Contents

1 Probability Space

There are often various approaches to probability each with its own advantages and disadvantages.

Experiment is a procedure that can be infinitely repeated and has a well-defined set of possible outcomes, known as the sample space.

The observation/result of the experiment are termed as outcomes.

1.1 Classical Approach

Probability of an event E is defined to be:

$$P(E) = \frac{\text{Number of outcomes in } E}{\text{Total number of outcomes}}$$

Some examples are tossing a coin or rolling a die. Disadvantages:

- Unable to model biases. It says nothing about cases where no physical symmetry exists.
- Doesn't deal with cases where total outcomes are infinite.

1.2 Frequentist Approach

Also known as the relative frequency approach or frequentism. It defines an event's probability as the limit of its relative frequency in many trials.

Probability is defined to be:

$$P(E) = \lim_{n \to \infty} \frac{n_E}{n}$$

where an experiment is conducted n times and event E occurs n_E times. Disadvantages:

- It isn't efficient to conduct an experiment multiple times just to find the probability of an event occuring.
- It is unable to deal with subjective belief. Eg: Suppose a cricket expert says there is a 50% of RCB winning the IPL this year. It doesn't mean that the RCB has won half the titles in the past.

1.3 Axiomatic Approach

1.3.1 Probability Space

The triple (S, F, P) is referred to as a probability space where:

- S : Sample space, set of all possible outcomes of the experiment.
- F : Event Space, subset of the sample space
- P : Probability Measure

1.3.2 Sample Space

S can either be finite or countably infinite or uncountably infinite.

Examples for S:

- Finite Sample Space: Single coin toss $S = \{H, T\}$ and two coin tosses $S = \{HH, HT, TT, TH\}$.
- Countably Infinite Sample Space: Keep tossing a coin till you get a head $S = \{H, TH, TTH...\}$.
- Uncountably Infinite Sample Space: We have a circular dart board and we are measuring the angle at which a dart hits the board. $S = [0, 2\pi]$

1.4 Event Space

Collection of events is called an event space, there are some properties to be satisified such as: it has to be a "Sigma Field".

1.4.1 Sigma Field

A sigma field (or sigma algebra) F is a collection of subsets of S which satisfies the following properties:

- $S \in F$
- If $E \in F$, then $E^C \in F$
- If $E_1, E_2, E_3 \dots \in F$, then $\bigcup_{i=1}^{\infty} E_i \in F$

1.4.2 Examples for Event Space

- Smallest possible event space:

$$F = \{\phi, S\}$$

- Next non-trivial event space:

$$F = \{\phi, E, E^c, S\}$$

- If $E_1 \in F$ and $E_2 \in F$, then $E_1 \cap E_2 \in F$ (Proof in 1.4.3)
- For S=$\{1, 2, 3, 4, 5, 6\}$, $E_1 = \{1, 2\}$ and $E_2 = \{3, 4\}$. The smallest event space containing E_1 and E_2 is:

$$F = \{\phi, S, E_1, E_1^C, E_2, E_2^C, E_1 \cup E_2, (E_1 \cup E_2)^C\}$$

1.4.3 Proposition 1

A$_1$, A$_2$, A$_3$,....A$_n$ $\in F$, then $\bigcap_{i=1}^{n}$ A$_i$ $\in F$.

Proof. If A$_1$, A$_2$, A$_3$,....A$_n$ $\in F$, then A$_1^c$, A$_2^c$, A$_3^c$,....A$_n^c$ $\in F$ and $\bigcup_{i=1}^{n}$ A$_i^c$ $\in F$. Then by property 2, $(\bigcup_{i=1}^{n} \mathrm{A}_i^c)^c = \bigcap_{i=1}^{n} \mathrm{A}_i \in F$. ☐

1.4.4 Proposition 2

A, B $\in F$, then A\B = A - B $\in F$.

Proof. If B $\in F$, then B^c $\in F$ by property 2. So, A $\cap$ B^c = A\B $\in F$ (As seen in proposition 1). ☐

1.5 Probability Measure

The probability measure P is a function returning an event's probability. A probability is a real number between zero and one.

$$P : F \to [0, 1]$$

P has to satisfy the following 3 axioms:

- $P(E) \geq 0$
- $P(S) = 1$
- If $E_1, E_2 \cdots \in F$ such that $E_i \cap E_j = \phi$ then:

$$P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$$

For two disjoint sets: $P(E_1 \cup E_2) = P(E_1) + P(E_2) + \sum P(\phi)$
We will later see that $P(\phi)$ is indeed 0.

1.6 Derived Properties of Probability

1.
$$P(E^C) = 1 - P(E)$$

Proof.
$$E \cup E^C = S$$
$$P(E) + P(E^C) = 1$$

☐

2. For any two events E_1 and E_2,

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

Proof.

$$P(E_1 \cup E_2) = P(E_1) + P(E_2 \cap E_1^C)$$

Now,

$$E_2 = (E_2 \cap E_1) \cup (E_2 \cap E_1^C)$$

$$\Rightarrow P(E_2) = P(E_1 \cap E_2) + P(E_1^C \cap E_2)$$

Also,

$$P(E_1 \cup E_2) = P(E_1) + P(E_1^C \cap E_2)$$

Substitute the required value in the final equation. □

Question: $S = \{1, 2, 3, 4, 5, 6\}$. 1 and 5 are equally likely and probability of getting a 6 is one-third.

Find minimum and maximum probability that we get an even number.

Answer:

$$Minimum\ Prob. = \frac{1}{3}\ \ when\ \ P_2 = P_4 = 0$$

$$Maximum\ Prob. = 1,\ \ P_2 + P_4 = \frac{2}{3}$$

2 Conditional Property

Given that an event A has occured.

$(S, F, P) \to$ Original probability space

If additional info has been given that A has occured, probability space need to be suitably modified.

eg: $S = \{1, 2, 3, 4, 5, 6\},\ E_1 = \{1, 2\},\ E_2 = \{3, 4\}$

$$F = \{\phi, S, E_1, E_1^C, E_2, E_2^C, E_1 \cup E_2, (E_1 \cup E_2)^C\}$$

and event $A = E_1^c = \{3, 4, 5, 6\}$ has occured.

Then, $F_A = \{\phi, A, \{3, 4\}, \{5, 6\}\}$.(shown later)

2.1 Modified probability space

Then,

- $S_A = A$. (Modified Sample Space)
- $F_A = \{(E \cap A) | E \in F\} \to E \cap A \in F$. (Modified Event Space)

  (Also if some event $C \cap A = \phi$, then C won't occur)

  To prove: F_A also satisfies event space axioms.(see sec. 1.4.1)

1. $A \in F_A \to S \cap A = A$(S was original sample space) $\Rightarrow A \in F_A$
2. $D \in F_A \Rightarrow D = E \cap A, D^c \in F_A$
   As,
   $D^c = A \setminus D = E^c \cap A \in F \quad (Because\ E^c \in F)$
3.
$$D_1, D_2, \cdots \in F_A$$
$$(E_1 \cap A), \cdots \in F_A$$
$$E_1, E_2 \cdots \in F$$
$$\Rightarrow \bigcup_{i=1}^{\infty} E_i \in F_A$$
$$\Rightarrow (\bigcup_{i=1}^{\infty} E_i) \cap A \in F_A$$
$$\Rightarrow (\bigcup_{i=1}^{\infty} E_i \cap A) \in F_A \Rightarrow \bigcup_{i=1}^{\infty} D_i \in F_A$$

Hence, F_A is an event space.

- Modified probability measure

$$P(E/A) = \frac{P(E \cap A)}{P(A)}$$

This definition is called conditional probability measure for any $E \in F$.

eg: $F_A = \{\phi, \{3,4,5,6\}, \{3,4\}, \{5,6\}\}$, then $P(\{3,4\}/\{3,4,5,6\}) = 1/2$

Now, we need to prove $P(E/A)$ satisfies the 3 axioms of probability measure.(see sec. 1.5)

  - $P(E/A) \geq 0$ (as ratio of two nos. which are positive)
  - $P(S/A) = 1$
  - $B_1, B_2 \cdots$ are all mutually disjoint.

$$P(\bigcap_{i=1}^{\infty} B_i/A) = \frac{P(\bigcup_{i=1}^{\infty} B_i \cap A)}{P(A)}$$
$$= \frac{\sum_{i=1}^{\infty} P(B_i \cap A)}{P(A)} = \sum_{i=1}^{\infty} P(B_i/A)$$
$$\Rightarrow P(\bigcap_{i=1}^{\infty} B_i/A) = \sum_{i=1}^{\infty} P(B_i/A)$$

3 Total probability theorem

Events $A_1, \cdots, A_n \in F$ which are all mutually exclusive/disjoint and exhaustive. Then,

$$A_i \cap A_j = \phi \quad \forall\, i, j$$

$$\bigcup_{i=1}^{n} A_i = S$$

$$P(B) = \sum_{i=1}^{n} P(B/A_i)P(A_i)$$

the total probability theorem expresses $P(B)$ in terms of conditional probability $P(B/A_i)$ & prior probability $P(A_i)$

Proof.

$$B = \bigcup_{i=1}^{n} (B \cap A_i)$$

A_i's are disjoint, so $(B \cap A_i)$ are also disjoint.

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B/A_i)P(A_i)$$

□

3.1 Question

Two factories manufacture zoggles. 20% of F_1 are defective. 5% of F_2 are defective.

In any week, F_1 produces twice the number of zoggles as F_2. What is the probability that a zoggle chosen randomly in a week is defective?

$$P(D) = P(F_1)P(D/F_1) + P(F_2)P(D/F_2)$$
$$= \frac{2}{3} \times \frac{1}{5} + \frac{1}{3} \times \frac{1}{20} = \frac{3}{20}$$

4 Bayes Theorem

$A_1, \cdots, A_n$ are events which are mutually exclusive and exhaustive. B be an arbitrary event. Then,

$$P(A_i/B) = \frac{P(B/A_i)P(A_i)}{\sum_{i=1}^{n} P(B/A_i)P(A_i)}$$

- $P(A_i)$- Prior probability
- $P(B/A_i)$- Likelihood

- $P(A_i/B)$- Posterior probability

Bayes theorem expresses posterior probabilities $P(A_i/B)$ in terms of prior probabilities $P(A_i)$ and likehoods $P(B/A_i)$.

In some experiments $P(A_i)$ are all same, $P(A_i) = 1/n$. Then posterior probabilities are proportional to likelyhoods:

$$P(A_i/B) = \frac{P(B/A_i)}{\sum_{i=1}^{n} P(B/A_i)}$$

$$\Rightarrow P(A_i/B) \propto P(B/A_i)$$

Proof.

$$P(A_i/B) = \frac{P(A_i \cap B)}{P(B)}$$

$$P(A_i/B) = \frac{P(B/A_i)P(A_i)}{\sum_{i=1}^{n} P(B/A_i)P(A_i)}$$

□

4.0.1 Example

In answering a question in a multiple choice test, a student knows the answer with probability p and guesses the answer otherwise. If he/she guesses from m choices, the probability of being correct is $\frac{1}{m}$. Find the conditional probability that the student knew the answer if he/ she answered correctly.

Ans: A_1 = knowing the answer. A_2 = Guessing the answer.

B =Answer is correct. $A_1 \subseteq B$

$P(A_1) = p, P(A_2) = 1 - p, P(B/A_1) = 1, P(B/A_2) = \frac{1}{m}$

$$P(A_1/B) = \frac{P(B/A_1)P(A_1)}{P(B/A_1)P(A_1) + P(B/A_2)P(A_2)}$$

$$= \frac{p}{p + (1-p)\frac{1}{m}}$$

5 Indepedent Events

A & B are said to be independent. If

$$P(A \cap B) = P(A)P(B)$$

In terms of conditional probability,

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)P(B)}{P(B)} = P(B)$$

$$\Rightarrow P(B/A) = P(B)$$

Probability of event B remains same with or without conditioning on A. Hence, B is said to be independent of A. Knowledge of occurence of event A does not give any information about B.

If A & B are independent, then A & B^c are also independent.

Proof.

$$P(B^c/A) = \frac{P(B^c \cap A)}{P(A)} = \frac{P(A) - P(A \cap B)}{P(A)} = \frac{P(A) - P(A)P(B)}{P(A)} = P(B^c)$$

□

5.1 Important Results

1. $P(A/B)$ may be greater than, less than or equal to $P(A)$.
2. Independent events and mututally exclusive events are different.
   Indepedence: $P(A \cap B) = P(A)P(B)$
   Mututally exclusive: $A \cap B = \phi$
   (a) (Indepedent but not mutually exclusive). Coin toss followed by throwing dice experiment.
   $$S = \{(H,1),(H,2),\cdots,(H,6),(T,1),\cdots,(T,6)\}$$
   F = Power set of S (Always an event space)
   $A = \{(H,1),\cdots,(H,6)\}$, $B = \{(H,2),(H,4),(H,6),(T,2),(T,4),(T,6)$
   $P(A) = 1/2$ & $P(B) = 1/2$
   $$P(A \cap B) = P(\{(H,2),(H,4),(H,6)\}) = \frac{3}{12} = \frac{1}{4}$$
   $$P(A \cap B) = P(A)P(B)$$
   (b) (Not independent but mutually exclusive) If the events are mutually exclusive $\Rightarrow$ they are not independent.
   Single Coin Toss: $A = \{H\}$, $B = \{T\} \rightarrow$ Mututally exclusive.
   $P(A \cap B) = 0$, $P(A)P(B) = \frac{1}{4}$

5.2 Conditionally indepedent events

A & B are said to be conditionally indepedent given C if

$$P((A \cap B)/C) = P(A/C)P(B/C)$$

In terms of conditional probabilities,

$$P(B/C) = \frac{P((A \cap B)/C)}{P(A/C)} = \frac{\frac{P(A \cap B \cap C)}{P(C)}}{\frac{P(A \cap C)}{P(C)}} = \frac{P(A \cap B \cap C)}{P(A \cap C)} = P(B/(A \cap C))$$

Indepedent Events: $P(A \cap B) = P(A)P(B)$ & $P(B/A) = P(B)$.

Conditionally indepedent events: $P(A \cap B/C) = P(A/C)P(B/C)$ & $P(B/(A \cap))$

5.2.1 Example

Two fair coins are tossed, $S = \{HH, HT, TH, TT\}$, $A = \{HH, HT\}$, $B = \{HH, TH\}$, $C = \{HH\}$ & $D = \{HT, TH\}$

1. Are A & B independent?

$$P(A \cap B) = P(\{HH\}) = \frac{1}{4} = P(A)P(B)$$

2. Are A & B conditionally indepedent given C?

$$P((A \cap B)/C) = P(A/C) = P(B/C) = 1$$

$$P(A/C) = \frac{P(A \cap C)}{P(C)} = \frac{\frac{1}{4}}{\frac{1}{4}} = 1$$

3. Are A & B conditionally indepedent given D?

$$P(A/D) = \frac{P(A \cap D)}{P(D)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$$

$$P(B/D) = \frac{P(B \cap D)}{P(D)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$$

$$P(A \cap B/D) = 0$$

Hence, they are not conditionally indepedent given D.

If A and B are independent, it doesn't imply A and B will be conditionally indepedent given C and vice-versa.

5.3 Indepedence of collection of events

Three events A_1, A_2 & A_3 are said to be indepedent if:

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$$

$$P(A_1 \cap A_2) = P(A_1)P(A_2) \quad P(A_2 \cap A_3) = P(A_2)P(A_3) \quad P(A_3 \cap A_1) = P(A_3)P(A_1)$$

5.3.1 Chain rule of Probability

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2)$$

$$P(\bigcap_{i=1}^{n} A_i) = P(A_1)\prod_{i=2}^{n} P(A_i/A_1 \cdots A_{i-1})$$

For independent events A_1, A_2, A_3.

$$P(A_1)P(A_2/A_1)P(A_3/(A_1 \cap A_2)) = P(A_1)P(A_2)P(A_3)$$

$$\Rightarrow P(A_2/A_1)P(A_3/(A_1 \cap A_2)) = P(A_2)P(A_3)$$

Hence, if A_1, A_2, A_3 are independent events, the first condition doesn't imply the other conditions.

Eg: Pair-wise independence does not imply that three events are independent. Consider example of two coin tosses: $S = \{HH, HT, TH, TT\}, A_1 = \{HH, HT\}, A_2 = \{HH, TH\}, A_3 = \{HT, TH\}$.

A_1 & A_2 are indepedent $P(A_1 \cap A_2) = P(A_1)P(A_2)$

A_2 & A_3 are indepedent $P(A_3 \cap A_2) = P(A_3)P(A_2)$

A_1 & A_3 are indepedent $P(A_1 \cap A_3) = P(A_1)P(A_3)$

A_1, A_2 & A_3 are not still not independent as $P(A_1 \cap A_2 \cap A_3) \neq P(A_1)P(A_2)P(A_3)$.

6 Continuity of Probability

Ω is an equivalent notation for sample space.

Consider the probability space (S, F, P).

$E_i \in F$. Let $E_1 \subseteq E_2 \subseteq \cdots$ be countably infinite sequence of events (increasing sequence of events). Then

$$P(\bigcup_{i=1}^{\infty} E_i) = \lim_{i \to \infty} P(E_i)$$

.

$$\bigcup_{i=1}^{\infty} E_i \in F$$

Pushing the limit from inside the probability expression to outside.

$$P(\bigcup_{i=1}^{\infty} E_i) = \lim_{i \to \infty} P(E_i)$$

$$P(\lim_{n \to \infty} \bigcup_{i=1}^{n} E_i) = \lim_{i \to \infty} P(E_i)$$

Exchanging limits with probability $\to$ non trivial operation. Exchanging limits with differentiation, integration etc. $\to$ require a proof.

- $$E_1 \subseteq E_2 \subseteq \cdots$$

  $$P(\bigcup_{i=1}^{\infty} E_i) = \lim_{i \to \infty} P(E_i)$$

  Proof. We shall use the third axiom. $D_1, D_2 \cdots \to$ mutually exclusive events.

  $$P(\bigcup_{i=1}^{\infty} D_i) = \sum_{i=1}^{\infty} P(D_i)$$

$E_1 \subseteq E_2 \subseteq \rightarrow$ not mutually exclusive.

Hence we construct a sequence which are mututally exclusive and unions of both sequences have to be the same.

$A_1 = E_1,\ A_2 = E_2 \setminus E_1,\ A_3 = E_3 \setminus E_2\ ,\ \cdots.$

$$P(\bigcup_{i=1}^{\infty} E_i) = P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) = \lim_{n\to\infty} \sum_{i=1}^{\infty} P(A_i)$$

$$\Rightarrow \lim_{n\to\infty} \sum_{i=1}^{n} P(A_i) = \lim_{n\to\infty} P(\bigcup_{i=1}^{n} A_i) = \lim_{n\to\infty} P(E_n)$$

If there exists a subsequence of events, which is increasing.

$\square$

- Let $E_1 \supseteq E_2 \supseteq \cdots$ be a countably infinte sequence of events(decreasing sequence of events)

$$P(\bigcap_{i=1}^{\infty} E_i) = \lim_{i\to\infty} P(E_i)$$

Proof. We can use the result of the previous proof.

$$E_1^c \subseteq E_2^c \subseteq \cdots$$

which is an increasing set of events

$$P(\bigcup_{i=1}^{\infty} E_i^c) = \lim_{i\to\infty} P(E_i^c)$$

as,

$$\left(\bigcap_{i=1}^{\infty} E_i\right) = \left(\bigcup_{i=1}^{\infty} E_i^c\right)^c$$

$$\Rightarrow 1 - P(\bigcup_{i=1}^{\infty} E_i^c) = \lim_{i\to\infty} 1 - P(E_i^c)$$

$$P\left(\bigcap_{i=1}^{\infty} E_i\right) = \lim_{i\to\infty} P(E_i)$$

$\square$

7 Random Variables

Consider a probability space (S, F, P) and an experiment is conducted.

Probability spaces vary a lot based on the experiment. We use random variables to be able to develop a theory of probability which is indepedent of the actual experiment which is performed.

$$S \xrightarrow{\text{R.V.}} \mathbb{R}$$
$$F \xrightarrow{\text{R.V.}} B$$

Random Variable is a function which maps the sample space to the real line. It also maps the event space to the Borel σ-algebra, B.

7.1 Borel σ-algebra

It is the smallest σ- algebra which contains sets of the form $(-\infty, x]$ $\forall x \in \mathbb{R}$.

$$B = \{\mathbb{R}, (-\infty, x], (x, \infty), \phi, (-\infty, x), (x, y), [x, y], \{x\}, (x, y], [x, y)\}$$

$$E_i = (-\infty, x_i] \quad x_i = x - \frac{1}{i}$$
$$\bigcup_{i=1}^{\infty} (-\infty, x_i] = \bigcup_{i=1}^{\infty} E_i = (-\infty, x)$$

Random Variable $X : S \to R$

X has to be a measurable function.

A function is said to be measurable if pre-image of $(-\infty, x]$ $\forall x \in \mathbb{R}$ is in the event space.

$$X : S \to R$$
$$\text{inverse image} \leftarrow (-\infty, x]$$
$$X^{-1}((-\infty, x]) \subseteq S \qquad X^{-1}((-\infty, x]) \in F$$

eg: $X : \{1, 2, 3, 4, 5, 6\} \to \{0, 1, 0, 1, 0, 1\}$

$$X^{-}1\left((-\infty, 0.5]\right) = \{1, 3, 5\}$$

In the above example, X is not an invertible mapping. We are using X^{-1} as a notation.

X is a map from S to $\mathbb{R}$ such that

$$X : S \to \mathbb{R}$$
$$X^{-1}\left((-\infty, x]\right) \leftarrow (-\infty, x] \qquad X^{-1}\left((-\infty, x]\right) \in F$$

7.2 Examples

- X which is a random variable.

  $S = \{a, b, c\}\ ,\ F = \{\phi, \{a\}, \{b, c\}, S\}\ ,\ X : S \to \mathbb{R}$

  $X(\omega) = 0\ ,\ \omega = a$

  $X(\omega) = 1\ ,\ \omega = b, c$

  1. $(-\infty, x]\ ,\ x < 0\ ,\ X^{-1}((-\infty, x]) = \phi \in F$
  2. $(-\infty, x]\ ,\ 0 \le x < 1\ ,\ X^{-1}((-\infty, x]) = \phi \in F$
  3. $(-\infty, x]\ ,\ x \ge 1\ ,\ X^{-1}((-\infty, x]) = S \in F$

- X which is not a random variable $S = \{a, b, c\}\ ,\ F = \{\phi, \{a\}, \{b, c\}, S\}\ ,\ X : S \to \mathbb{R}$

  $X(\omega) = 0\ ,\ \omega = b$

  $X(\omega) = 1\ ,\ \omega = a, c$

  But $X^{-1}((-\infty, x])\ ,\ 0 \le x < 1 = \{b\}\ ,\ \{b\} \notin F$

  X is not a random variable.

If S is finite & F is a power set of S, then every function is a random variable.

$$X^{-1}((-\infty, x]) \subseteq S$$

F is set of all subsets of S.

7.3 Conditions on random variable

Theorem 7.1. *Let a probability space be (S, F, P) and X is a random variable $X : S \to \mathbb{R}$. The following conditions hold:*

- $X^{-1}((-\infty, x)) \in F$
- $X^{-1}(\{x\}) \in F$
- $X^{-1}((x_1, x_2]) \in F$
- $X^{-1}((x_1, x_2)) \in F$

Proof. Applying the axioms of the event space we want to show the above conditions are true.

$A_i = X^{-1}((-\infty, x])$ $\qquad x_i = x - \frac{1}{i}$. Now,

$$\bigcup_{i=1}^{\infty} (-\infty, x_i] = (-\infty, x)$$

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} X^{-1}((-\infty, x_i]) = X^{-1}((-\infty, x)) \in F$$

$$X^{-1}((-\infty, x] \cap (-\infty, x)^c) = X^{-1}(\{x\}) \in F$$

Similarly the other conditions can also be proved. $\square$

8 Cumulative distribution function

It is denoted by $F_X(x)$.

$$F_X(x) = P(X \leq x) = P((-\infty, x]) = P(X^{-1}((-\infty, x]))$$

The above definitions are different notations used, the last one is well defined from the probability space.

8.1 Properties of CDF

1. $F_X(x)$ is a monotonically non-decreasing function of x.

   Proof. If $x_2 \geq x_1$, then $F_X(x_2) \geq F_X(x_1)$.

   $$F_X(x_2) = P(X \leq x_2) = P(X \leq x_1) + P(x_1 < X \leq x_2)$$
   $$F_X(x_2) = P(X \leq x_2) \geq P(X \leq x_1) - F_X(x_1)$$

   $\square$

2. $\lim_{x\to\infty} F_X(x) = 1$

   Proof. Construct a decreasing set of events. $A_i = (-\infty, i]$, $B_i = X^{-1}(A_i)$.

   $$\bigcap_{i=1}^{\infty} A_i = \mathbb{R}$$

   $$\bigcap_{i=1}^{\infty} B_i = S$$

   Applying continuity of probability,

   $$\lim_{i\to\infty} P(B_i) = P(\bigcup_{i=1}^{\infty} B_i) = P(S) = 0$$

   $$\lim_{i\to\infty} P(X \leq i) = \lim_{x\to\infty} P(X \leq x) = \lim_{x\to\infty} F_X(x)$$

   The change from integers to real numbers x in the last step is valid.

   $\square$

3. $\lim_{x\to-\infty} F_X(x) = 0$

Proof. Construct a decreasing set of events. $A_i = (-\infty, -i]$, $B_i = X^{-1}(A_i)$.

$$B_1 \supseteq B_2 \supseteq \cdots$$

$$\bigcap_{i=1}^{\infty} B_i = \phi$$

$$P(\bigcap_{i=1}^{\infty} B_i) = \lim_{i\to\infty} P(B_i) = 0$$

$$\lim_{i\to\infty} P(B_i) = \lim_{i\to\infty} P((-\infty, i]) = \lim_{x\to\infty} P((-\infty, x]) \to (\text{x = all real nos.})$$

$$\lim_{x\to\infty} P((-\infty, x]) = \lim_{x\to\infty} F_X(x) = 0$$

□

4. $F_X(x)$ is a right continuous function.

$$\lim_{x\to x_0^+} F_X(x) = F_X(x_0)$$

Right continuous function: $\lim_{x\to x_0^+} f(x) = f(x_0)$

Left continuous function: $\lim_{x\to x_0^-} f(x) = f(x_0)$

Continuous function(Approaching from either left or right): $\lim_{x\to x_0} f(x) = f(x_0)$

Proof. Decreasing sequence of events $B_i = (X \leq x + \frac{1}{i})$

$$B_1 \supseteq B_2 \supseteq \cdots$$

$$\bigcap_{i=1}^{\infty} B_i = \{X \leq x\}$$

$$P(\bigcap_{i=1}^{\infty} B_i) = P(X \leq x) = F_X(x)$$

$$\Rightarrow \lim_{i\to\infty} P(B_i) = \lim_{i\to\infty} P(X \leq x + \frac{1}{i})$$

We are changing the variable from an integer to real.

$$= \lim_{\epsilon\to 0} P(X \leq x + \epsilon) = \lim_{\epsilon\to 0^+} F_X(x + \epsilon)$$

$$= \lim_{x_0\to x^+} F_X(x_0) \quad \text{Change of variable: } x_0 = x + \epsilon$$

□

8.2 Indicator Random Variable

$$I_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A \end{cases}$$

$$I_A : S \to \mathbb{R}$$

$$F_{I_A}(x) = P(I_A \leq x)$$

If $x < 0$, $P(I_A \leq x) = 0$.
If $0 \leq x < 1$, $P(I_A \leq x) = P(A^c) = 1 - P(A)$.
If $x \geq 1$, $P(I_A \leq x) = 1$.

$$\lim_{x \to 0^+} F_{I_A}(x) = P(A^c) \qquad F_{I_A}(0) = P(A^c)$$

$$\lim_{x \to 1^+} F_{I_A}(x) = 1 \qquad F_{I_A}(1) = 1$$

8.3 Examples of CDF's and non-CDF's

- Is $F_1(x)$ a valid CDF?

$$F_1(x) = \begin{cases} 1, & \text{if } x \leq 0 \\ 0.5, & \text{if } 0 < x \leq 1 \\ 0.25 + 0.25x, & \text{if } 1 < x \leq 3 \\ 1, & \text{if } 3 < x \end{cases}$$

No, as it is not right continuous at $x = 0$.

$$\lim_{x = 0^+} F_1(x) = 0.5, \quad F_1(0) = 0$$

- Is $F_2(x)$ a valid CDF?

$$F_2(x) = \begin{cases} 0, & \text{if } x < 0 \\ 0.5, & \text{if } 0 \leq x < 1 \\ 0.75, & \text{if } 1 \leq x < 3 \\ 1, & \text{if } 3 \leq x \end{cases}$$

It is valid as it satisfies all 4 properties.

- Is $F_3(x)$ a valid CDF?

$$F_3(x) = \begin{cases} 1, & \text{if } x < 0 \\ 0.5, & \text{if } 0 \leq x < 1 \\ 0.25, & \text{if } 1 \leq x < 3 \\ 1, & \text{if } 3 \leq x \end{cases}$$

$F_3(x)$ is not a valid CDF as it is not non-decreasing.

Lemma 8.1. *For any $x \in \mathbb{R}$,*

$$P(X = x) = P(X \leq x) - P(X < x)$$

$$P(X = x) = F_X(x) - \lim_{\epsilon \to 0} F_X(x - \epsilon)$$

Upon applying continuity of probability.

Proof.

$$B_1 \subseteq B_2 \subseteq \cdots$$

$$B_i = \{X \leq x - \frac{1}{i}\} \qquad \bigcup_{i=1}^{\infty} B_i = \{X < x\}$$

$$P(X < x) = P(\bigcup_{i=1}^{\infty} B_i) = \lim_{i \to \infty} B_i = \lim_{\epsilon \to \infty} P(X \leq x - \epsilon) = \lim_{\epsilon \to 0} F_X(x - \epsilon)$$

□

Corollary 8.1.1. *$F_X(.)$ is left continuous if and only if $P(X = x) = 0 \; \forall \; x \in \mathbb{R}$. From lemma,*

$$P(X = x) = F_X(x) - \lim_{\epsilon \to 0^+} F_X(x - \epsilon)$$

A function is left continuous if

$$\lim_{\epsilon \to 0^+} F_X(x - \epsilon) = F_X(x) \quad \forall x \in \mathbb{R}$$

$$\Rightarrow P(X = x) = 0 \quad \forall x \in \mathbb{R}$$

$F_X(.)$ is continuous, means that it is both right continuous, and in particular, left continuous.

9 Types of Random Variables

- Continuous random variable
- Discrete random variable
- Mixed random variable

9.1 Continuous random variables

A random variable X with cummulative distribution function $F_X(.)$ is said to be a continuous random variable if $F_X(.)$ is continuous.

$P(X = x) = 0 \quad \forall\, x \in \mathbb{R}$, probability of every point is zero.

An example for a continuous random variable is:

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ x, & \text{if } 0 \leq x < 1 \\ 1, & \text{if } x \leq 1 \end{cases}$$

In the context of a continuous random variable, we need to ask what is the probability of intervals or unions of intervals: $P(a \leq X \leq b)$

If $F_X(.)$ is differentiable, given that it is continuous then we can find another function for the continuous random variable.

$$f_X(x) = \frac{dF_X(x)}{dx}$$

$f_X(x)$ is called the probability density function (pdf).

$$f_X(x) = \begin{cases} 1, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

$$F_X(a) = P(X \leq a) = \int_{-\infty}^{a} f_X(x)dx$$

$$P(a \leq X \leq b) = F_X(b) - F_X(a) = \int_{a}^{b} f_X(x)dx$$

$$P(x \leq X \leq x + \Delta x) \to \Delta x \text{ is very small(infintesimal)}$$

$$P(x \leq X \leq x + \Delta x) \approx \text{area of rectangle with b=}\Delta x \text{ \& h= } f_X(x)$$

$$P(x \leq X \leq x + \Delta x) \approx f_X(x)\Delta x$$

9.1.1 Probability density function

When is a function a valid probability density function?

Two properties of the probability density function are:

- $f_X(x) \geq 0$

  As CDF is monotonically non decreasing, it's derivative is always non-negative.

- $\int_{-\infty}^{\infty} f_X(x)dx = 1 \to \quad P(X \leq \infty) = F_X(\infty) = 1$

Note: $f_X(.) \not\leq 1$

The probability density function itself doesn't indicate any probability. Only the area under the curve indicates probability.

9.2 Discrete random variable

X is said to be a discrete random variable if the range of X is either finite or countably infinite in $\mathbb{R}$.

$$X : S \to \mathbb{R}$$

CDF of a discrete random variable will be constant everywhere and jumps at some points(finite or countably infinite number of points).

Range of $X = x_1, x_2, \cdots$

Discrete random variables can take non-zero values unlike the continuous r.v's.

$$F_X(a) = \sum_{x_i \leq a} P(X = x_i)$$

$$P(X = x_i) = P_X(x_i) \qquad x_i \text{ is in the range of X.}$$

P_X is said to be the probability mass function.

9.2.1 Probability mass function

When is a function a valid probability mass function?

$$S = \text{ Range of X } = \{x_1, x_2, \cdots\}$$

Two properties of the probability mass function are:

- $P_X(x_i) \geq 0$
- $\sum_{x_i \in S} P_X(x_i) = 1$

9.3 Mixed random variable

$F_X(.)$ is continuous in some intervals and also jumps at some points. The probability density function has some impulses included in it. Let, X_1 be a continuous random variable, X_2 be a discrete random variable, then Y, a linear combination, is a mixed variable.

$$Y = \alpha X_1 + \beta X_2$$

9.4 Functions of random variables

X is a random variable. $X : S \to \mathbb{R}$.

$$Y = g(X) \qquad g : \mathbb{R} \to \mathbb{R} \qquad Y = goX : S \to \mathbb{R}$$

We require a condition on g to find out when Y is a random variable.

$$Y^{-1}((-\infty, y]) \in F \Rightarrow X^{-1}\left(g^{-1}((-\infty, y])\right) \in F$$

Let $\mathfrak{B}$ be an arbitrary set in the Borel σ-algebra generated by intervals of the form $(-\infty, x] = \mathfrak{B}$.

Then, since X is a random variable, $X^{-1}(\mathfrak{B}) \in F$.

$g^{-1}((-\infty, y]) \in \mathfrak{B}$, (Borel σ-algebra generated by intervals of the form $(-\infty, x]$).

$$X^{-1}\left(g^{-1}((-\infty, y])\right) \in F \Rightarrow \text{Y is a random variable}$$

Any common functions satisfy the condition needed for g.

- Suppose X is a discrete random variable, $X : S \to \mathbb{R}$, range of X is discrete.

  g is a function, such that $Y = g(X)$. Then Y is a discrete random variable.

  X has a probability mass function P_X, $Y \to P_Y$. Then,

  $$\begin{aligned} P_Y(y) &= P(Y = y) = P(g(X) = y) \\ &= P(\{x_i | g(x_i) = y\}) \\ &= \sum_{x_i : g(x_i) = y} P_X(x_i) \end{aligned}$$

  Eg: Tossing a coin till first head.

  $$S = \{H, TH, TTH, \cdots\}$$

  $$X : S \to \mathbb{R} \qquad H \to 1 \qquad TH \to 2 \qquad TTH \to 3 \cdots$$

  $Y = g(X) = X \bmod 4 \Rightarrow$ Range of Y $= \{0, 1, 2, 3\}$

  $$P_Y(0) = P(Y = 0) = P(X \bmod 4 = 0) = P(\{4, 8, \cdots\}) = \sum_{x \bmod 4 = 0} P_X(x)$$

- X is a continuous random variable, $Y = g(X)$.

  Based on g, it can either be continuous or discrete.

  1. When X is continuous & Y is also continuous.
     Let $F_X(.)$ be the CDF of X and $f_X(.)$ be the PDF of Y.

     $$Y = aX + b \qquad a, b \in \mathbb{R}$$

     For the CDF's:

     $$F_Y(y) = P(Y \leq y)$$

     $$\{Y \leq y\} = \{aX + b \leq Y\}$$

(a) If $a > 0$,

$$\{Y \leq y\} = \{X \leq \frac{y-b}{a}\}$$

$$F_Y(y) = P(Y \leq y) = P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right)$$

(b) If $a < 0$,

$$\{Y \leq y\} = \{X \geq \frac{y-b}{a}\}$$

$$F_Y(y) = P(Y \leq y) = P\left(X \geq \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right)$$

For the PDF's:

(a) If $a > 0$

$$F_Y(y) = F_X\left(\frac{y-b}{a}\right)$$

$$\begin{aligned} f_Y(y) = \frac{dF_Y(y)}{dy} &= \frac{d}{dy}F_X\left(\frac{y-b}{a}\right) \\ &= f_X\left(\frac{y-b}{a}\right)\frac{d}{dy}\left(\frac{y-b}{a}\right) \\ &= \frac{1}{a}f_X\left(\frac{y-b}{a}\right) \end{aligned}$$

(b) If $a > 0$

$$F_Y(y) = 1 - F_X\left(\frac{y-b}{a}\right)$$

$$f_Y(y) = \frac{-1}{a}f_X\left(\frac{y-b}{a}\right)$$

Hence, $\boxed{f_Y(y) = \frac{1}{|a|}f_X\left(\frac{y-b}{a}\right)}$

2. X is continuous and Y is continuous, but a many to one function.

$$Y = X^2$$

Now let us find the CDF & PDF of Y in therms of CDF & PDF of X.

(a) $y < 0$, $F_Y(y) = 0$

(b) $y \geq 0$,

$$\begin{aligned} F_Y(y) = &= P(Y \leq y) \\ &= P(X^2 \leq y) = P(-\sqrt{y} \leq x \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) + P(x = -\sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \frac{dF_Y(y)}{dy} = \frac{d}{dy}(F_X(\sqrt{y}) - F_X(-\sqrt{y})) \\ &= \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) - f_X(-\sqrt{y})] \end{aligned}$$

3. X is continuous, Y is continuous and a many to one function.

$$f_X(x) = \begin{cases} \frac{1}{2\pi}, & \text{if } 0 \leq x \leq 2\pi \\ 0, & \text{otherwise} \end{cases}$$

$$F_X(x) = \int_{-\infty}^{x} f_x(x)dx$$

$$Y = \sin X$$

Let us find the PDF of Y.

(a) Consider $0 \leq y \leq 1$

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(\sin X \leq Y) \\ &= P(\{0 \leq X \leq \sin^{-1} y\} \cup \{\pi - \sin^{-1} y \leq X \leq 2\pi\}) \\ &= P(0 \leq X \leq \sin^{-1} y) + (\pi - \sin^{-1} y \leq X \leq 2\pi) \\ &= F_X(\sin^{-1} y) - F_X(0) + F_X(2\pi) - F_X(\pi - \sin^{-1} y) \\ &= \frac{\sin^{-1} y}{2\pi} - 0 + 1 - \left(\frac{\pi - \sin^{-1} y}{2\pi}\right) \\ &= \frac{1}{2} + \frac{\sin^{-1} y}{\pi} \end{aligned}$$

(b) $-1 \leq y < 0$

$$\begin{aligned} F_y(y) &= P(Y \leq y) = P(\sin X \leq y) \\ &= P(\pi - \sin^{-1} y \leq X \leq 2\pi + \sin^{-1} y) \end{aligned}$$

4. X is continuous and Y is discrete.
X has CDF $F_X(.)$ & PDF $f_X(.)$
Y has CDF $F_Y(.)$ & PMF $P_Y(.)$

$$Y = g(x) = k \; if \; k \leq x < k+1 \qquad k \in \mathbb{Z}$$

The above operation is also called quantization. Range of g is countably infinite & thus Y is a discrete random variable.
We would like to experss the PMF of Y in terms of the PDF of X.

$$\begin{aligned} P_Y(y) &= P(Y = y) = P(y \leq X < y+1) \\ &= \int_y^{y+1} f_X(x)dx \end{aligned}$$

In general,

$$P_Y(y) = \int f_X(x)dx \qquad S = \{x : g(x) = y\}$$

9.5 General formula for determining PDF of $Y = g(X)$

Where g is differentiable,

$$P(y \leq Y \leq y + \Delta y) \approx f_Y(y)\Delta y$$

The left hand side can be written in terms of PDF of X.

$$g(x_i) = y$$

$$g^{-1}[y, y+\Delta y] = [x_1, x_1 + \Delta x_1] \cup [x_2, x_2 + \Delta x_2] \cdots [x_n, x_n + \Delta x_n]$$

What is the sign of slope of $g(x)$ at x_i?

- If $g^{'}(x_i) \geq 0 \qquad [x_i, x_i + \delta x_i]$
- If $g^{'}(x_i) < 0 \qquad [x_i - \delta x_i, x_i]$

$$\begin{aligned} f_Y(y)\Delta y &= \sum_{i=1}^{n} P(x_i \leq X \leq x_i + \Delta x_i) \\ &= \sum_{i=1}^{n} f_x(x_i)\Delta x_i \end{aligned}$$

$$f_Y(y) = \sum_{i=1}^{n} f_X(x_i)\frac{\Delta x_i}{\Delta y} = \sum_{i=1}^{n} f_X(x_i)\frac{1}{(\Delta y/\Delta x_i)}$$

For infintesimally small Δy & corresponding Δx_i:

$$\frac{\Delta y}{\Delta x_i} = |g^{'}(x_i)|$$

We take modulus as LHS is positive, i.e. magnitude of the slope of tangent at x_i.

$$f_Y(y) = \sum_{i=1}^{n} \frac{f_X(x_i)}{|g'(x_i)|}$$

Examples:

- $Y = X^2$

$$f_Y(y) = \frac{f_X(\sqrt{y})}{2\sqrt{y}} + \frac{f_X(-\sqrt{y})}{2\sqrt{y}}$$

- $Y = \sin X$, we take $\sin^{-1} y (2n\pi + \sin^{-1} y)$ & $\pi - \sin^{-1} y (2n\pi + \pi - \sin^{-1} y)$

$$f_Y(y) = \sum_n \left[\frac{f_X(2n\pi + \sin^{-1} y)}{\sqrt{1-y^2}} + \frac{f_X(2n\pi + \pi - \sin^{-1} y)}{\sqrt{1-y^2}} \right]$$

Theorem 9.1. *If $Y = F_X(X)$, where X itself is a random variable with CDF $F_X(.)$, the CDF of Y:*

$$F_Y(y) = P(Y \leq y) = P(F_X(x) \leq y)$$

- If $y < 0$,

$$\{X \leq F_X^{-1}((-\infty, y])\} = \phi \qquad P(F_X(x) \leq y) = 0$$

- If $y \geq 1$,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(F_X(x) \leq y) \\ &= 1 \qquad (\text{max value of } F_X \text{ is } 1,\ F_X^{-1}((-\infty, y]) = \mathbb{R}) \end{aligned}$$

- If $0 \leq y < 1$,

$$\begin{aligned} F_Y(y) = P(Y \leq y) &= P(F_X(x) \leq y) \\ &= P(X \leq F_X^{-1}(y)) \\ &= F_X(F_X^{-1}(y)) \\ &= y \end{aligned}$$

$$f_Y(y) = \begin{cases} 0, & \text{if } y < 0 \\ y, & \text{if } 0 \leq y < 1 \\ 1, & \text{if } y \geq 1 \end{cases}$$

If we take any random variable X & apply $F_X(.)$ as a function of X, then resulting random variable is a uniform random variable.

Suppose we start with an uniform random variable Y and consider a function of Y which is $X = F_X^{-1}(Y)$, then X has a CDF $F_X(.)$.

10 Expectation of a random variable

It is also called mean (or average).

For a discrete random variable X, let S be the range of X, which is either finite or countably infinte.

$$E(X) = \sum_{x_i \in S} x_i P_X(x_i)$$

where P_X is the PMF of random variable X.

The expectation, is based on the weighted average, where weights are the probabilities of the random variable taking a particular value unlike the general notion of average.

For a continuous random variable X,

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

The integral can be thought of as the limit of a summation. Hence,

$$E(X) = \lim_{\Delta x_i \to 0} \sum_{i=-\infty}^{\infty} x_i f_X(x_i) \Delta x_i$$

10.1 Expectation of a function of a random variable

$Y = g(X)$, X & Y are discrete.

$$E(Y) = \sum_{y_i} y_i P_Y(y_i)$$

We can express $E(Y)$ in terms of PMF of X, P_X,

$$\begin{aligned}
E(Y) &= \sum_{y_i} y_i P_Y(y_i) \\
&= \sum_{y_i} y_i \sum_{x_{i,j}: g(x_{i,j}) = y_i} P_X(x_{i,j}) \\
&= \sum_{y_i} \sum_{x_{i,j}: g(x_{i,j}) = y_i} y_i P_X(x_{i,j}) \\
&= \sum_{y_i} \sum_{x_{i,j}: g(x_{i,j}) = y_i} g(x_{i,j}) P_X(x_{i,j}) \\
&= \sum_{x_j} g(x_j) P_X(x_j) \\
\Rightarrow E(Y) &= \sum_{y_i} y_i P_Y(y_i) \\
&= \sum_{x_i} g(x_i) P_X(x_i)
\end{aligned}$$

For continuous random variable,

$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty} y f_Y(y) dy \\ &= \int_{-\infty}^{\infty} g(x) f_X(x) dx \end{aligned}$$

10.2 Variance of a random variable

It is defined as,

$$E[g(X)] \qquad g(X) = (X - E(X))^2$$

$E(X)$ indicates the average value of a random variable. So there is a variation in the values the random variable takes around mean, now we would like to measure this variation and hence the variance comes into play.

$(X - E(X))$ measures variation from the mean, and we square it so that the positive and negative values don't cancel each other.

If X is discrete:
$E(X) = \mu$
$E(X) = E(X - E(X)^2) = \sum_{x_i} (x_i - \mu)^2 P_X(x_i)$
If X is continuous:
$E(X) = \mu$
$E(X) = E(X - E(X)^2) = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx$

- $Var(X) \geq 0$

  $Var(X) = 0$ iff $X = E(X)$ with probability 1.

10.3 Examples of discrete random variables

1. Bernoulli random variable:

   Experiment: Coin toss experiment.

   $$X : S = \{H, T\} \qquad H \to 1 \qquad T to 0$$

   PMF of X, $P_X(1) = p,\ P_X(0) = 1 - p$

   $E(X) = 0(1 - p) + 1(p) = p.$

   $Var(X) = (0 - p)^2(1 - p) + (1 - p)^2 p = p(1 - p)$

2. Binomial random variable:

   Experiment: Toss a biased coin n times. X is defined as the number of heads in that string.

   $$S = \{HH \cdots H, TH \cdots H, \cdots\}$$

   $X : S \to R$

   Range of X: $\{0, 1, cdots, n\}$

PMF of X: $P_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$

$$
\begin{aligned}
E(X) &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\
&= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!(n-1-k)!} p^k (1-p)^{n-1-k} \\
&= np
\end{aligned}
$$

$Var(X) = np(1-p)$

3. Geometric random variable:

Experiment: Keep tossing a biased coin till we get a head.

$$S = \{H, TH, TTH, \cdots\} \qquad X : S \to R$$

$P(H) = p,\ P(T) = 1-p$
$P_X(k) = p(1-p)^{k-1}$

$$E(X) = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = \frac{1}{p}$$

$$
\begin{aligned}
Var(X) &= E((X - E(X)^2) \\
&= E(X^2 - 2XE(X) + (E(X))^2) \\
&= E(X^2) + E(-2XE(X)) + E((E(X))^2 \\
&= E(X^2) - 2E(X)E(X) + (E(X))^2 \\
&= E(X^2) - (E(X))^2
\end{aligned}
$$

We know, $Var(X) \geq 0 \Rightarrow E(X^2) \geq (E(X))^2$

Now, $Var(x)$ for a geometric random variable:

$$
\begin{aligned}
Var(X) &= E(X^2) - E(X)^2 \\
&= \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1} - \left(\frac{1}{p}\right)^2 \\
&= \frac{2-p}{p^2} - \frac{1}{p^2} \\
&= \frac{1-p}{p^2}
\end{aligned}
$$

4. Poisson random variable:

   Experiment: When we have to count rare events within a time frame.

$$X = \{0, 1, 2, \cdots\}$$

   PMF of Poisson RV:

$$P_X(k) = \frac{\lambda^k}{k!}e^{-\lambda}$$

   where λ is a parameter and is the average value.

   $Var(X) = \lambda$

$$\begin{aligned}\sum_{k=0}^{\infty} P_X(k) &= e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} \\ &= e^{-\lambda}e^{\lambda} \\ &= 1\end{aligned}$$

10.4 Examples of continuous random variables

1. Uniform random variable:

   Experiment: Measuring voltages, quantization error (original signal - quantized signal) is modelled as a uniform random variable.

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

   $E(X) = \frac{a+b}{2}$

   $Var(X) = \frac{(b-a)^2}{12}$

2. Exponential random variable:

   Experiment: Used to model completion time of a process.

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

$$E(X) = \int_0^{\infty} x\lambda e^{-\lambda x}dx = \frac{1}{x}$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{1}{\lambda^2}$$

3. Gaussian random variable:

   Experiment: Noise is generally modelled as a Gaussian random variable. We shall discuss this in detail while doing the central limit theorem.

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

   Mean of X: μ

   $Var(x) = \sigma^2$

11 Random vectors and joint CDF

Consider 2 individual random variables X_1 X_2, with (S, F, P) as the probability space. $X_1 : S \to \mathbb{R}$ $\quad X_2 : S \to \mathbb{R}$

$$(X_1, X_2) : S \to \mathbb{R}^2$$

Then (X_1, X_2) form a random vector and the joint CDF of X_1 and X_2:

$$F_{X_1,X_2}(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2)$$

$$X_1 \leq x_1 = \{w : X_1(w) \leq x_1\} = A_1 \in F$$

$$X_2 \leq x_2 = \{w : X_2(w) \leq x_2\} = A_2 \in F$$

$$(X_1 \leq x_1, X_2 \leq x_2) = \{\omega : X_1(\omega) \leq x_1 \,\&\, X_2(\omega) \leq x_2\} = A_1 \cap A_2 \in F$$

$$P(X_1 \leq x_1, X_2 \leq x_2) \;\forall\; (x_1, x_2) \in \mathbb{R}^2$$

By definition, $F_{X_1,X_2}(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2)$.
Then we can expand this definition till n, i.e.
$X_1, X_2, \cdots, X_n$ $\quad X_i S \to \mathbb{R}$
We can say that the joint CDF of n random variables:

$$F_{X_1,\cdots,X_n}(x_1, \cdots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \cdots X_n \leq x_n)$$

11.1 Properties of joint CDF

1. $F_{X,Y}(.,.)$ is monotonically non decreasing in x,y.

   We can compare them for $\{\{x_1 = x_2, y_1 \leq y_2\}, \{x_1 \leq x_2, y_1 = y_2\}, \{x_1 < x_2, y_1 < y_2\}\}$ but not for $\{x_1 > x_2, y_1 < y_2\}$ etc.

2.

$$\lim_{x\to-\infty, y\to-\infty} F_{X,Y}(x,y) = 0$$

$$\lim_{x\to-\infty} F_{X,Y}(x,y) = 0 \qquad \lim_{y\to-\infty} F_{X,Y}(x,y) = 0$$

3.

$$\lim_{x\to\infty, y\to\infty} F_{X,Y}(x,y) = 1$$

$$\lim_{x\to\infty} F_{X,Y}(x,y) = F_Y(y) \qquad \lim_{y\to\infty} F_{X,Y}(x,y) = F_X(x)$$

4. $F_{X,Y}(.,.)$ is right continuous individually in x & y.

$$\lim_{x\to x_0^+} F_{X,Y}(x,y) = F_X(x_0,y)$$

$$\lim_{(x,y)\to(x_0^+,y_0^+)} F_{X,Y}(x,y) = F_X(x_0,y_0)$$

Different cases possible for X & Y:

- X & Y both discrete.
- X & Y both continuous.
- X is discrete & Y is continuous.

1. X & Y both discrete.

   Joint PMF:

$$P_{X,Y}(x_i,y_i) = P(X = x_i, Y = y_i)$$

   Expressing joint CDF in terms of the joint PMF,

$$F_{X,Y}(x,y) = \sum_{x_i\le x}\sum_{y_i\le y} P_{X,Y}(x_i,y_i)$$

   $X \to P_X \qquad Y \to P_Y \qquad (X,Y) \to P_{X,Y}$

   We can form a relation between $P_{X,Y}$ and P_X,

$$P_X(x_i)\sum_{y_i} P_{X,Y}(x_i,y_i)$$

$$P(X = x_i)\sum_{y_i} P_{X,Y}(X = x_i, Y = y_i)$$

   In the context of the joint PMF, P_X & P_Y are also called as marginal PMF's. And the process of summing over the range of the other random variable to get the marginal PMF, is known as marginalization.

11.2 Conditional PMF

Let us recall conditional probabilities (in the probability space (S, F, P)),

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \qquad P(B) > 0$$

In the context of random variables, how do we define conditional probabilities?

Given an event $B \in F$, $P(B) > 0$. Conditional CDF conditioned on the event B would be:

$$F_{X/B}(x/B) = \frac{P((X \leq x) \cap B)}{P(B)}$$

$X \leq x = \{w : X(w) \leq x\} = A \in F$

If B is of the form $\{Y \leq y\}$

$$F_{X/B}(x/B) = \frac{P((X \leq x) \cap B)}{P(B)}$$

$$F_{X/\{Y \leq y\}}(x/\{Y \leq y\}) = \frac{P(\{X \leq x\} \cap \{Y \leq y\})}{P(\{Y \leq y\})} = \frac{F_{X,Y}(x,y)}{F_Y(y)}$$

We shall denote this, conditional PMF, by:

$$F_{X/Y}(x/y) = \frac{F_{X,Y}(x,y)}{F_Y(y)}$$

If both X & Y are discrete, for defining conditional PMF, we will not condition on $\{Y \leq y_i\}$ but instead condition on $\{Y = y_i\}$

$$P_{X/\{Y=y_i\}}(x_i/\{Y = y_i\}) = \frac{P((X = x_i) \cap (Y = y_i))}{P(Y = y_i)}$$

Conditional PMF:

$$P_{X/Y}(x_i/y_i) = \frac{P_{X,Y}(x_i, y_i)}{P_Y(y_i)} = \frac{\text{Joint PMF}}{\text{Marginal PMF}}$$

If X & Y are both continuous random variables, we can define joint PDF of X & Y as

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$$

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(x,y)dxdy$$

Marginal PDF's of X & Y can be expressed in terms of joint PDF's as:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$

A joint PDF is said to be valid if,

$$f_{X,Y}(x,y) \geq 0 \quad \& \quad \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(x,y)\partial x \partial y = 1$$

When both X & Y are continuous, $P(X = x) = 0,\ P(Y = y) = 0$ i.e. probability that the random variable takes a single value, finite or countably infinite set of values is zero.

Conditional PDF can be expressed in terms of joint PDF's and marginal PDF's,

$$f_{X/Y}(x/y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

$$f_{X/Y}(x/y) = \lim_{\Delta x \to 0, \Delta y \to 0} \frac{P(x < X \leq x + \Delta x / y < Y \leq y + \Delta y)}{\Delta x}$$

$$f_{X/B}(x/B) = \lim_{\Delta x \to 0} \frac{P(x < X \leq x + \Delta x / B)}{\Delta x}$$

If $B = \{y < Y \leq y + \Delta y\}$, and let $\Delta y \to 0$, $P(B) > 0$.

$$\Rightarrow f_{X/Y}(x/y) = \lim_{\Delta x \to 0, \Delta y \to 0} \frac{P(x < X \leq x + \Delta x, y < Y \leq y + \Delta y)}{P(y < Y \leq y + \Delta y)\Delta x}$$

$$= \lim_{\Delta x \to 0, \Delta y \to 0} \quad =$$

11.3 Bayes rule for discrete random variable

$$P_{Y/X}(y_i/x_i) = \frac{P_{X/Y}(x_i/y_i)P_Y(y_i)}{P_X(x_i)}$$

$$= \frac{P_{X/Y}(x_i/y_i)P_Y(y_i)}{\sum_{y_i} P_{X/Y}(x_i/y_i)P_Y(y_i)}$$

11.4 Bayes rule for continuous random variable

$$P(x < X \leq x+\Delta x / y < Y \leq y+\Delta y) = \frac{P(x < X \leq x + \Delta x, y < Y \leq y + \Delta y)P(x < X \leq x + \Delta x)}{P(y < Y \leq y + \Delta y)}$$

$$f_{X/Y}(x/y)\Delta x = \frac{f_{Y/X}(y/x)\Delta y f_X(x)\Delta x}{f_Y(y)\Delta y}$$

$$\Rightarrow f_{X/Y}(x/y) = \frac{f_{Y/X}(y/x)f_X(x)}{f_Y(y)}$$

11.5 Bayes rule when X is discrete and Y is continuous

If X is discrete and Y is continuous, then we can't talk about either joint PMF or PDF. We can talk about the joint CDF,

$$P_{X/Y}(x/y) = \frac{f_{Y/X}(y/x)P_X(x)}{f_Y(y)}$$

$$\begin{aligned}
P_{X/Y}(x/y) &= \lim_{\Delta y \to 0} P(X = x/y < Y \leq y + \Delta y) \\
&= \lim_{\Delta y \to 0} \frac{P(X = x, y < Y \leq y + \Delta y)}{P(y < Y \leq y + \Delta y)} \\
&= \lim_{\Delta y \to 0} \frac{P(y < Y \leq y + \Delta y/X = x)P(X = x)}{P(y < Y \leq y + \Delta y)} \\
&= \lim_{\Delta y \to 0} \frac{f_{Y/X}(y/x)\Delta y P(X = x)}{f_Y(y)\Delta y} \\
&= \frac{f_{Y/X}(y/x)P_X(x)}{f_Y(y)}
\end{aligned}$$

12 Independent random variables

- Discrete case: Two random variables are said to be indepedent if

  $$P_{X,Y}(x,y) = P_X(x)P_Y(y)$$

  i.e. joint PMF is the product of the marginal PMF's.

  OR,

  $$P_{X/Y}(x/y) = P_X(x)$$

- Continuous case: Two random variables are sid to be indepedent if

  $$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

  i.e. joint PDF is the product of the marginal PDF's.

  OR,

  $$f_{X/Y}(x/y) = f_X(x)$$

If both X & Y are discrete and indepedent, show that $F_{X,Y}(x,y) = F_X(x)F_Y(y)$.

Proof. For the discrete case,

$$\begin{aligned}
F_{X,Y}(x,y) &= P(X \leq x, Y \leq y) \\
&= \sum_{x_i \leq x} \sum_{y_i \leq y} P_{X,Y}(x_i, y_i) \\
&= \sum_{x_i \leq x} \sum_{y_i \leq y} P_X(x_i) P_Y(y_i) \\
&= \left[\sum_{x_i \leq x} P_X(x_i) \right] \left[\sum_{y_i \leq y} P_y(y_i) \right] \\
&= F_X(x) F_Y(y)
\end{aligned}$$

Similarly, using the same steps it can be proved for the continuous case too. $\square$