MATH1020 General Mathematics **Fall 2019**

Chapter 9: Probability

9.1 Sample Space of an experiment

In probability theory, we consider processes with uncertain results. For example, in rolling a dice, we may get any number from 1 to 6. We call such a process a random experiment. Its possible results are called outcomes. The set of all outcomes, usually denoted by S, is called the sample space. An event is a subset of sample space. Some examples are

- In the experiment of rolling a dice, the possible outcomes are the numbers from 1 to 6. The sample space is the set $S = \{1, 2, 3, 4, 5, 6\}$. The event "The number is even" is the subset $\{2, 4, 6\}$ of S.
- In another experiment, a coin with faces head(H) and tail(T) is tossed. The sample space is the set $\{\mathrm{H}, \mathrm{T}\}$ with two elements.
  If the coin is tossed twice, the sample space is
  $$S = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}.$$
  The outcome HT, for example, means getting a head in the first toss and a tail in the second toss.
- We consider the scores of a football match. Let (m, n) denote the outcome that Team A scores m goals and Team B scores n goals in a football match. Then the sample space is
  $$S = \{(m, n) : m, n \text{ are non-negative integers.}\}$$
  The event "Team A wins" is the subset $\{(m, n) \in S : m > n\}$.

9.2 Random Variables

A random variable is a variable whose values are numerical outcome of a random experiment. It may also be considered as a function $X : S \to \mathbb{R}$.

Example 1. A coin is tossed three times. Then sample space is

$$S = \{\text{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}\}$$

Let X be the number of times a head is showed. Then

$$\begin{aligned}
&X(\text{HHH}) = 3 \\
&X(\text{HHT}) = X(\text{HTH}) = X(\text{THH}) = 2 \\
&X(\text{HTT}) = X(\text{THT}) = X(\text{TTH}) = 1 \\
&X(\text{TTT}) = 0
\end{aligned}$$

Example 2. The sample space for the scores of a football match can be given by $S = \{(m,n) : m,n \geq 0\}$. Let Y be the random variable of total number of goals. Then

$$Y(m,n) = m + n.$$

Example 3. Let Z be the total rainfall next Sunday. Then Z is a random variable whose values can be any non-negative real numbers.

Below are the possible values of the random variables in the last three examples

$$\begin{aligned}
&X \in \{0,1,2,3\} \\
&Y \in \{0,1,2,3,\ldots\} = \{m \in \mathbb{Z} : m \geq 0\} \\
&Z \in [0,\infty)
\end{aligned}$$

Since the sets of possible values of X and Y are discrete, X and Y are called discrete random variables. On the other hand, Z takes value in an interval, which is not discrete. Z is called a continuous random variable. Our focus will be on discrete random variables.

Many operations on numbers can be done on random variables too. For example, we can add, subtract or multiply a random variable with a number or another random variables, for examples,

- Two coins, A and B, are each tossed 5 times. Let X_1 and X_2 be the number of times a head is shown by coin A and coin B respectively. Then $10 - X_1 - X_2$ is the random variable of the total number of times a tail is shown.
- If Y is the random variable of the number of couples will get married next year. then $2Y$ is the random variable of the number of persons will get married next year.

Probability mass function is often used to describe a random variable.

Definition 1. Let X be a discrete random variable. Define its probability mass function (PMF) to be

$$f(x) = P(X = x).$$

Example 4. Let X be the number showed from rolling a fair dice. Since the dice is a fair one, the probabilities of getting any of the six numbers 1 to 6 are the same and equal $\frac{1}{6}$. Also, it is impossible to get any other numbers. Hence,

$$f(x) = P(X = x) = \begin{cases} 1/6 & \text{if } x = 1, 2, \ldots \text{ or } 6; \\ 0 & \text{otherwise.} \end{cases}$$

From the properties of probability, it is clear that

> Let X be a discrete random variable and $f(x)$ be its probability mass function. Then
>
> 1. $0 \leq f(x) \leq 1$ for any $x \in \mathbb{R}$.
> 2. $\sum_x f(x) = 1$, where the sum is taken over all possible values x of X.

Mean, Variance, Standard deviation

> **Definition 2.** Let X be a discrete random variable and $f(x)$ be its probability mass function. Define its expected value to be
>
> $$E(X) = \sum_x x f(x),$$
>
> where the sum is over all x with $f(x) \neq 0$. $E(X)$ is also called the mean of X and can be denoted by μ. Also, define the variance of X to be
>
> $$Var(X) = E((X - \mu)^2) = \sum_x (x - \mu)^2 f(x)$$
>
> and the standard deviation of X to be
>
> $$\sigma = \sqrt{Var(X)}.$$

For a random variable X, the expected value $E(X)$ is roughly the average value of X obtained if the experiment is repeated many times. The variance and the standard deviation measure how spread out a distribution is. A smaller standard deviation indicates that the value of X is concentrated near its mean.

Example 5. Let X be the number shown from a fair dice. Then

$$\begin{aligned} E(X) &= \sum_x f(x) \\ &= (1)f(1) + 2f(2) + \cdots + 6f(6) \\ &= 1 \cdot \tfrac{1}{6} + 2 \cdot \tfrac{1}{6} + \cdots + 6 \cdot \tfrac{1}{6} = \tfrac{7}{2} \end{aligned}$$

$$\begin{aligned} Var(X) &= \sum_x (x-\mu)^2 f(x) \\ &= \left(1 - \tfrac{7}{2}\right)^2 \tfrac{1}{6} + \left(2 - \tfrac{7}{2}\right)^2 \tfrac{1}{6} + \cdots + \left(6 - \tfrac{7}{2}\right)^2 \tfrac{1}{6} = \tfrac{35}{12} \end{aligned}$$

$$\sigma = \sqrt{Var(X)} = \sqrt{\tfrac{35}{12}} \approx 1.708$$

Let X be a discrete random variable. Then

$$Var(X) = E(X^2) - E(X)^2 = \sum_x x^2 f(x) - \left(\sum_x x f(x)\right)^2$$

Proof.

$$\begin{aligned} Var(X) &= \sum_x (x-\mu)^2 f(x) \\ &= \sum_x (x^2 - 2\mu x + \mu^2) f(x) \\ &= \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x) \\ &= E(X^2) - 2\mu E(X) + \mu^2(1) \\ &= E(X^2) - 2E(X)E(X) + E(X)^2 \\ &= E(X^2) - E(X)^2 \end{aligned}$$

□

We end this section with a few more formulas on expected values and variance.

Let X, Y be discrete random variables and c be a constant. Then

1. $E(X+Y) = E(X) + E(Y)$
2. $E(X+c) = E(X) + c$
3. $E(cX) = cE(X)$
4. $Var(cX) = c^2 Var(X)$

9.3 Binomial distribution

We first consider the following example.

Example 6. A fair coin with faces head(H) and Tail(T) is tossed four times. Let X be the random variables of the number of times a head is showed. Find the probability mass function of X.

Solution. There are $2^4 = 16$ outcomes with equal probability. We list all of them below

0 head:	TTTT	$C^4_0 = 1$ combination
1 head:	HTTT, THTT, TTHT, TTTH	$C^4_1 = 4$ combinations
2 heads:	HHTT, HTHT, HTTH, THHT, THTH, TTHH	$C^4_2 = 6$ combinations
3 heads:	HHHT, HHTH, HTHH, THHH	$C^4_3 = 4$ combinations
4 heads:	HHHH	$C^4_4 = 1$ combination

For a fair coin, all these 16 outcomes has equal probability $\dfrac{1}{16}$. Hence, the probability mass function is given by

$$f(k) = P(X=k) = \begin{cases} \frac{1}{16} \text{ if } k=0 \\ \frac{4}{16} \text{ if } k=1 \\ \frac{6}{16} \text{ if } k=2 \\ \frac{4}{16} \text{ if } k=3 \\ \frac{1}{16} \text{ if } k=4 \\ 0 \text{ otherwise.} \end{cases}$$

■

What happens if a biased coin with 60% of showing a head is used instead in the example above? We will still have $2^4 = 16$ total possible outcomes. However, they do not have the same probability. For instance, each of the $C_1^4 = 4$ outcomes with exactly 1 head

HTTT, THTT, TTHT, TTTH

has probability $(0.6)(0.4)^3$ and so $f(1) = C_1^4(0.6)(0.4)^3 = 0.1536$ instead.

More generally, suppose an experiment with only two outcomes, either success with probability p or failure with probability $1-p$, is repeated n times. Let X be the random variable of the number of successes. Then its probability mass function

$$f(k) = C_k^n p^k (1-p)^{n-k} \quad \text{for } k = 0, 1, 2, \ldots, n.$$

The random variable X is said to follow the **binomial distribution** and is called the **binomial random variable** with parameters n and p, denoted by $X \sim B(n,p)$.

Let $X \sim B(n,p)$ be the binomial random variable with parameters n and p. Then its probability mass function, expected value (mean) and variance are

$$\begin{aligned} f(k) &= \begin{cases} C_k^n p^k (1-p)^{n-k} & \text{if } k = 0, 1, \cdots, n; \\ 0 & \text{otherwise.} \end{cases} \\ E(X) &= \mu = np \\ Var(X) &= \sigma^2 = np(1-p) \end{aligned}$$

Proof for the formula of $E(X)$. We first prove a useful formula of C_r^n. For $n \geq k \geq 1$,

$$kC_k^n = k \cdot \frac{n!}{k!(n-k)!} = \frac{n!}{(k-1)!(n-k)!} = n \cdot \frac{(n-1)!}{(k-1)![(n-1)-(k-1)]!} = nC_{k-1}^{n-1}$$

Hence,

$$\begin{aligned} \mu &= \sum_{k=0}^{n} kf(k) && \text{(The term with } k = 0 \text{ is zero.)} \\ &= \sum_{k=1}^{n} kC_k^n p^k (1-p)^{n-k} \\ &= \sum_{k=1}^{n} nC_{k-1}^{n-1} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^{n} C_{k-1}^{n-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\ &= np \sum_{l=0}^{n-1} C_l^{n-1} p^l (1-p)^{(n-1)-l} \\ &= np[p + (1-p)]^{n-1} && \text{(By binomial theorem)} \\ &= np \end{aligned}$$

□

Example 7. A biased coin with 60% chance head(H) and 40% chance tail(T) is tossed four times. Let Y be the random variables of the number of times H is showed. Find the probability mass function, mean and standard deviation of Y.

Solution. Note $n = 4$, $p = 0.6$ and $Y \sim B(4, 0.6)$. By $f(k) = C_k^n p^k (1-p)^{n-k}$,

$$f(k) = \begin{cases} 1 \cdot (0.4)^4 = 0.0256 & \text{if } k = 0; \\ 4 \cdot (0.6)(0.4)^3 = 0.1536 & \text{if } k = 1; \\ 6 \cdot (0.6)^2(0.4)^2 = 0.3456 & \text{if } k = 2; \\ 4 \cdot (0.6)^3(0.4) = 0.3456 & \text{if } k = 3; \\ 1 \cdot (0.6)^4 = 0.1296 & \text{if } k = 4; \\ 0 & \text{otherwise.} \end{cases}$$

The mean and variance are

$$\mu = np = 4(0.6) = 2.4$$
$$Var(Y) = np(1-p) = 4(0.6)(1-0.6) = 0.96$$

Standard deviation is $\sigma = \sqrt{Var(Y)} = \sqrt{0.96} \approx 0.980$. ■

Example 8. A supermarket sells apples with 10 in a bag. If there are at least 3 bad apples in the bag, a customer can return the whole bag. Suppose that there is a 5% chance that an apple is bad. Find the probability that a randomly chosen bag of apples can be returned.

Solution. Let X be the number of bad apples in a randomly chosen bag of apples. Then $X \sim B(10, 0.05)$ and has probability mass function $f(k) = C_k^{10}(0.05)^k(1-0.05)^{10-k} = C_k^{10}(0.05)^k(0.95)^{10-k}$ for $k = 0, 1, \ldots, 10$. In particular,

$$f(0) = C_0^{10}(0.05)^0(0.95)^{10} \approx 0.5987$$
$$f(1) = C_1^{10}(0.05)^1(0.95)^9 \approx 0.3151$$
$$f(2) = C_2^{10}(0.05)^2(0.95)^8 \approx 0.0746$$

Hence, the probability that a randomly chosen bag of apples can be returned is

$$P(X \geq 3) = 1 - f(0) - f(1) - f(2) \approx 0.0115 = 1.15\%$$

■

9.4 Geometric distribution

Suppose an experiment with success rate p is carried out repeatedly until the first success is achieved. Let X be the random variable of the number of trials needed. Then $X = k$ means failures in the first $k-1$ attempts and success in the k-th attempt. The probabilities for $X = 1, 2, 3, 4$ are

Values of X	Trails (Success/Failure)	Probability
$X = 1$	S	p
$X = 2$	FS	$(1-p)p$
$X = 3$	FFS	$(1-p)^2 p$
$X = 4$	FFFS	$(1-p)^3 p$

Clearly, X has probability mass function

$$f(k) = P(X = k) = (1-p)^{k-1} p \quad \text{for any positive integer } k.$$

Such a X is said to follow the geometric distribution and is called the geometric random variable with parameter p.

Example 9. A basketball player attempts 3-point shots until a shot is made. Suppose he makes 20% of his shots on average. Let X be the number of attempts needed. Find the probability that

1. He needs exactly 4 attempts.
2. He needs more than 5 attempts.

Solution.

1. Note X is a geometric random variable with parameter $p = 0.2$. Hence, the probability that he needs exactly 4 attempts is

$$P(X = 4) = (1-0.2)^3(0.2) = 0.1024.$$

2. If more than 5 attempts are needed, that means he misses all the first 5 shots. Hence, the required probability is

$$P(X > 5) = (1-0.2)^5 = 0.32768.$$

■

Here are some formulas for geometric random variables.

Let X be the geometric random variable with parameter p. Then its probability mass function, expected value (mean) and variance are

$$f(k) = \begin{cases} (1-p)^{k-1}p & \text{if } k \text{ is a positive integer;} \\ 0 & \text{otherwise.} \end{cases}$$

$$E(X) = \mu = \frac{1}{p}$$

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

Proof for the formula of $E(X)$.

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

To compute it, we first derive a formula. For $|x| < 1$,

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + \cdots ,$$

By differentiation,

$$\left(\frac{1}{1-x}\right)^2 = 1 + 2x + 3x^2 + 4x^3 + \cdots = \sum_{k=1}^{\infty} kx^{k-1} \qquad (*)$$

Put $x = 1 - p$, then

$$\frac{1}{p^2} = \sum_{k=1}^{\infty} k(1-p)^{k-1}$$

Hence,

$$E(X) = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = p \cdot \sum_{k=1}^{\infty} k(1-p)^{k-1} = p \cdot \frac{1}{p^2} = \frac{1}{p}.$$

□

Remark. The formula for $Var(X)$ can also be proved similarly. Multiplying x to the equation (*) above, we obtain

$$\frac{x}{(1-x)^2} = \sum_{k=1}^{\infty} kx^k \qquad (*)$$

Differentiating this gives a formula for $\sum_{k=1}^{\infty} k^2x^{k-1}$, which can be used to compute $Var(X)$ from the formula $Var(X) = E(X^2) - E(X)^2$. The details are left to readers.

Example 10. A fair dice is thrown until six is shown. Find the expected number of throws needed and the standard deviation.

Solution. Let X be the number of throws needed. Then X is a geometric random variable with parameter $p = \frac{1}{6}$. Hence, its expected value and standard deviation are

$$E(X) = \frac{1}{p} = 6$$

$$\sigma = \sqrt{Var(X)} = \sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{1-\frac{1}{6}}{\left(\frac{1}{6}\right)^2}} = \sqrt{30} \approx 5.477.$$

■

9.5 Poisson Distribution

If a company receives 5 phone enquiries per hour on average, what is the probability that exactly 2 phone enquiries will be received within the next hour? This type of questions can be answered using the Poisson distribution to be discussed in this section.

Suppose an event occurs λ times in a given time interval on average. Let X be the random variable of the number of times the event will occur in a particular time period of that length. Under the model of the Poisson distribution, the probability mass function of X is given by

$$f(k) = P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{for any integer } k \geq 0.$$

The random variable X is called the Poisson random variable with parameter λ. For instance, to answer the question in last paragraph, we set $\lambda = 5$ and $k = 2$ to obtain the required probability

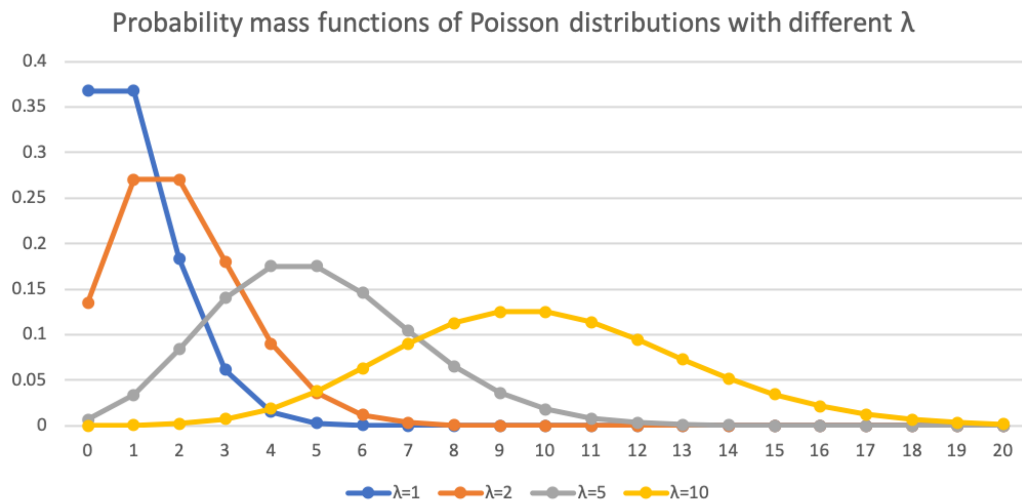
$$f(2) = \frac{5^2 e^{-5}}{2!} = \frac{25}{2e^5} \approx 0.0842.$$

The Poisson distribution is derived based on the following assumptions.

- Occurrences of the event are independent of each other. For instance, in the example of phone enquiries above, we assume that person A making enquiry does not affect whether person B would make enquiry.
- The average rate λ is constant.
- The event cannot occur simultaneously.

Very often, these conditions are not totally fulfilled in the situations we consider. However, the Poisson distribution can still be applied to give good enough predictions.

Below is the graphs of the probability mass functions of the Poisson distribution with parameters $\lambda = 1, 2, 5$ and 10. These functions attain their maxima near λ as expected.



Here are some formulas for Poisson random variables.

Let X be the Poisson random variable with parameter λ. Then its probability mass function, expected value (mean) and variance are

$$f(k) = \begin{cases} \dfrac{\lambda^k e^{-\lambda}}{k!} & \text{for any integer } k \geq 0.; \\ 0 & \text{otherwise.} \end{cases}$$

$$E(X) = \mu = \lambda$$

$$Var(X) = \lambda$$

Example 11. Suppose a flood occurs once every 100 years on average. Find the probability that

1. there are 2 floods in the next 100 years;
2. there are 4 floods in the next 200 years.

Solution. 1. Let X be the number of floods in the next 100 years. Then X is the Poisson random variable with parameter $\lambda = 1$ and probability mass function

$$f(k) = \frac{\lambda^k e^{-\lambda}}{k!} = \frac{e^{-1}}{k!}$$

Hence, the required probability is

$$P(X = 2) = f(2) = \frac{e^{-1}}{2!} \approx 0.184.$$

2. Let Y be the number of floods in the next 200 years. Since 2 flood occur every 200 years on average, Y is the Poisson random variable with parameter $\lambda = 2$ and probability mass function

$$f(k) = \frac{2^k e^{-2}}{k!}$$

Hence, the required probability is

$$P(Y = 4) = f(4) = \frac{2^4 e^{-2}}{4!} \approx 0.090.$$

■

Example 12. Suppose a secretary makes 1 typo in 7 pages on average. Let X be the number of typos she will make in a document of 10 pages.

1. What are the mean and variance of X?
2. Find the probability that there will be more than two typos in the document.

Solution. 1. The secretary makes $10(\frac{1}{7}) = \frac{10}{7}$ typos in a 10-page document on average. Hence, X follows the Poisson distribution with parameter $\lambda = \frac{10}{7}$. By the formulas for Poisson distribution, both the mean and variance of X are $\lambda = \frac{10}{7}$.

2. The probability mass function of X is

$$\frac{(\frac{10}{7})^k e^{-10/7}}{k!}$$

The probability that there will be more than two typos in the document is

$$\begin{aligned} P(X > 2) &= 1 - P(X = 0) - P(X = 1) - P(X = 2) \\ &= 1 - f(0) - f(1) - f(2) \\ &= 1 - \frac{(\frac{10}{7})^0 e^{-10/7}}{0!} - \frac{(\frac{10}{7})^1 e^{-10/7}}{1!} - \frac{(\frac{10}{7})^2 e^{-10/7}}{2!} \\ &\approx 1 - 0.240 - 0.342 - 0.245 \\ &= 0.173 \end{aligned}$$

■

At the end, we want to explain how to obtain the probability mass function of the Poisson distribution.

Given a time period. We divide it into n equal subintervals. The average rate for the event to occur in each of these subintervals is $\frac{\lambda}{n}$. Then X can then be approximated by the binomial random variable $B(n, \frac{\lambda}{n})$. The approximation is better as n increases. Indeed, X can be considered as the "limit" of $B(n, \frac{\lambda}{n})$ with $n \to \infty$. Hence

$$\begin{aligned}
f(k) &= \lim_{n\to\infty} C_k^n \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \lim_{n\to\infty} \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!} \lim_{n\to\infty} \left[\frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k}\right] \left(1-\frac{\lambda}{n}\right)^{n} \left(1-\frac{\lambda}{n}\right)^{-k} \\
&= \frac{\lambda^k}{k!} \lim_{n\to\infty} \left[\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\right] \left(1-\frac{\lambda}{n}\right)^{n} \left(1-\frac{\lambda}{n}\right)^{-k} \\
&= \frac{\lambda^k}{k!} \left[(1-0)(1-0)\cdots(1-0)\right] e^{-\lambda} (1-0)^{-k} \\
&= \frac{\lambda^k e^{-\lambda}}{k!}
\end{aligned}$$