Mathematics 3 (SM 211): Probability and Statistics

Amit Chattopadhyay

IIIT-Bangalore

Ch. 3: Probability Distribution - I

Syllabus: Outline

Probability:

1. The Concept of Probability
2. Compound or Joint Experiment
3. Probability Distributions-I
4. Mathematical Expectation-I
5. Probability Distributions-II
6. Mathematical Expectation-II
7. Some Important Continuous Univariate Distributions
8. Convergence of a Sequence of Random Variables and Limit Theorems

Statistics:

1. Random Samples
2. Sampling Distributions
3. Estimation of Parameters
4. Testing of Hypothesis
5. Regression

Reference Books

1. Mathematical Probability by A. Banerjee, S.K. De and S. Sen
2. Mathematical Statistics by S.K. De and S. Sen
3. Groundwork of Mathematical Probability and Statistics by Amritava Gupta
4. Introduction to Probability and Statistics for Engineers and Scientists by S.M. Ross
5. Introduction to Probability Models, by S.M. Ross
6. Probability and Statistics, (Schaum's Outlines) by Murray R Spiegel, John J Schiller and R Alu Srinivasan

Probability Distribution - I

Objective

- Random Variable
- Cumulative Distribution Function
- Probability Mass/Density Function
- Transformation of Random Variables
- Introduction to Poisson Process

Random Variable

$A \xleftrightarrow{X}$ a b $\quad f : \mathbb{R}^2 \to \mathbb{R}$

$\Delta \longleftrightarrow \mathbb{R}$

Probability Space: (S, Δ, P)

Motivation:

- Developing *mathematical theory of probability* using the probability function $P : \Delta \to \mathbb{R}$ is difficult because its domain Δ consists of collection of events
- An event corresponding to a random experiment consists of objects which are not necessarily real numbers
- **Way out:**
  - **Random Variable:** Finds a correspondence between events in Δ with real numbers
  - **Distribution Function:** Replaces probability function by a real-valued function defined on real line

$F_X : \mathbb{R} \to \mathbb{R}$

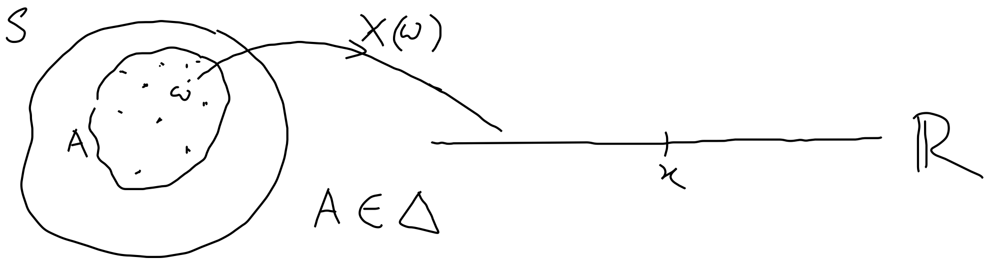
Definition

A real-valued function $X : S \to \mathbb{R}$ is called a *random variable* if for any $x \in \mathbb{R}$,

$$\{\omega \in S : -\infty < X(\omega) \le x\} \in \Delta.$$

(i.e. is an event)

- **Spectrum of X:** The range of the random variable X
- X is said to be *discrete* or *continuous* if the spectrum of X is discrete or continuous, respectively.



Random Variable: Examples

E : A coin is tossed twice

$S = \{\omega_1 = (H,H), \omega_2 = (H,T), \omega_3 = (T,H), \omega_4 = (T,T)\}$

1. Define $X_1 : S \to \mathbb{R}$ as

$$X_1(\omega_i) = \text{ number of heads}$$

Spectrum of $X = \{0, 1, 2\}$

2. Define $X_2 : S \to \mathbb{R}$ as

$$X_2(\omega_i) = \text{ square of number of heads}$$

Spectrum of $X = \{0, 1, 4\}$

3. Define $X_3 : S \to \mathbb{R}$ as

$$X_3(\omega_i) = \text{ number of heads } - \text{ number of tails}$$

Spectrum of $X = \{-2, \cancel{-1}, 0, \cancel{1}, 2\}$

Random Variable: Examples

Q: What are the events $\{\omega : X_i(\omega) \leq 1\}$, for $i = 1, 2, 3$?

$$\{\omega : X_1(\omega) \leq 1\} = \{\omega_2, \omega_3, \omega_4\}$$

$$\{\omega : X_2(\omega) \leq 1\} = \{\omega_2, \omega_3, \omega_4\}$$

$$\{\omega : X_3(\omega) \leq 1\} = \{\omega_2, \omega_3, \omega_4\}$$

Notations: Expressing Events using Random Variable

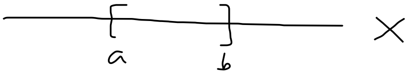
- By properties of σ-algebra the following are events:

$A = \{\omega \in S : a \leq X(\omega) \leq b\}$

$B = \{\omega \in S : X(\omega) = a\}$

etc. (HW)

$A^c = (-\infty < X < a) \cup (b < X < \infty)$



$(a - \frac{1}{n} < X \leq a) \in \Delta \quad \forall n \in \mathbb{N}$

$\bigcap_{n=1}^{\infty} (a - \frac{1}{n} < X \leq a) \in \Delta$

$= (X = a) \in \Delta$

- For brevity, we use following notations:

$\{\omega \in S : a < X(\omega) \leq b\} \equiv (a < X \leq b)$

$\{\omega \in S : -\infty \leq X(\omega) \leq b\} \equiv (X \leq b)$

$\{\omega \in S : X(\omega) = a\} \equiv (X = a)$

$(-\infty < x \leq b)$

If $A_1, A_2, \dots, A_n \in \Delta$ then $\bigcap_{i=1}^{n} A_i \in \Delta$

Pf: $\left(\bigcap_{i=1}^{n} A_i\right)^c = \left(\bigcup_{i=1}^{n} A_i^c\right)$

$(-\infty < X \leq a) \in \Delta$, $(-\infty < X \leq b) \in \Delta$

$\Rightarrow (a < X < \infty) \in \Delta$, $(-\infty < X \leq b) \in \Delta$

$\Rightarrow (a < X \leq b) \in \Delta$ (intersection $\in \Delta$)

$\Rightarrow (a \leq X \leq b) \in \Delta$ (taking union)

Cumulative Distribution Function

Definition

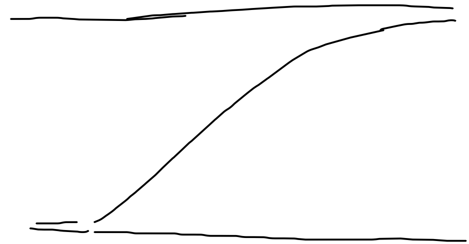
Let (S, Δ, P) be a probability space with the probability function $P : S \to \mathbb{R}$ and let $X : S \to \mathbb{R}$ be a random variable. Then the (Cumulative) distribution function of X is a real-valued function $F_X : \mathbb{R} \to \mathbb{R}$ defined as:

$$F_X(x) = P(\underbrace{-\infty < X \leq x}_{\{\omega \in S : -\infty < X(\omega) \leq x\}}), \ \forall x \in (-\infty, \infty).$$

Distribution Function: Properties

1. $0 \leq F_X(x) \leq 1$ (bounded function)

since $F_X(x) = P\left(-\infty < X \leq x\right)$



1. $0 \leq F_X(x) \leq 1$

2. F_X is monotonic increasing.

To prove:
if $x_1, x_2 \in \mathbb{R}$, $x_1 \leq x_2$ then $F_X(x_1) \leq F_X(x_2)$

$$F_X(x_1) \equiv P(-\infty < X \leq x_1) \qquad F_X(x_2) \equiv P(-\infty < X \leq x_2)$$

$$(-\infty < X \leq x_2) = (-\infty < X \leq x_1) \cup (x_1 < X \leq x_2)$$

(check $(x_1 < X \leq x_2)$ is an event)

(are mutually exclusive set of events)

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

[Axiom (iii) in Def^n]

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

$$\Rightarrow \boxed{P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)} \; (*) \geq 0 \Rightarrow F_X(x_2) \geq F_X(x_1).$$

1. $0 \le F_X(x) \le 1$
2. F_X is monotonic increasing.
3. F_X is continuous from right at every point, i.e.

   $F_X(a+0) = F_X(a)\ \forall a \in \mathbb{R}$.

Continuity Th of Probability

$A_1 \supseteq A_2 \supseteq \ldots$ $\Rightarrow$ $\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n$

$A_1 \subseteq A_2 \subseteq \ldots$ $\Rightarrow$ $\lim A_n = \sum A_n$

$$\lim_{n\to\infty} P(A_n) = P\left(\lim_{n\to\infty} A_n\right)$$

Pf: Let, $A_n = \left(a < X \le a + \frac{1}{n}\right),\ \forall n \in \mathbb{N}$

$$\left.\begin{array}{l} A_1 = (a < X \le a+1) \\ A_2 = \left(a < X \le a + \frac{1}{2}\right) \\ \vdots \end{array}\right\} \Rightarrow A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots$$

$$\boxed{\text{Claim: } \lim_{n\to\infty} A_n = \phi}$$

$$\lim_{n\to\infty} P(A_n) = P\left(\lim_{n\to\infty} A_n\right) = P(\phi) = 0$$

$$\Rightarrow \lim_{n\to\infty} P\left(a < X \le a + \tfrac{1}{n}\right) = 0 \Rightarrow \lim_{n\to\infty} F_X\left(a + \tfrac{1}{n}\right) = F_X(a)$$

$\circledast$ since F_X is mon. & bdd

$F_X(a+0) = F_X(a)$.

Claim: To show $\bigcap_{n=1}^{\infty} A_n = \phi$

If possible let, $\omega \in \bigcap_{n=1}^{\infty} A_n$ for some $\omega \in S$

$$\Rightarrow \omega \in A_n \quad \forall n \in \mathbb{N}$$

$$\Rightarrow a < \underbrace{X(\omega)}_{= b} \le a + \frac{1}{n} \quad \forall n \in \mathbb{N} \qquad (1)$$

Then for $b-a\ (>0)$ & 1, using Archimedean prop, $\exists\, m \in \mathbb{N}$ s.t. $(b-a)m > 1$.

$\Rightarrow b > a + \frac{1}{m}$ which is contradicting (1)

$$\Rightarrow \bigcap_{n=1}^{\infty} A_n = \phi$$

Distribution Function: Properties

1. $0 \leq F_X(x) \leq 1$
2. F_X is monotonic increasing.
3. F_X is continuous from right at every point, i.e.

   $F_X(a+0) = F_X(a) \ \forall a \in \mathbb{R}.$
4. $F_X(a) - F_X(a-0) = P(X = a) \ \forall a \in \mathbb{R}$ (HW)

$A_n = \left(a - \frac{1}{n} < X \leq a\right)$ ~~claim~~ $\lim_{n \to \infty} A_n = (X = a)$

Distribution Function: Properties

1. $0 \leq F_X(x) \leq 1$
2. F_X is monotonic increasing.
3. F_X is continuous from right at every point, i.e.
   $F_X(a+0) = F_X(a) \ \forall a \in \mathbb{R}$.
4. $F_X(a) - F_X(a-0) = P(X = a) \ \forall a \in \mathbb{R}$
5. $F_X(\infty) = 1$

$F_X(\infty) = P(-\infty < X < \infty) = P(S) = 1$, (informal way)

<u>Formal</u> $A_n = (-\infty < X \leq n)$

Distribution Function: Properties

1. $0 \leq F_X(x) \leq 1$
2. F_X is monotonic increasing.
3. F_X is continuous from right at every point, i.e.

   $F_X(a+0) = F_X(a) \ \forall a \in \mathbb{R}$.
4. $F_X(a) - F_X(a-0) = P(X = a) \ \forall a \in \mathbb{R}$
5. $F_X(\infty) = 1$
6. $F_X(-\infty) = 0$

informally $F_X(-\infty) = P(-\infty < X < -\infty) = P(\phi) = 0$

$A_n = (-\infty < X \leq -n) \downarrow \qquad \lim A_n = \phi$

Distribution Function: Properties

1. $0 \leq F_X(x) \leq 1$
2. F_X is monotonic increasing.
3. F_X is continuous from right at every point, i.e.
   $F_X(a+0) = F_X(a) \;\; \forall a \in \mathbb{R}$.
4. $F_X(a) - F_X(a-0) = P(X = a) \;\; \forall a \in \mathbb{R}$
5. $F_X(\infty) = 1$
6. $F_X(-\infty) = 0$
7. The set of points of discontinuity of F_X is at most enumerable.

Distribution Function: Remark

Converse: True

Any function F with domain $(-\infty, \infty)$ and range $[0, 1]$ is a distribution function of a random variable with respect to a probability space (S, Δ, P) iff properties 1-7 are satisfied.

Note: The curve $y = F_X(x)$ is called the *distribution curve* corresponding to the random variable X.

Converse: True

Any function F with domain $(-\infty, \infty)$ and range $[0, 1]$ is a distribution function of a random variable with respect to a probability space (S, Δ, P) iff properties 1-7 are satisfied.

Note: The curve $y = F_X(x)$ is called the *distribution curve* corresponding to the random variable X.

1. The spectrum of the random variable X consists of the points $\{1, 2, \ldots, n\}$ and $P(X = i) \propto \frac{1}{i(i+1)}$. (i) Determine the distribution function of X and (ii) compute $P(3 < X \leq n)$ and $P(X > 5)$.

Sol. Let, $P(X=i) = \frac{k}{i(i+1)}$

$$\sum_{i=1}^{n} P(X=i) = 1 \Rightarrow \sum_{i=1}^{n} \frac{k}{i(i+1)} = 1$$

$$\Rightarrow k\sum_{i=1}^{n}\left(\frac{1}{i} - \frac{1}{i+1}\right) = 1$$

$$\Rightarrow k\left(1 - \frac{1}{n+1}\right) = 1$$

$$\Rightarrow k = \frac{n+1}{n}.$$

$$P(X=i) = \frac{n+1}{n}\frac{1}{i(i+1)}.$$

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

If $x < 1$

$F_X(x) = 0$.

If $r \leq x < r+1$ for $r = 1, 2, \ldots, n-1$

$$F_X(x) = \sum_{i=1}^{r} \frac{n+1}{n} \cdot \frac{1}{i(i+1)}$$

If $x \geq n$

$F_X(x) = 1$

$$P(3<X\le n) = 1 - P(X=1) - P(X=2) - P(X=3)$$

$$\|$$

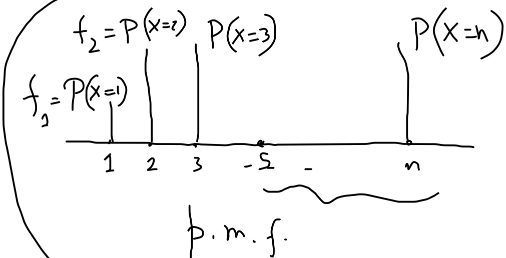
$$1 - P\left((3<X\le n)^c\right) = \dots \qquad \text{(check!)}$$

$$P(X>5) = 1 - P(X\le 5).$$

$$= 1 - \sum_{i=1}^{5} \frac{n+1}{n}\cdot\frac{1}{i(i+1)}.$$

$$\|$$

$$P(5<X\le n)$$



spectrum of $X = \{1, 2, 3, \dots, n\}$

$$(3<X\le n) = (X=4) \cup (X=5) \cup \dots$$

$$(3<X\le n)^c = (X=1) \cup (X=2) \cup (X=3)$$

2. A number is chosen at random from each of the two sets $\{0, 1, 2, 3\}$ and $\{0, 1, 2, 3\}$. Find the probability distribution of the random variable denoting the sum of the numbers chosen.

Sol. $E \to$ random experiment of choosing the numbers from the sets.

S = sample space corr. E

$= \{(x, y) : x, y = 0, 1, 2, 3\}$

$|S| = 16$.

$P(X=0) = \frac{1}{16}$

$P(X=1) = \frac{2}{16}$

$\vdots$

r.v. : $X : S \to \mathbb{R}$: $X(x,y) = x + y \quad \forall (x,y) \in S$

Spectrum of $X = \{0, 1, 2, 3, 4, 5, 6\}$

3. Find the probability distribution of the number of failures preceding the first success in an infinite sequence of Bernoulli trials with probability of success p.

Sol: $E \rightarrow$ infinite seq of Bernoulli trials

$(s, f, f, s, \ldots)$: sample points

r.v. $X : S \longrightarrow \mathbb{R}$ $\quad X(\omega)$ = no. of failures preceding the first success

spectrum of $X = \{0, 1, 2, \ldots\}$

$P(X = i) = (1-p)^i p \quad \forall i = 0, 1, 2, \ldots$

(Geometric Distribution)

Check! $\sum_{i=0}^{\infty} P(X=i) = 1$