Basic Concepts of Probability

Definitions

Statistical Experiment : Any procedure that produces data or observations
Sample Space (S) : The set of all possible outcomes of a statistical experiment
Sample Point : An outcome (element) in the sample space
Event : Subset of the sample space

Note : The same experiment can have different sample spaces and outcomes for the same event

E.g. Consider the experiment of rolling a die

(i) If the problem of interest is "the number that shows on the top face", then

- Sample space: $S = \{1, 2, 3, 4, 5, 6\}$.
- Sample point: 1 or 2 or 3 or 4 or 5 or 6.

(ii) If the problem of interest is "whether the number is even or odd", then

- Sample space: $S = \{\text{even}, \text{odd}\}$.
- Sample point: "even" or "odd".

Event where an odd number occurs : (i) $\{1, 3, 5\}$
(ii) $\{\text{odd}\}$

Event Operations :

1. Union $A \cup B$
2. Intersection $A \cap B$
3. Complement A'

Event Relationships :

1. Contained $A \subset B$
2. Equivalent $A = B$
3. Mutually Exclusive $A \cap B = \emptyset$
4. Independence $P(A \cap B) = P(A)P(B)$
OR equivalently
$P(A \mid B) = P(A)$

Event Laws & Properties

MORE EVENT OPERATIONS

(a) $A \cap A' = \emptyset$ (b) $A \cap \emptyset = \emptyset$

(c) $A \cup A' = S$ (d) $(A')' = A$

(e) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

(f) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

(g) $A \cup B = A \cup (B \cap A')$

(h) $A = (A \cap B) \cup (A \cap B')$

DE MORGAN'S LAW

For any n events $A_1, A_2, \ldots, A_n$,

(i) $(A_1 \cup A_2 \cup \ldots \cup A_n)' = A_1' \cap A_2' \cap \ldots \cap A_n'$.
A special case: $(A \cup B)' = A' \cap B'$.

(j) $(A_1 \cap A_2 \cap \ldots \cap A_n)' = A_1' \cup A_2' \cup \ldots \cup A_n'$.
A special case: $(A \cap B)' = A' \cup B'$.

Independence Properties :

(a) Suppose $P(A) > 0$, $P(B) > 0$. If $A \perp B$, then A and B are not mutually exclusive.

(b) Suppose $P(A) > 0$, $P(B) > 0$. If A and B are mutually exclusive, then $A \not\perp B$.

(c) S and $\emptyset$ are independent of any other event.

(d) If $A \perp B$, then $A \perp B'$, $A' \perp B$, and $A' \perp B'$.

Counting Methods

If the experiments are not independent, then we use the Multiplication Principle otherwise, we use the Addition Principle

How many **even three-digit numbers** can be formed from the digits 1, 2, 5, 6 and 9 if there is **no restriction** on how many times a digit can be used?

Even 3-digit numbers : can only choose 2 or 6 for the 'Ones' place

$\therefore$ Hundred's place : 5 choices - 1, 2, 5, 6, 9
Ten's place : 5 choices - 1, 2, 5, 6, 9
One's place : 2 choices - 2, 6

No restriction on how many times a digit can be used

$\therefore$ we can form $5 \times 5 \times 2 = 50$ possibilities

Consider the digits 0, 1, 2, 3, 4, 5 and 6. If each digit can be used **at most once**, how many 3-digit numbers greater than 420 can be formed?

3-digit : 0 cannot be in the hundred's place
Each digit can be used at most once.

Cases :

1. 4 is in the hundreds place
   1.1 2 is in the tens place
   → 4 choices for ones place - 1, 3, 5, 6
   → 4 possibilities
   1.2 3, 5, 6 is in the hundreds place (3 choices)
   → 5 choices for ones place - 0, 1, 2, Two of (3, 5, 6)
   → $3 \times 5 = 15$ possibilities
2. 5 or 6 is in the hundreds place (2 choices)
   → 6 choices for tens place - 0, 1, 2, 3, 4, One of (5, 6)
   → 5 choices for ones place
   → $6 \times 5 \times 2 = 60$ possibilities

$\therefore$ Total number of possibilities $= 4 + 15 + 60$
$= 79$

When Counting, we also have to take into account whether the order of the samples selected matters.
That is, whether we are talking about Permutation or Combination

Permutation
A **permutation** is a selection and arrangement of r objects out of n. In this case, *order* is taken into consideration.

The number of ways to choose and arrange r objects out of n, where $r \leq n$, is denoted by P_r^n, where

$$P_r^n = \frac{n!}{(n-r)!} = n(n-1)(n-2)\ldots(n-(r-1)).$$

Combination
A **combination** is a selection of r objects out of n, *without regard to the order.*

The number of combinations of choosing r objects out of n, denoted by C_r^n or $\binom{n}{r}$, is given by as

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} = \binom{n}{n-r}$$

Note : when $r = n$, ${}^nP_n = n!$

Probability

INTERPRETATION OF PROBABILITY: RELATIVE FREQUENCY

Suppose that we repeat an experiment E for a total of n times.

Let n_A be the number of times that the event A occurs.

Then $f_A = n_A/n$ is called the **relative frequency** of the event A in the n repetitions of E.

Clearly, f_A may not equal to $P(A)$ exactly. However when n grows large, we expect f_A to be close to $P(A)$; in the sense that $f_A \approx P(A)$. Or mathematically,

$$f_A \to P(A), \quad \text{as } n \to \infty.$$

Thus f_A "mimics" $P(A)$, and has the following properties:

(a) $0 \leq f_A \leq 1$.

(b) $f_A = 1$ if A occurs in every repetition.

(c) If A and B are mutually exclusive events, $f_{A\cup B} = f_A + f_B$.

*

FINITE SAMPLE SPACE WITH EQUALLY LIKELY OUTCOMES

Consider a sample space $S = \{a_1, a_2, \ldots, a_k\}$.

Assume that all outcomes in the sample space are **equally likely** to occur, i.e.,

$$P(a_1) = P(a_2) = \cdots = P(a_k).$$

Then for any event $A \subset S$,

$$P(A) = \frac{\text{number of sample points in } A}{\text{number of sample points in } S}.$$

Equivalently,

For any event A, $0 \leq P(A) \leq 1$

For sample space S, $P(S) = 1$

For any two mutually exclusive events A and B, $P(A \cup B) = P(A) + P(B)$

Properties :

- $P(\emptyset) = 0$
- If $A_1, A_2, \ldots, A_n$ are mutually exclusive events, then
$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$$
- $P(A') = 1 - P(A)$
- $P(A) = P(A \cap B) + P(A \cap B')$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- If $A \subset B$, then $P(A) \leq P(B)$

Conditional Probability : $P(B \mid A) = \dfrac{P(A \cap B)}{P(A)}$, interpret as the sample space being the event A

$\Rightarrow$ The probability that B occurs given that A occurred.

We also have the Inverse Probability Formula : $P(A \mid B) = \dfrac{P(A)P(B \mid A)}{P(B)}$

Random Variables

Let S be the sample space of an experiment. A function X, which assigns a real number to every $s \in S$ is called a **Random Variable**. That is,

$$X : S \mapsto \mathbb{R}$$, a mapping from the Sample Space to the real line

Typically, upper case letters are used to denote R.V., with lower case letters used to denote their observed values

The **Range Space** of X is the set of real numbers

$$R_x = \{ x \mid x = X(s),\ s \in S \}$$

There are two main types of R.V.s:

1. Discrete
   - ⇒ Number of values in R_x is finite or countable

Note: $R_x = \{1, 2, 3, \ldots\}$ is Discrete R.V.

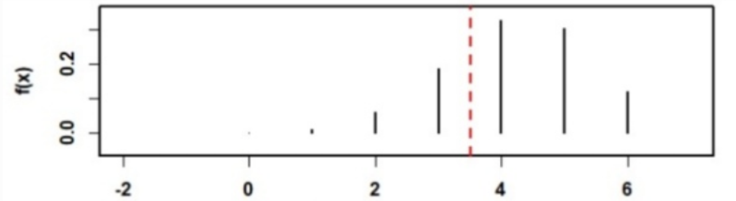
2. Continuous
   - ⇒ Interval, or a Collection of Intervals.
   - ⇒ Uncountable

In ST2334, we do not consider hybrid types

* If the number of possible values of $f(x)$ is finite, then x must be a Discrete R.V.

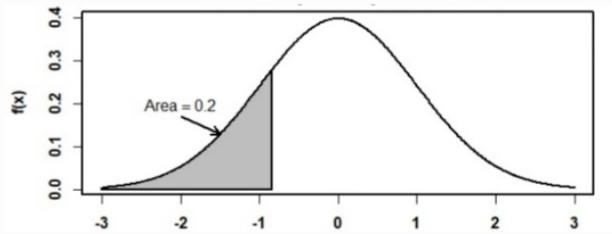
False.

Probability Distributions



Probability Mass Function (PMF)

- Discrete R.V
- $$f(x) = \begin{cases} P(X = x) & , \text{ for } x \in R_x \\ 0 & , \text{ for } x \notin R_x \end{cases}$$
  - Must satisfy the following : * To check if $f(x)$ is a valid PMF
    1. $f(x_i) \geq 0$ for all $x_i \in R_x$
    2. $f(x) = 0$ for all $x \notin R_x$
    3. $\sum_{x_i \in R_x} f(x_i) = 1$
  - For any set $B \subset \mathbb{R}$, we have
    $$P(X \in B) = \sum_{x_i \in B \cap R_x} f(x_i)$$



Probability Density Function (PDF)

- Continuous R.V
- Must satisfy the following : * To check if $f(x)$ is a valid PDF
  1. $f(x) \geq 0$ for all $x \in R_x$
  2. $f(x) = 0$ for $x \notin R_x$
  3. $\int_{R_x} f(x)\, dx = 1$
- For any a and b s.t. $a \leq b$,
  $$P(a \leq X \leq b) = \int_a^b f(x)\, dx$$

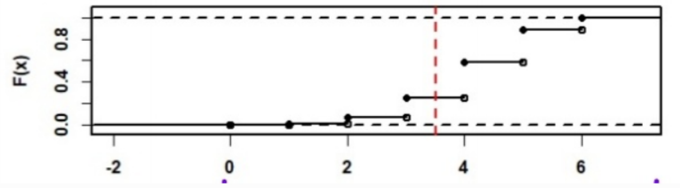
* For any specific value, x_0, $P(X = x_0) = \int_{x_0}^{x_0} f(x)\, dx = 0$

Cumulative Distribution Function

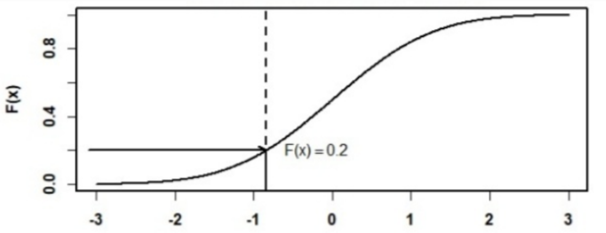
For any R.V. X, its Cumulative Distribution Function is defined by :

$$F(x) = P(X \leq x)$$

* Note : $0 \leq F(x) \leq 1$



Discrete R.V.

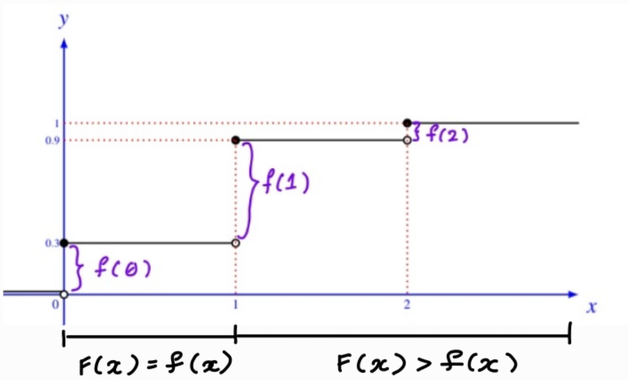


Continuous R.V.

To verify if a CDF is valid (regardless of Discrete or Continuous) :

1. $0 \leq F(x) \leq 1$
2. $F(x)$ is non-decreasing
3. $F(x)$ is right-continuous

Discrete R.V.



- $F(x) = \sum_{t \in R_x ; t \leq x} f(t) = \sum_{t \in R_x ; t \leq x} P(X = t)$
- The jump in height from 1 point to the other = probability at that point
- $P(X \leq b) = F(b)$
- $P(X < b) = F(b-)$

Continuous R.V.

- $F(x) = \int_{-\infty}^{x} f(t)\, dt$
- $P(a \leq X \leq b) = P(a < X < b) = F(b) - F(a)$
- $F(a-) = F(a_0)$, where a_0 is the largest $x < a$

Expectation and Variance

Let X be a discrete R.V. with $R_x = \{x_1, x_2, x_3, \dots\}$,
or a continuous R.V with p.f. $f(x)$.

Then, the expectation or mean of X is defined by : (Also denoted as μ_x)

Discrete : $E(X) = \sum_{x_i \in R_x} x_i f(x_i)$

Continuous : $E(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{x \in R_x} x f(x)\, dx$

* μ_x is not necessarily a possible value of the R.V. X

PROPERTIES OF EXPECTATION

(1) Let X be a random variable, and let a and b be any real numbers. Then

$$E(aX+b) = aE(X)+b.$$

(2) Let X and Y be two random variables. We have

$$E(X+Y) = E(X)+E(Y).$$

(3) Let $g(\cdot)$ be an arbitrary function.

- If X is a **discrete** random variable with probability mass function $f(x)$ and range R_X,
$$E[g(X)] = \sum_{x \in R_X} g(x) f(x).$$
- If X is a **continuous** random variable with probability density function $f(x)$ and range R_X,
$$E[g(X)] = \int_{R_X} g(x) f(x)\,\mathrm{d}x.$$

Let X be a R.V. The variance of X is defined as :

$\sigma_X^2 = V(X) = E(X - \mu_x)^2 \leftarrow$ Let $g(x) = (X - \mu_x)^2$

- If X is a **discrete** random variable with probability mass function $f(x)$ and range R_X,
$$V(X) = \sum_{x \in R_X} (x-\mu_X)^2 f(x).$$
- If X is a **continuous** random variable with probability density function $f(x)$,
$$V(X) = \int_{-\infty}^{\infty} (x-\mu_X)^2 f(x)\,\mathrm{d}x.$$

$\therefore$

- $V(X) \geq 0$ for any X. Equality holds if and only $P(X = E(X)) = 1$, that is, when X is a **constant**.
- Let a and b be any real numbers, then $V(aX+b) = a^2 V(X)$.
- The variance can also be computed by an alternative formula:
$$V(X) = E(X^2) - [E(X)]^2.$$
- The positive square root of the variance is defined as the **standard deviation** of X:
$$\sigma_X = \sqrt{V(X)}.$$

* Variance tells us the spread of data

$\downarrow$ Variance : Thinner, higher

$\uparrow$ Variance : Fatter, lower

There is no r/s btw. Variance & Expected Value

Integration Rules

$\int f(x)^n dx = \frac{f(x)^{n+1}}{n+1} + C \quad , n \neq -1$

$\int c \, dx = cx + C_1$

$\int \frac{1}{x} dx = \ln x + C$

$\int a^x dx = \frac{a^x}{\ln a} + C$

$\int e^x dx = e^x + C$

$\int e^{ax+b} dx = \frac{1}{a} e^{ax+b} + C$

$\int \sin(kx) \, dx = -\frac{\cos(kx)}{k} + C$

$\int \cos(kx) \, dx = \frac{\sin(kx)}{k} + C$

$\int \sec(x) \tan(x) \, dx = \sec x + C$

$\int \csc(x) \cot(x) \, dx = \csc x + C$

$\int \sec^2 x \, dx = \tan x + C$

$\int \csc^2 x \, dx = -\cot(x) + C$

Definite Integrals (Let $F(x) = \int f(x) \, dx$)

$$\int_a^b f(x) \, dx = [F(x)]_a^b = F(b) - F(a)$$

$$\int_a^b (f(x) + g(x)) \, dx = \int_a^b f(x) \, dx + \int_a^b g(x) \, dx$$