Mean/expected value/expectation of a discrete random variable/distribution

Let X be a discrete random variable defined on the probability space $(S, \mathcal{C}, P)$. Its range (set of possible values) is countable. The **mean** or **expected value** or **expectation** of X, denoted $E(X)$ (or sometimes just EX) or μ_X (or sometimes just μ), is defined by the sum

$$E(X) := \sum_{s \in S} X(s)P(\{s\}).$$

When the sum has a countable infinity of terms, we impose the technical requirement that the infinite series be absolutely convergent. In advanced calculus or real analysis it is shown that then the terms may be added in any order and will always yield the same sum. For a conditionally convergent series, this is not true!

Example. Let X be the number of heads one gets when spinning a penny three times. Let p be the probability of heads in one spin and $q = 1 - p$.

s	$X(s)$	$P(\{s\})$	$X(s)P(\{s\})$
HHH	3	ppp	$3p^3$
HHT	2	ppq	$2p^2q$
HTH	2	pqp	$2p^2q$
THH	2	qpp	$2p^2q$
HTT	1	pqq	$1pq^2$
THT	1	qpq	$1pq^2$
TTH	1	qqp	$1pq^2$
TTT	0	qqq	$0q^3$
Totals		1	$3p$

Thus $E(X) = 3p$. (Use $q = 1 - p$ and algebra to simplify the sum above.) In particular, when $p = 1/2$, $E(X) = 3/2$.

Observe that the sum in the final column is equal to $3P(X = 3) + 2P(X = 2) + 1P(X = 1) + 0P(X = 0)$. More generally,

$$E(X) = \sum_{s \in S} X(s)P(\{s\}) = \sum_{x} x \cdot \sum_{s : X(s) = x} P(\{s\}) = \sum_{x} xP(X = x).$$

Here $\sum_x$ denotes the sum over all x in the range of X.

Thus we have

$$E(X) = \sum_x xP(X = x).$$

Many texts give this as the definition of $E(X)$. This formula shows that the mean is the weighted average of possible values of X using their probabilities as weights.

Example. If X has the uniform distribution on $\{x_1, x_2, \ldots, x_n\}$, then

$$E(X) = \sum_{j=1}^{n} x_j P(X = x_j) = \sum_{j=1}^{n} x_j \cdot \frac{1}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \overline{x}$$

the (arithmetic) mean of $x_1, x_2, \ldots, x_n$.

The reasoning used to derive the alternative definition of $E(X)$ may also be used to prove the extremely useful

Law of the Unconscious Statistician (LotUS). If we define a new r.v. Y as a function of X, $Y = g(X)$, then

$$E(Y) = E(g(X)) = \sum_x g(x)P(X = x).$$

This theorem is named in honor of the statistician who thinks that this formula is a direct application of the definition of $E(Y)$ rather than a formula deduced from the definition. It has a generalization to functions of n random variables, $Y = g(X_1, \ldots, X_n)$:

$$E(Y) = \sum_{x_1,\ldots,x_n} g(x_1, \ldots x_n)P(X_1 = x_1, \ldots, X_n = x_n).$$

The special case $g(x_1, \ldots, x_n) = c_1x_1 + \cdots + c_nx_n$ where $c_1, \ldots, c_n$ are constants yields the following frequently used result.

Linearity: $E(c_1X_1 + \cdots + c_nX_n) = c_1E(X_1) + \cdots + c_nE(X_n)$.

Random variables are said to be **independent** if all events defined in terms of them individually are independent events. In particular, for discrete random variables X and Y, if X and Y are independent, then $P(X = x, Y = y) = P(X = x)P(Y = y)$ for every x, y. The converse is also true.

Theorem. If X and Y are independent random variables, then $E(XY) = E(X)E(Y)$