Contents

Random Variables

- So far we have considered **random events**. An event can take any kind of value, e.g. heads/tails, colour of your eyes, age
- That means we can't do calculations using events. It's meaningless to add heads and tails for example, or blue and green
- This is akin to variable "typing" in programming. We need to define a quantity that is associated with a random event but which is real-valued, so that we can carry out arithmetic operations, etc.
- We use **random variables** for this. A random variable effectively maps every event to a real number

A **random variable** is a function that maps from the sample space S to the real line $\mathbb{R}$

- Write $X(\omega)$ where $\omega \subset S$ is an event
- $X(\omega) \subset \mathbb{R}$ in general, but we'll mostly think of $X(\omega)$ being single-valued
- Very often ω is dropped and just write X. This is just convenience though.
- When X can take only discrete values, e.g. $\{1, 2\}$ then it is called a **discrete** random variables
- Otherwise its a **continuous** random variable

Out of 2 coin tosses, how many came up heads. Let's this call random variable X (usual convention is to use upper case for RVs)

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

In general

- The set of outcomes for which $X = x$ is $E_x = \{\omega \mid X(\omega) = x, \omega \in S\}$
- So $P(X = x)$ is the probability that event E_x occurs, i.e. $P(X = x) = P(E_x)$

All the ideas regarding the probability of random events carry over to random variables (since random variables are just a mapping from events to numerical values)

Indicator Random Variable

Indicator Random Variables: take value 1 is event E occurs and 0 if event E does not occur.

$$I = \begin{cases} 1 & \text{if } E \text{ occurs} \\ 0 & \text{if } E \text{ doesn't occurs} \end{cases}$$

I is a random variable, a function of events in sample space S that takes values 0 or 1

Conditional Probability

- Recall for events we defined conditional probability $P(E \mid F) = \frac{P(E \cap F)}{P(F)}$
- For RVs, $P(X = x \mid Y = y) = \frac{P(X=x \text{ and } Y=y)}{P(Y=y)}$
- In fact $P(X = x \mid Y = y) = P(E_x \mid E_y)$ by noting that $P(X = x \text{ and } Y = y) = P(E_x \cap E_y)$ and $P(Y = y) = P(E_y)$

Example:

- Roll two dice. What is the probability that second dice is both 1 if both dice sum to 3?
- Let random variable X equal first value rolled, Y equal the sum. Want $P(X = 1 \mid Y = 3)$
- $P(X = 1 \text{ and } Y = 3) = P(\{1, 2\}) = 1/36$
- $P(Y = 3) = P(\{(1, 2), (2, 1)\}) = 2/36$
- So $P(X = 1 \mid Y = 3) = \frac{1/36}{2/36} = \frac{1}{2}$

Marginalisation

Discrete random variable Y takes values on $\{y_1, y_2, \ldots, y_m\}$ Then

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

Chain Rule, Bayes and Independence

Since

- $P(X = x \mid Y = y) = P(E_x \mid E_y)$
- $P(X = x \text{ and } Y = y) = P(E_x \cap E_y)$
- $P(Y = y) = P(E_y)$

we also have:

- Chain rule: $P(X = x \text{ and } Y = y) = P(X = x \mid Y = y)P(Y = y)$
  - cf $P(E_x \cap E_y) = P(E_x \mid E_y)P(E_y)$
- Bayes Rule: $P(X = x \mid Y = y) = \frac{P(Y=y|X=x)P(X=x)}{P(Y=y)}$
  - cf $P(E_x \mid E_y) = \frac{P(E_y|E_x)P(E_x)}{P(E_y)}$
- 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
  - cf Events E_x and E_y are indepdenent when $P(E_x \cap E_y) = p(E_x)P(E_y)$

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, or sometimes to $P_x(c)$ or $p_x(x)$

Suppose discrete random variable X can take values $x_1, x_2, \ldots, x_n$

- We have probability $P(X = x_1), P(X = x_2), \ldots, P(X = x_n)$

- This is called the **probability mass function** (PMF) of X

Example: The number of heads from two coin flips

- $P(X = 0 = \frac{1}{4}$ (event $\{(T, T)\}$)
- $P(X = 1 = \frac{1}{2}$ (event $\{(H, T), (T, H)\}$)
- $P(X = 2 = \frac{1}{4}$ (event $\{(H, H)\}$)

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 $D = \{x_1, x_2, \ldots, x_n\}$ the CDF is $F(a) = P(X \leq a) = \sum_{x_i \leq a} P(X = x_i)$
- If $a \leq b$ then $F(a) \leq F(y)$

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…

A discrete random variable X has a CDF $F(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{10} & 1 \leq x < 2 \\ \frac{3}{10} & 2 \leq x < 3 \\ \frac{6}{10} & 3 \leq x < 4 \\ 1 & 4 \leq x \end{cases}$

Why are these Important?

- Random variables: convenient way to represent events in the real world
- PMF and CDF: concise way to represent the probability og 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 $P(x, y)$