Contents

1 Probability Theory

1.1 Elemantary Probability

Definition 1.1.1 (Cardano's Principle)**.** A be a random outcome of an experiment that may proceed in various ways. Assume each of these ways is **equally likely**, then, probability $P[A]$ of outcome A is

$$P[A] = \frac{number\ of\ ways\ leading\ to\ outcome\ A}{number\ of\ ways\ the\ experiment\ can\ be\ proceeded} \tag{1}$$

1.1.1 Basic principles of Counting

- Permutation:
$$A_n^k = \frac{n!}{(n-k)!} \tag{2}$$

- Combination:
$$\tbinom{n}{k} = \frac{n!}{(n-k)!k!} \tag{3}$$

- Permutation of k **Indisguishable** Objects:

$$\frac{n!}{n_1!n_2!n_3!\cdots n_k!} = \binom{n}{n_1} \cdot \binom{n-n_1}{n_2} \cdots \binom{n-(n_1+n_2+\cdots+n_{k-1})}{n_k} \tag{4}$$

Remark:
In permutation of k indistinguishable objects, the elements has noorder within a group but are differnent from each other; the groups either has order or are different from each other.
Example:
Consider 10 balls, 5 red, 3 green and 2 blue. How many ways can they be arranged on a line?

$$\frac{10!}{5! \cdot 3! \cdot 2!} = 2520 \tag{5}$$

1.1.2 Sample Points, Sample Space and σ-Field

Definition 1.1.2 (Sample Points)**.** Mathematical objects are called sample points.

Definition 1.1.3 (Sample Space)**.** The sample space S is large enough to accommodate all the sample points.

Definition 1.1.4 (Event)**.** An outcome in the sense of Cardano's principle is interpreted as a subset A of a sample space S abd called an event.

Definition 1.1.5 (Mutually exclusive)**.** Two events A_1, A_2 are called mutual exclusive if $A_1 \cap A_2 = \emptyset$

Definition 1.1.6 (σ-Field)**.** Suppose that a non-empty set S is given. A *σ-field* $\mathcal{F}$ on S is a family of subsets of S such that:

- $\emptyset \in \mathcal{F}$
- If $A \in \mathcal{F}$, then $S \backslash A \in \mathcal{F}$
- If $A_1, A_2, \cdots \in \mathcal{F}$ is a finit sequence of subsets, then the union $\cup_k A_k \in \mathcal{F}$

1.1.3 Probability Measures and Spaces

Definition 1.1.7 (Probability Measure)**.** Let S be the sample space and $\mathcal{F}$ be a σ-field. Then a function

$$P : \mathcal{F} \to [0, 1], A \to P[A], \tag{6}$$

is called *probability measure / probability function* on S if

- $P[S] = 1$
- For any set of events $A_k \subset mathcalF$ such that $A_j \cap A_k = \emptyset$ for $j \neq k$,

$$P[\cup_k A_k] = \sum_k a_n P[A_k] \tag{7}$$

Theorem 1.1.1 (Basic Properties)**.** $P[A_1 \cup A_2] = p[A_1] + P[A_2] - P[A_1 \cap A_2]$

1.2 Conditional Probability

Definition 1.2.1 (Conditional Probability)**.** B occurs given that A has occured

$$P[B \mid A] := \frac{P[A \cap B]}{P[A]} \tag{8}$$

1.2.1 Independence of Events

Definition 1.2.2 (independent)**.** Two events are *independent* if

$$P[A \cap B] = P[A]P[B] \tag{9}$$

equivalent to

$$P[A \mid B] = P[A], P[B] \neq 0 \tag{10}$$

Definition 1.2.3 (Total Probability)**.**

$$P[B] = P[B \mid A_1] * P[A_1] + \cdots + P[B \mid A_n]P[A_n] = \sum_{k=1}^{n} P[B \mid A_k] \cdot P[A_k] \tag{11}$$

is called total propability formula for $P[B]$

1.2.2 Bayes' Theorem

Theorem 1.2.1 (Bayes's Theoremm)**.** Let $A_1, \cdots, A_n \subset S$ be a set of pairwise mutually exclusive events whose union is S and who each have non zero probability of occuring. Let $B \subset S$ be any events such that $P[B] \neq 0$. Then for any $A_k, k = 1, \cdots, n$

$$P[A_k \mid B] = \frac{P[B \cap A_k]}{P[B]} = \frac{P[B \mid A_k] \cdot P[A_k]}{\sum_{j=1}^{n} P[B \mid A_j] \cdot P[A_j]} \tag{12}$$

1.3 Discrete Random Variables

Definition 1.3.1 (Random Variable)**.** Such function X

$$X : S \to \mathbb{R} \tag{13}$$

is a random variable. X has numerical values that are derived from the outcome of a random experiment.

Two types:

- *Discrete Random Variables*: countable range in $\mathbb{R}$

  Definition 1.3.2 (Discrete Random Variable)**.** Let S be sample space, Ω a countable subset of $\mathbb{R}$. A *discrete random variable* is a map

  $$X : S \to \Omega \tag{14}$$

  $$f_X : \Omega \to \mathbb{R} \tag{15}$$

A random variable is often given by the pair (X, f_X)

- *Continuous Random Variables*: having a range equal to $\mathbb{R}$

Example:
Flip a coin three times, sample space can be given by :

$$S = (t,t,t), (t,t,h), (t,h,t), \cdots, (h,h,h) \tag{16}$$

Then we can define X as follows:

$$X(t,t,t) = 0, X(t,t,h) = 1, \cdots, X(h,h,h) = 3 \tag{17}$$

X denotes the number of heads

$$P[X = 1] = P[(t,t,h), (t,h,t), (h,t,t)] \tag{18}$$

We can write

$$P[X = x] = P[A] \tag{19}$$

where $x \in R$ and $A \subset S$ is the event containing all sample points p such that $X(p) = x$.

1.3.1 PDF and CDF

Random variable comes with a probability density function or probability distribution f_X that allows the calculation of probability directly.
Follows:

- $f_X(x) > 0$ for all x
- $\sum_{X \in \Omega} f_X(x) = 1$

Definition 1.3.3 (Cumulative distribution function)**.** Cumulative distribution function of a random variable is defined as

$$F_X : \mathbb{R} \to \mathbb{R}, F_X(x) := P[X \leq x] \tag{20}$$

For discrete random variable, $F_X(x) = \sum_{y \leq x} f_X(y)$

1.3.2 Bernoulli Random Variable

Definition 1.3.4 (Bernoulli Trial)**.** Consider an experiment can only results in two possible outcomes, such as success or failure.
and probability of success $0 < p < 1$

Definition 1.3.5 (Bernoulli Random Variable)**.** Let S be a sample space and $X : S \to 0, 1 \in \mathbb{R}$, Let $0 < P < 1$, then define the density function

$$f_X : 0, 1 \to R \quad f_X(x) = \begin{cases} 1 - p & for\ x = 0 \\ p & for\ x = 1 \end{cases} \tag{21}$$

Then X is said to be a *Bernoulli random variable* or follow a *Bernoulli distribution* with parameter p. We indicate this by writing

$$X \sim Bernoulli(p) \tag{22}$$

1.3.3 Independent and Identical Trials

- *independent* means the outcome of a trial will not influence the following trial
- *identical* means each trial has same probability of success

1.3.4 Counting Successes in a Saquance of Trials

$$P[x\ successes\ in\ n\ trials] = \tbinom{n}{x} p^x (1-p)^{n-x} \tag{23}$$

1.3.5 Binomal Random Variable

Definition 1.3.6 (Binomal Random Variable)**.** S: sample space, $n \in \mathbb{N} \backslash 0$ and

$$X : S \to \Omega = \{0, \cdots, n\} \in \mathbb{R} \tag{24}$$

Let $0 < p < 1$ and define the density function

$$f_X : \Omega \to \mathbb{R}, \quad f_X(x) = \tbinom{n}{x} p_x (1-p)^{n-x} \tag{25}$$

Then X is said to be a binomial random variable with parameters n and p. Indicate this by writing

$$X \sim B(n, p) \tag{26}$$

Also:

$$B(1, p) = Bernoulli(p) \tag{27}$$

1.3.6 Cumulative Distribution Function

Definition 1.3.7 (Cumulative Distribution Function)**.**

$$F_X : \mathbb{R} \to \mathbb{R}, \qquad F_X(x) := P[X \leq x] \tag{28}$$

For a discrete random variable

$$F_X(x) = \sum_{y \leq x} f_X(y) \tag{29}$$

In case of binomial distribution,

$$F_X(x) = \sum_{y=0}^{[x]} \tbinom{n}{y} p^y (1-p)^{n-y} \tag{30}$$

$[x]$ denote the largest integer not greater than x.

MMA command:
CDF[BinomalDistribution[n,p],x]

1.3.7 The Geometric Distribution

Another example: suppose er perform a sequence of Bernoulli trials which continues until a success is obtained. We then define the *geometric random variable* X to denote the number of trials needed to obtain the first success.
Example:
Result (t, t, t, h), geometric random variable X attains value $X = 4$

Definition 1.3.8 (Geometric Random Variable)**.**

$$X : S \to \Omega = \mathbb{N}\backslash\{0\} \tag{31}$$

Let $0 < p < 1$ and define density function $f_X(x) : \mathbb{N}\backslash\{0\} \to \mathbb{R}$

$$f_X(x) = (1-p)^{x-1}p \tag{32}$$

then, X is a geometric random variable with parameter p

$$X \sim Geom(p) \tag{33}$$

The cumulative distribution fuction for a geometrically distributed random variable (X, f_X) with parameter p is given by

$$F(x) = P[X \le x] = 1 - (1-p)^{[x]} \tag{34}$$

1.4 Expectation, Variance and Moments

1.4.1 Expectation

Definition 1.4.1 (Expectation)**.**

$$E[X] := \sum_{x \in \Omega} x \cdot f_X(x). \tag{35}$$

Lemma 1.4.1. Let (X, f_X) be a discrete random variable and $\varphi : \Omega \to \mathbb{R}$ some function. Then the expected value of $\varphi \circ X$ is

$$E[\varphi \circ X] = \sum_{x \in \Omega} \varphi(x) \cdot f_X(x) \tag{36}$$

Theorem 1.4.1.

$$E[X + Y] = E[X] + E[Y] \tag{37}$$

1.4.2 Location

It indicated where the values of a random variable are concentrated.

1.4.3 Variance and Standard Deviation

Definition 1.4.2 (variance)**.** A possible way to measure the dispersion of a random variable.
Which is the *mean squre deviation from the mean*
Given X, the deviation from mean is $X - E(X)$, then

$$Var[X] := E[(X - E[X])^2] \tag{38}$$

Definition 1.4.3 (Standard Deviation)**.**

$$\sigma_X = \sqrt{Var[X]} \tag{39}$$

Useful formula:

$$Var[X] = E[X^2] - E[X]^2 \tag{40}$$

Important Properties

- $Var[cX] = c^2 Var[X]$
- If X and Y are independent, then $Var[X + Y] = Var[X] + Var[Y]$

The expectation and variance of some distribution

- binomial distribution($X \sim B(n, p)$):
  - $E[X] = np$
  - $Var[X] = np(p - 1)$

  proof: https://www.cuemath.com/data/variance-of-binomial-distribution/
- geometric distribution($X \sim Geom(p)$)
  - $E[X] = \frac{1}{p}$
  - $Var[X] = \frac{1-p}{p^2}$

1.4.4 Standard Random Variables

$$Y = \frac{X - \mu}{\sigma} \tag{41}$$

We find that

$$\begin{aligned} E[Y] &= E[\frac{X - \mu}{\sigma}] = \frac{1}{\sigma} E[X - \mu]] \\ &= \frac{1}{\sigma}(E[X] - \mu) \\ &= 0 \end{aligned}$$

$$
\begin{aligned}
Var[Y] &= E[(\frac{X-\mu}{\sigma})^2] - E[Y]^2 \\
&= \frac{1}{\sigma^2} \cdot E[(X-\mu)^2] \\
&= 1
\end{aligned}
$$

1.4.5 Ordinary and Central Moments

Definition 1.4.4 (Ordinary Moments)**.** $E[X^n]$ is called n^{th} ordinary moments of X.

Definition 1.4.5 (Central Moments)**.** $E[(\frac{X-\mu}{\sigma})^n], \quad n = 3, 4, 5, \cdots$ are called n^{th} central moments of X.

1.4.6 The Moment-Generating Function

Definition 1.4.6 (moment-generating function)**.** Let (X, f_X) be a random variable and such that the sequence of moments $E[X^n], n \in \mathbb{N}$, exists.
If the power series

$$m_X(t) := \sum_{k=0}^{\infty} \frac{E[X^k]}{k!} t^k \tag{42}$$

has radius of convergence $\varepsilon > 0$, the defined function

$$m_X : (-\varepsilon, \varepsilon) \to \mathbb{R} \tag{43}$$

is called the moment-generating function for X.

Theorem 1.4.2. Let $\varepsilon > 0$ be given such that $E[e^t X]$ exits and has a power series expansion in t that converges for $|t| < \varepsilon$. Then moment-generating function exits and

$$m_X(t) = E[e^{tX}] \qquad for\ |t| < \varepsilon \tag{44}$$

furthermore,

$$E[X^k] = \frac{d^k * m_X(t)}{dt^k}|_{t=0} \tag{45}$$

We can hence calculate the moments of X by differentiating the moment-generating-function.

$$m_X(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} E[X^n] = E[\sum_{n=0}^{\infty} \frac{t^n X^n}{n!}] \tag{46}$$

Differentiating term-by-term

$$\frac{d^k m_X(t)}{dt^k} = \sum_{n=0}^{\infty} \frac{d_k}{dt^k} \frac{t^n}{n!} E[X^n] = \sum_{n=k}^{\infty} \frac{t^{n-k}}{(n-k)!} E[X^k] \tag{47}$$

At $t = 0$ only $n = k$ survives, so $E[X^k] = \frac{d^k * m_X(t)}{dt^k}|_{t=0}$

1.5 Pascal and Poisson Distribution