Overview

Recall module is roughly split into four parts:

1. Random events: counting, events, axioms of probability, Bayes, independence
2. Random variables: discrete RVs, mean and variance, correlation, conditional expectation
   Mid-term
3. Inequalities and laws of large numbers: Markov, Chebyshev bounds, sample mean, weak law of large numbers, central limit theorem, confidence intervals, bootstrapping
4. Statistical models: continuous random variables, logistic regression, least squares

Overview

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

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. Its 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.

Example: Indicator Random Variable

Indicator Random Variable: takes value 1 if event E occurs and 0 if event E does not occur.

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

Some other examples:

- Out of 2 coin tosses, how many came up heads (so if the event is (H, T) then the random variable takes the value 1, and so on)
- When I throw two dice, what is the sum

Random Variables

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.
- Very often ω is dropped and just write X. This is just a convenience though.
- When X can take only discrete values e.g. $\{1, 2\}$ then it is called a **discrete** random variable.
- Otherwise its a **continuous** random variable.

Indicator 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 events in sample space S that takes values 0 or 1.

- Sometimes we might see e.g. for lunch today random variable $X = \mathit{Sandwich}$.
- 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. $X = \mathit{Sandwich}$ is the same as $I = 1$.

Random Variables

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).

- 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 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.

In general,

- 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. $P(X = x) = P(E_x)$.

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

Conditional Probability

- Recall for events we defined conditional probability $P(E|F) = \frac{P(E \cap F)}{P(F)}$
- For RVs $P(X = x|Y = y) = \frac{P(X=x \text{ and } Y=y)}{P(Y=y)}$
- 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:

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

Marginalisation

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

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

Proof is same as before:

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

$$\begin{aligned}\sum_{i=1}^{m} P(X = x \text{ and } Y = y_i) &= \sum_{i=1}^{m} P(Y = y_i | X = x) P(X = x) \\ &= P(X = x) \sum_{i=1}^{m} P(Y = y_i | X = x) \\ &= P(X = x)\end{aligned}$$

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

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: $P(X = x \text{ and } Y = y) = P(X = x|Y = y)P(Y = y)$
  - cf $P(E_x \cap E_y) = P(E_x|E_y)P(E_y)$
- Bayes rule: $P(X = x|Y = y) = \frac{P(Y=y|X=x)P(X=x)}{P(Y=y)}$
  - cf $P(E_x|E_y) = \frac{P(E_y|E_x)P(E_x)}{p(E_y)}$
- 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
  - cf Events E_x and E_y are independent 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 probabilities $P(X = x_1)$, $P(X = x_2), \ldots,$ $P(X = x_n)$
- This is called the **probability mass function** (PMF) of X.

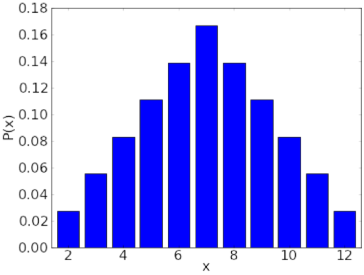
Example: The number of heads from two coin flips.

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

Probability Mass Function

Another example. The sum of two dice.

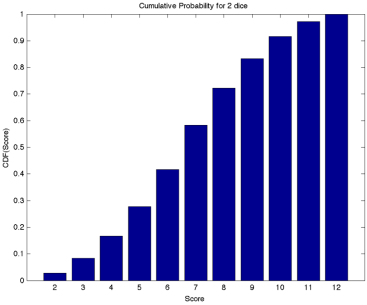
- $P(X = 2) = \frac{1}{36}$ (event $\{(1, 1)\}$)
- $P(X = 3) = \frac{2}{36}$ (event $\{((1, 2), (2, 1)\}$)
- $P(X = 4) = \frac{3}{36}$ (event $\{(1, 3), (2, 2), (3, 1)\}$)



PMF for sum of two dice

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 $2 \leq x < 3$,

$$\begin{aligned} F(x) &= \sum_{x_i \leq 2} P(X = x_i) = P(X = 0) + P(X = 1) + P(X = 2) \\ &= \frac{1}{10} + \frac{2}{10} = \frac{3}{10} \end{aligned}$$

- 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

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

What is its probability mass function ?

Cumulative Distribution 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 $P(x, y)$