Weighted mean

A *weighted mean* or *weighted average* of a set of data (or numbers) is a mean in which the different pieces of data are given different *weights*. "Weight" here has the sense of "importance", but can also be thought of as literal "weight".

The simplest case is where each piece of data x_i occurs with a corresponding frequency f_i, then we can think of the data as

$$x_1, \dots, x_1, x_2, \dots, x_2, \dots, x_n, \dots, x_n$$

with f_1 copies of x_1, and so on. Thus the arithmetic mean of these would be

$$\bar{x} = \frac{x_1 + \dots + x_1 + x_2 + \dots + x_2 + \dots + x_n + \dots + x_n}{f_1 + f_2 + \dots + f_n} = \frac{\sum f_i x_i}{\sum f_i}.$$

More generally, if we weight the data x_i with weight w_i, then the weighted arithmetic mean is $\dfrac{\sum w_i x_i}{\sum w_i}$.

This is the same formula as that for the centre of mass of a set of objects (where weight w_i is replaced by mass m_i): $\bar{x} = \dfrac{\sum m_i x_i}{\sum m_i}$, that is, the x-coordinate of the centre of mass is the mean of the x-coordinates of the objects, where the objects are weighted by their mass.

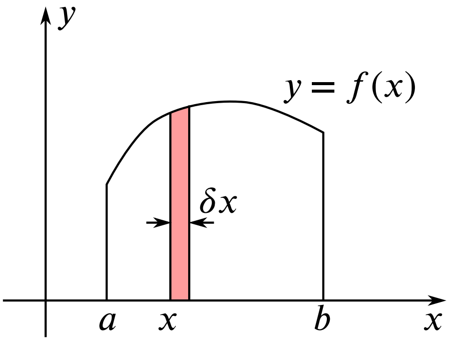
It is also the same formula as that for the expectation of a discrete random variable X. If X can take the values x_1, ..., x_n with probabilities p_1, ..., p_n respectively, where $p_1 + p_2 + \dots + p_n = 1$, then the expectation (mean) of X is given by $E(X) = p_1 x_1 + p_2 x_2 + \dots + p_n x_n = \sum p_i x_i$. (We don't need to explicitly divide by $p_1 + p_2 + \dots + p_n$ as this equals 1.)

Similarly, the weighted geometric mean of a set of data would be

$$\left(x_1^{w_1} x_2^{w_2} \dots x_n^{w_n}\right)^{1/(w_1 + w_2 + \dots + w_n)}.$$

Weighted means are used when calculating inflation: in the UK, the Retail Price Index (RPI) and Consumer Price Index (CPI) are calculated by first finding a weighted mean cost of a set of items in a typical shopping basket, weighted by the quantity in which they might typically be purchased. More can be found on these at the Office for National Statistics.

Weighted means can also be used for continuous mass distributions. For example, if a uniform lamina has the shape shown in this diagram:



then we can estimate the x-coordinate of its centre of mass by summing over the strips, one of which is highlighted in red:

$$\bar{x} \approx \frac{\sum(f(x)\delta x)x}{\sum(f(x)\delta x)}$$

since each strip has mass proportional to its area $f(x)\delta x$. In the limit as the strips become thinner, the sums become integrals, giving

$$\bar{x} = \frac{\int_a^b x f(x)\,dx}{\int_a^b f(x)\,dx}.$$

The same formula applies if we are finding the expectation (mean) of a continuous random variable, where this time $f(x)$ is the probability density function, so

$$E(X) = \int_a^b x f(x)\,dx.$$

(Again, we do not need to explicitly divide by the integral $\int_a^b f(x)\,dx$, as this equals 1.)