MA120 INTRODUCTION TO PROBABILITY THEORY AND STATISTICS
EXPECTED VALUES

Suppose N observation of a random variable consist of n_0 zeroes, n_1 ones, n_2 twos,.... Then the sample mean, or average, $\overline{x}$ can be written as

$$\begin{aligned}\overline{x} &= \frac{0n_0 + 1n_1 + 2n_2 + \dots}{N} \\ &= 0\left(\frac{n_0}{N}\right) + 1\left(\frac{n_1}{N}\right) + 2\left(\frac{n_2}{N}\right) + \dots \\ &= 0p_0 + 1p_1 + 2p_2 + \dots = \sum_x xp_x\end{aligned}$$

where $p_x = \frac{n_x}{N}$, the observed proportion of x's.
Now let $N \to \infty$. Then $p_x \top(x)$ for all x, so that

$$\overline{x} \longrightarrow_{N\to\infty} \sum_x xp(x).$$

Example 1 The number of complaints received by a shop in a day follows the distribution with the probability function

x	0	1	2	3	4	5
$p(4)$	$\frac{2}{20}$	$\frac{4}{20}$	$\frac{7}{20}$	$\frac{4}{20}$	$\frac{2}{20}$	$\frac{1}{20}$

Hence

$$E(X) = \sum_x xp(x) = 0\left(\frac{2}{20}\right)+1\left(\frac{4}{20}\right)+2\left(\frac{7}{20}\right)+\dots+5\left(\frac{1}{20}\right) = \frac{43}{20} = 2.15.$$

Example 2 - Bernoulli distribution Suppose X follows the Bernoulli distribution with probability function

$$p_X(x) = \begin{cases} \pi, & x = 1, \\ 1-\pi, & x - 0. \end{cases}$$

Then

$$E(X) = 1(\pi) + 0(1-\pi) = \pi.$$

Example 3 - Binomial distribution If $X_1, X_2, \ldots, X_n$ is a sequence of Bernoulli trials and $Y = X_1 + X_2 + \ldots X_n$, then Y follows a binomial distribution with probability function

$$p_Y(y) = \begin{pmatrix} n \\ y \end{pmatrix} \pi^y (1-\pi)^{n-y}, \ y = 0, 1, \ldots, n.$$

Consequently,

$$\begin{aligned} E(Y) &= \sum_{y=0}^{n} y \begin{pmatrix} n \\ y \end{pmatrix} \pi^y (1-\pi)^{n-y} \\ &= \sum_{y=1}^{n} y \begin{pmatrix} n \\ y \end{pmatrix} \pi^y (1-\pi)^{n-y} \\ &= n \sum_{y=1}^{n} \begin{pmatrix} n-1 \\ y-1 \end{pmatrix} \pi^y (1-\pi)^{n-y} \\ &= n\pi \sum_{y'=0}^{n'} \begin{pmatrix} n' \\ y' \end{pmatrix} \pi^{y'} (1-\pi)^{n'-y'}, \text{ where } y' = y-1 \text{ and } n' = n-1, \\ &= n\pi. \end{aligned}$$

Expected value of $h(X)$ Suppose X is a random variable whose distribution we know but that we wish to know the expected value of $h(X)$, an arbitrary function of X. It would seem that we need to first to derive the distribution of $Y = h(X)$ and then find $E(Y)$. However, this is not necessary. It can be proved that

$$E(Y) = E[h(X)] = \sum_x h(x) p_X(x).$$

As a result of the above, we have

$$E(aX + b) = \sum_x (ax + b)p(x) = a \sum_x xp(x) + b \sum_x p(x) = aE(X) + b,$$

where a and b are constants, and generally

$$E[ah(X) + b] = aE[h(X)] + b \text{ and } E\left[\sum_{i=1}^{k} a_i h_i(X)\right] = \sum_{i=1}^{k} a_i E[h_i(X)]$$

for arbitrary functions $h_i(X), i = 1, 2, \ldots, k$ and constants $a_i, i = 1, 2, \ldots, k$.

In example 3, we therefore have $E\left(\frac{Y}{n}\right) = \pi$.

Moments Let

$$\mu'_r = E(X^r).$$

Then μ'_r is called the rth *moment* of X (about the origin). In particular, $\mu'_1 = E(X)$ is the mean of X and is usually denoted by μ. It is the most important measure of location of a distribution.

Now let

$$\mu_r = E[(X - \mu)^r].$$

Then μ_1 is called the rth moment of X about the mean, or the rth *central moment* of X. Note that $\mu_1 = 0$. When $r = 2$, we have

$$\mu_2 = E[(X - \mu)^2] = \text{var}(X),$$

the *variance* of X, which is often denoted by σ^2, and is a measure of spread of a distribution. Further, $\sigma = \sqrt{\text{var}(X)} = \text{sd}(X)$ is called the *standard deviation* of X and is also a measure of spread but with the same dimension as X itself.

It is sometimes easier to calculate variance using the relation

$$\text{var}(X) = E[(X - \mu)^2] = E(X^2) - \mu^2.$$

For some distributions, however, the easiest second moment to calculate is $S[X(X-1)]$ (known as the second factorial moment), from which we may obtain

$$\text{var}(X) = E[X(X-1)] + \mu - \mu^2.$$

Example 1 For the number of complaints received by the shop,

$$E(X^2) = 0^2\left(\frac{2}{20}\right) + 1^2\left(\frac{4}{20}\right) + 2^2\left(\frac{7}{20}\right) + \ldots + 5^2\left(\frac{1}{20}\right) = \frac{125}{20} = \frac{25}{4}$$

so that

$$\text{var}(X) = \frac{25}{4} - \left(\frac{43}{20}\right)^2 = \frac{651}{400} = 1.6275 \text{ and}$$

$$\text{sd}(X) = \sqrt{1.6275} = 1.276$$

Example 3 Here $Y \sim b(n, \pi)$ so that

$$\begin{aligned} E[Y(Y-1)] &= \sum_{y=0}^{n} y(y-1) \binom{n}{y} \pi^y (1-\pi)^{n-y} \\ &= n(n-1) \sum_{y=2}^{n} \binom{n-2}{y-2} \pi^y (1-\pi)^{n-y} = n(n-1)\pi^2. \end{aligned}$$

Hence

$$\text{var}(Y) = n(n-1)\pi^2 + n\pi - (n\pi)^2 = n\pi(1-\pi).$$

Example 2 When in example 3, we have

$$\text{var}(X) = \pi(1-\pi)$$

for the Bernoulli distribution.

Note that, in general,

$$\begin{aligned} \text{var}(aX+b) &= E\{(aX+b) - [aE(X)+b]\}^2 \\ &= E\{a[X-E(X)]\}^2 = a^2\text{var}(X). \end{aligned}$$

So, in example 3, $\text{var}\left(\frac{Y}{n}\right) = \left(\frac{1}{n}\right)\text{var}(Y) = \frac{\pi(1-\pi)}{n}$.

Higher moments The third central moment of X, $\mu_3 = E[(X-\mu)^3]$, or its standardised form $\frac{\mu_3}{\mu_2^{\frac{3}{2}}}$, is often considered to give a measure of skewness of a distribution. A symmetric distribution has $\mu_3 = 0$. If $\mu_3 > 0$, the distribution is said to be positively skewed and, if $\mu_3 < 0$, negatively skewed. Note however that, for a distribution to be symmetric, it is necessary for all its odd central moments to be zero.

The fourth central moments of X, $\mu_4 = E[(X-\mu)^4]$, or its standardised form $\frac{\mu_4}{\mu_2^2}$, is sometimes considered to give a measure of *kurtosis* (from the Greek for Bulging) or peakedness of a distribution. The standard, or normal, value is taken to be 3.