5 Averages, Law of Large Numbers, and Central Limit Theorem

Expectation, standard deviation and correlation.

Expectation is the most widely used notion of average in statistics, because of its intuitive interpretations and convenient properties.

Linearity is the most important property of expectation.

The law of large numbers (LLN) and central limit theorem (CLT) are powerful results about the sample mean of a large number of IID RVs.

The LLN says that the sample mean is likely to be close to the theoretical expectation. The CLT says that the sample mean will be approximately Normal.

5.1 Expectation

Arithmetic mean of $x_1, x_2, \ldots, x_n$:

$$\overline{x} = \frac{1}{n} \sum_{j=1}^{n} x_j.$$

More generally, weighted mean of $x_1, x_2, \ldots, x_n$:

$$weighted\_mean(x) = \sum_{j=1}^{n} x_j p_j.$$

Where the weights $p_1, p_2, \ldots, p_n$ are pre-specified nonnegative numbers with $\sum_{j=1}^{n} p_i = 1$ (so for unweighted mean $p_j = 1/n \; \forall \; i$).

The definition of *expectation of DRV* is inspired by weighted mean, weights are probabilities:

The **expected value** (or **expected mean**) of DRV X whose possible values are $x_1, x_2, \ldots$ is:

$$E(X) = \sum_{j=1}^{\infty} x_j P(X = x_j).$$

If support is finite,

$$E(X) = \sum_x x P(X = x)$$

Where x is value and $P(X = x)$ is PMF at x.

The expectation is undefined if $\sum_{j=1}^{\infty} |x_j| P(X = x_j)$ diverges, since then the series for $E(X)$ diverges or depends on the order in which the x_j are listed.

Warning!

For any DRV X, $E(X)$ is a number (if exists). Common mistake: to replace an RV by its expectation which is wrong because X is a function, $E(X)$ is a constant, and ignores the variability of X.

Notation:

$E(X)$ is abbreviating to EX, similarly EX^2 is $E(X^2)$ not $(E(X))^2$!

The order of operations here is very important!

5.2 Linearity of expectation

The most important property of expectation is *linearity*: expected sum of RVs is the sum of expectations: $\forall X, Y \ \forall$ constant c:

$$E(X + Y) = E(X) + E(Y),$$
$$E(cX) = cE(X).$$

Averages can be calculated in two ways, *ungrouped* or *grouped*, is all that is needed to prove linearity!