17 Expectation and random variables

17.1 Average of a sample and mean of a distribution

Given a probability distribution $p_1, \ldots, p_k$, on the sample space $S = x_1, \ldots, x_k$, where the sample values x_i are numbers (not, for example, 'heads' or 'tails'), suppose that n independent[1] trials are made with outcomes $x_{i_j},\ 1 \leqslant j \leqslant n$.

The *sample average* is

$$\frac{x_{i_1} + \ldots + x_{i_n}}{n}$$

Remember the 'frequentist principle,' that if a large number of trials is made, n trials, say, and the number of times x_i is the outcome, for $1 \leqslant i \leqslant k$, is n_k, then

$$\frac{n_i}{n} \approx p_i.$$

And in some vague sense the approximation gets better as n gets bigger.

Then the sample average is

$$\frac{n_1 x_1 + n_2 x_2 + \ldots + n_k x_k}{n} \approx p_1 x_1 + \ldots + p_k x_k$$

since $n_1/n \approx p_1,\ n_2/n \approx p_2, \ldots$

If n is large, the sample average comes very close to

$$(*) \qquad p_1 x_1 + \ldots + p_k x_k$$

(17.1) Definition *The quantity (*) is called the* mean *of the distribution.*

Example Uniform distribution: $p_i = 1/k, 1 \leqslant i \leqslant k$. The mean is

$$\frac{1 + 2 + \ldots + k}{k} = \frac{k(k+1)}{2k} = \frac{k+1}{2}$$

Example Binomial $B(3,p)$. The sample space is $0, 1, 2, 3$, and $p_i = \binom{3}{i} p^i (1-p)^{3-i}$. Mean is

$$\begin{gathered} 0 \times (1-p)^3 + 1 \times 3 \times (1-p)^2 p + 2 \times 3 \times (1-p) p^2 + 3 \times p^3 = \\ 3p(1-p)^2 + 6p^2(1-p) + 3p^3 = \\ 3p - 6p^2 + 3p^3 + 6p^2 - 6p^3 + 3p^3 = \\ 3p \end{gathered}$$

The mean of $B(n,p)$ is np (see tables).

The mean of a distribution is one of many kinds of *expectation*, see below.

[1] Meaning *roughly* that the outcome of one trial does not affect other outcomes.

17.2 Random variables and expectation

We generally use X to represent something which takes values in a sample space according to a given distribution. Each trial produces a particular outcome for X.

I would call X a *random variable,* though the official definition of random variable is different and counter-intuitive.

To help make sense of this, I would call X a *basic* random variable, from which others can be constructed. The word 'basic' would not be recognised in the general literature.

We generally use $X_1, \ldots, X_n$ to represent a sequence of basic random variables all following the same distribution over the same sample space.

(17.2) Definition *Let X be a (basic) random variable taking values in a sample space $S = \{x_1, \ldots, x_k\}$ with probabilities $p_1, \ldots, p_k$. A random variable is any function $f : S \to \mathbb{R}$. We'll call it Y to allow for the notation $E(Y)$.*

The expectation $E(Y)$ *is*

$$p_1 f(x_1) + \ldots + p_k f(x_k)$$

It is consistent with these definitions that $E(X)$ is the mean of the distribution.

(17.3) Lemma *$E(Y)$ is the limiting value of*

$$\frac{f(x_{i_1}) + \ldots + f(x_{i_n})}{n}$$

for large n. ∎

17.3 Variance

For example, $B(3,p)$, each sample taking 3 trials with two outcomes A, B, A with probability p, B with probability $1-p$.

Let μ be its mean (this is conventional). Let $Y = (X-\mu)^2$. This is the *variance* $\mathrm{Var}(X)$ of the distribution. X counts the number of A-outcomes. The distribution is

i	0	1	2	3
p_i	$(1-p)^3$	$3p(1-p)^2$	$3p^2(1-p)$	p^3
$(i-3p)^2$	$27p^2$	$(1-3p)^2$	$(2-3p)^2$	$9(1-p)^2$

This is difficult to calculate directly, so we use the equation, which is quite easy to explain,

$$E((X-\mu)^2) = E(X^2) - \mu^2.$$

$E(X^2)$:

i	0	1	2	3
p_i	$(1-p)^3$	$3p(1-p)^2$	$3p^2(1-p)$	p^3
i^2	0	1	4	9

$$3p(1-2p+p^2) + 12p^2(1-p) + 9p^3 = 3p - 6p^2 + 3p^3 + 12p^2 - 12p^3 + 9p^3 =$$
$$3p + 6p^2. \qquad \text{Subtract } \mu^2 : \quad 3p + 6p^2 - (3p)^2 = 3p - 3p^2 = 3p(1-p)$$

Generally: $B(n,p)$ has mean np and variance $np(1-p)$, as given in the tables.

17.4 Standard deviation

The standard deviation of a distribution is

$$\sqrt{\text{Variance}}$$

Notation: σ for the standard deviation, and σ^2 for the variance.

17.5 Sample variance

The difficult, and incomplete, calculation in this section is included as a matter of interest, not as a requirement.

But it is important to know the result, i.e. the correct formula for sample variance, and to use the correct version on a calculator (there are usually two).

There is a small but interesting difference between variance and sample variance. In this section a partial explanation is given. The sample space must be real-valued.

Given independent basic random variables $X_1, \ldots, X_n$ following the same distribution, for any i, j the product $X_i X_j$ can be shown to follow a certain distribution, and to be another kind of random variable. This is a bit vague, but the upshot is:

(17.4) Lemma *If $i \neq j$ then $E(X_i X_j) = E(X_i)E(X_j)$.* ∎

Now let us write $\overline{X}$ for the sample average

$$\frac{X_1 + \ldots + X_n}{n}$$

This is yet another example of a random variable.

Again, let $\mu = E(X_i)$, same for each i, $1 \leqslant i \leqslant n$; μ is the mean of the distribution.

The variance is $E(X - \mu)^2$.

One might suppose that 'sample variance' should be defined as

$$\frac{(X_1 - \overline{X})^2 + \ldots + (X_n - \overline{X})^2}{n}$$

but this would be *wrong.*

Consider the following

$$E\left((X_1 - \overline{X})^2 + \ldots + (X_n - \overline{X})^2\right)$$

This is

$$\begin{aligned} & E(X_1^2 + \ldots + X_n^2 - 2\overline{X}(X_1 + \ldots + X_n) + E(n(\overline{X})^2)) \\ & = \sum(E(X_i^2) - \overline{X}(X_1 + \ldots + X_n)) \end{aligned}$$

which is

$$(*) \quad nE(X^2) - E(\overline{X}(X_1 + \ldots + X_n))$$

Look at

$$\overline{X}(X_1 + \ldots + X_n)$$

This is

$$\frac{(X_1 + \ldots + X_n)^2}{n}$$

There are n^2 terms $X_i X_j$. If $i \neq j$, $E(X_i X_j) = \mu^2$. The expectation is

$$E\left(\frac{\sum_{i=1}^n X_i^2}{n}\right) + \frac{n(n-1)\mu^2}{n} = E(X^2) + (n-1)\mu^2$$

So (*) equals

$$nE(X^2) - E(X^2) - (n-1)\mu^2 = (n-1)E(X^2) - \mu^2.$$

It has already been mentioned that $E(X-\mu)^2 = E(X^2) - \mu^2$, so (*) equals

$$(n-1)\sigma^2$$

where σ^2 is the variance.

For this reason, the sample variance of a sample is defined as

$$S^2 = \frac{(X_1 - \overline{X})^2 + \ldots + (X_n - \overline{X})^2}{n-1}$$

The sample standard deviation would be

$$\sqrt{\text{sample variance}}$$

Notation: S would be the sample standard deviation and S^2 the sample variance, by analogy with σ and σ^2. Then $E(S^2) = \sigma^2$. Again, S and S^2 are random variables.

The denominator is $n-1$, not n which one would expect. The difference is small, but there is a difference. It is defined in this way, in order that its expectation is the true variance $E(X-\mu)^2$. Calculators provide both versions of sample variance. One should be careful to use the correct one.