function:

<table>
<thead>
<tr>
<th></th>
<th>( x=0 )</th>
<th>( x=1 )</th>
<th>( P(Y=y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y=0 )</td>
<td>0.2</td>
<td>0.3</td>
<td>0.5</td>
</tr>
<tr>
<td>( y=1 )</td>
<td>0.1</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>( P(X=x) )</td>
<td>0.3</td>
<td>0.7</td>
<td>1</td>
</tr>
</tbody>
</table>

- \( P(X = 0 \text{ and } Y = 0) = 0.2, \ P(X = 0 \text{ and } Y = 1) = 0.3 \) etc.
Conditional Probability

• Recall for events we defined conditional probability
  \[ P(E|F) = \frac{P(E \cap F)}{P(F)} \]

• Similarly, for RVs we define
  \[ P(X = x|Y = y) = \frac{P(X=x \text{ and } Y=y)}{P(Y=y)} \]

• In fact
  \[ P(X = x|Y = y) = P(E_x|E_y) \]
  by noting that
  \[ P(X = x \text{ and } Y = y) = P(E_x \cap E_y) \]
  and
  \[ P(Y = y) = P(E_y) \]

• Example again:

<table>
<thead>
<tr>
<th></th>
<th>(x=0)</th>
<th>(x=1)</th>
<th>(P(Y=y))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(y=0)</td>
<td>0.2</td>
<td>0.3</td>
<td>0.5</td>
</tr>
<tr>
<td>(y=1)</td>
<td>0.1</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>(P(X=x))</td>
<td>0.3</td>
<td>0.7</td>
<td>1</td>
</tr>
</tbody>
</table>

• \(P(Y = 1|X = 0) = \frac{0.1}{0.2+0.1} = \frac{1}{3} \)
Marginal Distributions

Discrete random variable $Y$ takes values on $\{y_1, y_1, \cdots, y_m\}$. Then

- By chain rule $P(X = x \text{ and } Y = y_i) = P(Y = y_i | X = x)P(X = x)$
- So

\[
\sum_{i=1}^{m} P(X = x \text{ and } Y = y_i) = \sum_{i=1}^{m} P(Y = y_i | X = x)P(X = x) \\
= P(X = x) \sum_{i=1}^{m} P(Y = y_i | X = x) \\
= P(X = x)
\]

since $\sum_{i=1}^{m} P(Y = y_i | X = x) = 1$.

- Summing over the $y_i$ so that we are just left with $P(X = x)$ is called **marginalising** and $P(X = x)$ is called the **marginal distribution**.
Marginal Distributions

Discrete random variable $Y$ takes values on $\{y_1, y_1, \cdots, y_m\}$. Then

- By chain rule $P(X = x \text{ and } Y = y_i) = p(Y = y_i|X = x)P(X = x)$

Example again. Joint probability mass function:

<table>
<thead>
<tr>
<th></th>
<th>$x=0$</th>
<th>$x=1$</th>
<th>$P(Y=y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y=0$</td>
<td>0.2</td>
<td>0.3</td>
<td>0.5</td>
</tr>
<tr>
<td>$y=1$</td>
<td>0.1</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>$P(X=x)$</td>
<td>0.3</td>
<td>0.7</td>
<td>1</td>
</tr>
</tbody>
</table>

- $P(X = 0) = P(X = 0 \text{ and } Y = 0) + P(X = 0 \text{ and } Y = 1) = 0.2 + 0.1 = 0.3$
Chain Rule, Bayes and Independence

Since $P(X = x|Y = y) = P(E_x|E_y)$,
$P(X = x \text{ and } Y = y) = P(E_x \cap E_y)$, $P(Y = y) = P(E_y)$ we also have:

- **Chain rule**: $P(X = x \text{ and } Y = y) = P(X = x|Y = y)P(Y = y)$
  - $P(E_x \cap E_y) = P(E_x|E_y)P(E_y)$
- **Bayes rule**: $P(X = x|Y = y) = \frac{P(Y=y|X=x)P(X=x)}{P(Y=y)}$
  - $P(E_x|E_y) = \frac{P(E_y|E_x)P(E_x)}{P(E_y)}$
- **Independence**: two discrete random variables $X$ and $Y$ are independent if $P(X = x \text{ and } Y = y) = P(X = x)P(Y = y)$ for all $x$ and $y$
  - Events $E_x$ and $E_y$ are independent when $P(E_x \cap E_y) = p(E_x)p(E_y)$