

# Expected value

From Wikipedia, the free encyclopedia

In probability theory, the expected value of a random variable, intuitively, is the long-run average value of repetitions of the experiment it represents. For example, the expected value in rolling a six-sided die is 3.5 because, roughly speaking, the average of all the numbers that come up in an extremely large number of rolls is very nearly always quite close to three and a half. Less roughly, the law of large numbers states that the arithmetic mean of the values almost surely converges to the expected value as the number of repetitions approaches infinity. The expected value is also known as the expectation, mathematical expectation, EV, average, mean value, mean, or first moment.

More practically, the expected value of a discrete random variable is the probability-weighted average of all possible values. In other words, each possible value the random variable can assume is multiplied by its probability of occurring, and the resulting products are summed to produce the expected value. The same principle applies to a continuous random variable, except that an integral of the variable with respect to its probability density replaces the sum. The formal definition subsumes both of these and also works for distributions which are neither discrete nor continuous: the expected value of a random variable is the integral of the random variable with respect to its probability measure.<sup>[1][2]</sup>

The expected value does not exist for random variables having some distributions with large "tails", such as the Cauchy distribution.<sup>[3]</sup> For random variables such as these, the long-tails of the distribution prevent the sum/integral from converging.

The expected value is a key aspect of how one characterizes a probability distribution; it is one type of location parameter. By contrast, the variance is a measure of dispersion of the possible values of the random variable around the expected value. The variance itself is defined in terms of two expectations: it is the expected value of the squared deviation of the variable's value from the variable's expected value.

The expected value plays important roles in a variety of contexts. In regression analysis, one desires a formula in terms of observed data that will give a "good" estimate of the parameter giving the effect of some explanatory variable upon a dependent variable. The formula will give different estimates using different samples of data, so the estimate it gives is itself a random variable. A formula is typically considered good in this context if it is an unbiased estimator—that is, if the expected value of the estimate (the average value it would give over an arbitrarily large number of separate samples) can be shown to equal the true value of the desired parameter.

In decision theory, and in particular in choice under uncertainty, an agent is described as making an optimal choice in the context of incomplete information. For risk neutral agents, the choice involves using the expected values of uncertain quantities, while for risk averse agents it involves maximizing the expected value of some objective function such as a von Neumann–Morgenstern utility function. One example of using expected value in reaching optimal decisions is the Gordon–Loeb model of information security investment. According to the model, one can conclude that the amount a firm spends to protect information should generally be only a small fraction of the expected loss (i.e., the expected value of the loss resulting from a cyber/information security breach).<sup>[4]</sup>

## Contents

- 1 Definition
  - 1.1 Univariate discrete random variable, finite case
  - 1.2 Univariate discrete random variable, countable case

- 1.3 Univariate continuous random variable
  - 1.4 General definition
- 2 Properties
  - 2.1 Constants
  - 2.2 Monotonicity
  - 2.3 Linearity
  - 2.4 Iterated expectation
    - 2.4.1 Iterated expectation for discrete random variables
    - 2.4.2 Iterated expectation for continuous random variables
  - 2.5 Inequality
  - 2.6 Non-multiplicativity
  - 2.7 Functional non-invariance
  - 2.8 Relation to characteristic function
- 3 Uses and applications
- 4 Expectation of matrices
- 5 Formulas for special cases
  - 5.1 Discrete distribution taking only non-negative integer values
  - 5.2 Continuous distribution taking non-negative values
- 6 History
- 7 See also
- 8 Notes
- 9 Literature

## Definition

### Univariate discrete random variable, finite case

Suppose random variable  $X$  can take value  $x_1$  with probability  $p_1$ , value  $x_2$  with probability  $p_2$ , and so on, up to value  $x_k$  with probability  $p_k$ . Then the expectation of this random variable  $X$  is defined as

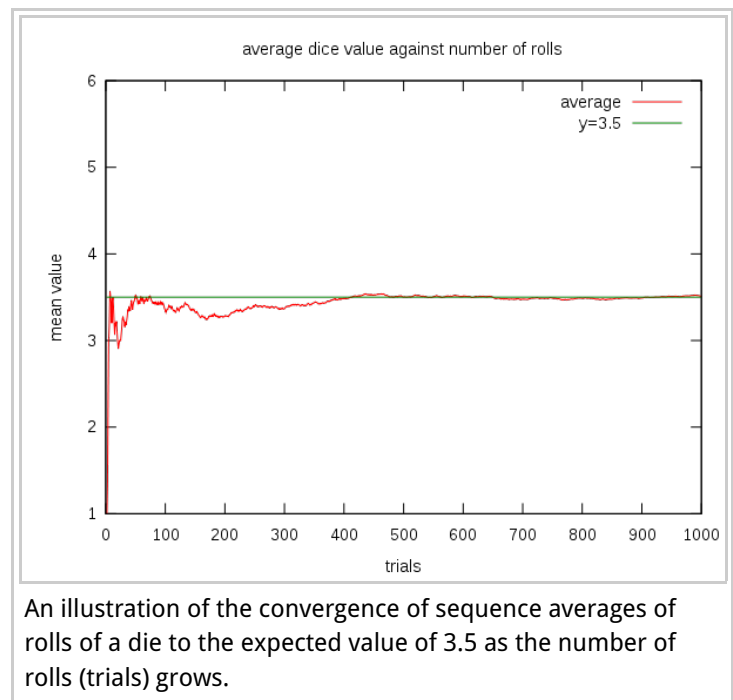
$$\mathbf{E}[X] = x_1p_1 + x_2p_2 + \cdots + x_kp_k .$$

Since all probabilities  $p_i$  add up to one ( $p_1 + p_2 + \dots + p_k = 1$ ), the expected value can be viewed as the weighted average, with  $p_i$ 's being the weights:

$$\mathbf{E}[X] = \frac{x_1p_1 + x_2p_2 + \cdots + x_kp_k}{1} = \frac{x_1p_1 + x_2p_2 + \cdots + x_kp_k}{p_1 + p_2 + \cdots + p_k} .$$

If all outcomes  $x_i$  are equally likely (that is,  $p_1 = p_2 = \dots = p_k$ ), then the weighted average turns into the simple average. This is intuitive: the expected value of a random variable is the average of all values it can take; thus the expected value is what one expects to happen *on average*. If the outcomes  $x_i$  are not equally probable, then the simple average must be replaced with the weighted average, which takes into account the fact that some outcomes are more likely than the others. The intuition however remains the same: the expected value of  $X$  is what one expects to happen *on average*.

Example 1. Let  $X$  represent the outcome of a roll of a fair six-sided die. More specifically,  $X$  will be the number of pips showing on the top face of the die after the toss. The possible values for  $X$  are 1, 2, 3, 4, 5, and 6, all equally likely (each having the probability of  $\frac{1}{6}$ ). The expectation of  $X$  is



$$\mathbf{E}[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5.$$

If one rolls the die  $n$  times and computes the average (arithmetic mean) of the results, then as  $n$  grows, the average will almost surely converge to the expected value, a fact known as the strong law of large numbers. One example sequence of ten rolls of the die is 2, 3, 1, 2, 5, 6, 2, 2, 2, 6, which has the average of 3.1, with the distance of 0.4 from the expected value of 3.5. The convergence is relatively slow: the probability that the average falls within the range  $3.5 \pm 0.1$  is 21.6% for ten rolls, 46.1% for a hundred rolls and 93.7% for a thousand rolls. See the figure for an illustration of the averages of longer sequences of rolls of the die and how they converge to the expected value of 3.5. More generally, the rate of convergence can be roughly quantified by e.g. Chebyshev's inequality and the Berry–Esseen theorem.

Example 2. The roulette game consists of a small ball and a wheel with 38 numbered pockets around the edge. As the wheel is spun, the ball bounces around randomly until it settles down in one of the pockets. Suppose random variable  $X$  represents the (monetary) outcome of a \$1 bet on a single number ("straight up" bet). If the bet wins (which happens with probability  $\frac{1}{38}$  in American roulette), the payoff is \$35; otherwise the player loses the bet. The expected profit from such a bet will be

$$\mathbf{E}[\text{gain from \$1 bet}] = -\$1 \cdot \frac{37}{38} + \$35 \cdot \frac{1}{38} = -\$0.0526.$$

i.e. the bet of \$1 stands to lose \$0.0526, so its expected value is \$0.9474.

### Univariate discrete random variable, countable case

Let  $X$  be a discrete random variable taking values  $x_1, x_2, \dots$  with probabilities  $p_1, p_2, \dots$  respectively. Then the expected value of this random variable is the infinite sum

$$\mathbf{E}[X] = \sum_{i=1}^{\infty} x_i p_i,$$

provided that this series converges absolutely (that is, the sum must remain finite if we were to replace all  $x_i$ 's with their absolute values). If this series does not converge absolutely, we say that the expected value of  $X$  does not exist.

For example, suppose random variable  $X$  takes values  $1, -2, 3, -4, \dots$ , with respective probabilities  $\frac{c}{1^2}, \frac{c}{2^2}, \frac{c}{3^2}, \frac{c}{4^2}, \dots$ , where  $c = \frac{6}{\pi^2}$  is a normalizing constant that ensures the probabilities sum up to one. Then the infinite sum

$$\sum_{i=1}^{\infty} x_i p_i = c \left( 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots \right)$$

converges and its sum is equal to  $\ln(2) \approx 0.69315$ . However it would be incorrect to claim that the expected value of  $X$  is equal to this number—in fact  $E[X]$  does not exist, as this series does not converge absolutely (see harmonic series).

### Univariate continuous random variable

If the probability distribution of  $X$  admits a probability density function  $f(x)$ , then the expected value can be computed as

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx.$$

### General definition

In general, if  $X$  is a random variable defined on a probability space  $(\Omega, \Sigma, P)$ , then the expected value of  $X$ , denoted by  $E[X]$ ,  $\langle X \rangle$ ,  $X$  or  $E[X]$ , is defined as the Lebesgue integral

$$E[X] = \int_{\Omega} X dP = \int_{\Omega} X(\omega) P(d\omega)$$

When this integral exists, it is defined as the expectation of  $X$ . Not all random variables have a finite expected value, since the integral may not converge absolutely; furthermore, for some it is not defined at all (e.g., Cauchy distribution). Two variables with the same probability distribution will have the same expected value, if it is defined.

It follows directly from the discrete case definition that if  $X$  is a constant random variable, i.e.  $X = b$  for some fixed real number  $b$ , then the expected value of  $X$  is also  $b$ .

The expected value of a measurable function of  $X$ ,  $g(X)$ , given that  $X$  has a probability density function  $f(x)$ , is given by the inner product of  $f$  and  $g$ :

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx.$$

This is sometimes called the law of the unconscious statistician. Using representations as Riemann–Stieltjes integral and integration by parts the formula can be restated as

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) dP(X \leq x) =$$

$$\begin{cases} g(a) + \int_a^\infty g'(x)P(X > x) \, dx & \text{if } P(g(X) \geq g(a)) = 1 \\ g(b) - \int_{-\infty}^b g'(x)P(X \leq x) \, dx & \text{if } P(g(X) \leq g(b)) = 1. \end{cases}$$

As a special case let  $\alpha$  denote a positive real number. Then

$$\mathbf{E}[|X|^\alpha] = \alpha \int_0^\infty t^{\alpha-1} P(|X| > t) \, dt.$$

In particular, if  $\alpha = 1$  and  $\Pr[X \geq 0] = 1$ , then this reduces to

$$\mathbf{E}[|X|] = \mathbf{E}[X] = \int_0^\infty \{1 - F(t)\} \, dt,$$

where  $F$  is the cumulative distribution function of  $X$ . This last identity is an instance of what, in a non-probabilistic setting, has been called the layer cake representation.

The law of the unconscious statistician applies also to a measurable function  $g$  of several random variables  $X_1, \dots, X_n$  having a joint density  $f$ .<sup>[5]</sup> <sup>[6]</sup>

$$\mathbf{E}[g(X_1, \dots, X_n)] = \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty g(x_1, \dots, x_n) f(x_1, \dots, x_n) \, dx_1 \cdots dx_n.$$

## Properties

### Constants

The expected value of a constant is equal to the constant itself; i.e., if  $c$  is a constant, then  $\mathbf{E}[c] = c$ .

### Monotonicity

If  $\mathbf{X}$  and  $\mathbf{Y}$  are random variables such that  $\mathbf{X} \leq \mathbf{Y}$  almost surely, then  $\mathbf{E}[\mathbf{X}] \leq \mathbf{E}[\mathbf{Y}]$ .

### Linearity

The expected value operator (or expectation operator)  $\mathbf{E}[\cdot]$  is linear in the sense that

$$\begin{aligned} \mathbf{E}[\mathbf{X} + c] &= \mathbf{E}[\mathbf{X}] + c \\ \mathbf{E}[\mathbf{X} + \mathbf{Y}] &= \mathbf{E}[\mathbf{X}] + \mathbf{E}[\mathbf{Y}] \\ \mathbf{E}[a\mathbf{X}] &= a \mathbf{E}[\mathbf{X}] \end{aligned}$$

The second result is valid even if  $\mathbf{X}$  is not statistically independent of  $\mathbf{Y}$ . Combining the results from the previous three equations, we can see that

$$\mathbf{E}[a\mathbf{X} + b\mathbf{Y} + c] = a \mathbf{E}[\mathbf{X}] + b \mathbf{E}[\mathbf{Y}] + c$$

for any two random variables  $\mathbf{X}$  and  $\mathbf{Y}$  (which need to be defined on the same probability space) and any real numbers  $a$ ,  $b$  and  $c$ .

### Iterated expectation

## Iterated expectation for discrete random variables

For any two discrete random variables  $X, Y$  one may define the conditional expectation:<sup>[7]</sup>

$$\mathbf{E}[X \mid Y = y] = \sum_x x \cdot \mathbf{P}(X = x \mid Y = y),$$

which means that  $\mathbf{E}[X \mid Y = y]$  is a function of  $y$ . Let  $g(y)$  be that function of  $y$ , then the notation  $\mathbf{E}[X \mid Y]$  is a random variable in its own right, equal to  $g(Y)$ .

Lemma. Then the expectation of  $X$  satisfies:<sup>[8]</sup>

$$\mathbf{E}[X] = \mathbf{E}[\mathbf{E}[X \mid Y]].$$

Proof.

$$\begin{aligned} \mathbf{E}[\mathbf{E}[X \mid Y]] &= \sum_y \mathbf{E}[X \mid Y = y] \cdot \mathbf{P}(Y = y) \\ &= \sum_y \left( \sum_x x \cdot \mathbf{P}(X = x \mid Y = y) \right) \cdot \mathbf{P}(Y = y) \\ &= \sum_y \sum_x x \cdot \mathbf{P}(X = x \mid Y = y) \cdot \mathbf{P}(Y = y) \\ &= \sum_y \sum_x x \cdot \mathbf{P}(Y = y \mid X = x) \cdot \mathbf{P}(X = x) \\ &= \sum_x x \cdot \mathbf{P}(X = x) \cdot \left( \sum_y \mathbf{P}(Y = y \mid X = x) \right) \\ &= \sum_x x \cdot \mathbf{P}(X = x) \\ &= \mathbf{E}[X] \end{aligned}$$

The left-hand side of this equation is referred to as the *iterated expectation*. The equation is sometimes called the *tower rule* or the *tower property*; it is treated under law of total expectation.

## Iterated expectation for continuous random variables

In the continuous case, the results are completely analogous. The definition of conditional expectation would use inequalities, density functions, and integrals to replace equalities, mass functions, and summations, respectively. However, the main result still holds:

$$\mathbf{E}[X] = \mathbf{E}[\mathbf{E}[X \mid Y]]$$

## Inequality

If a random variable  $X$  is always less than or equal to another random variable  $Y$ , the expectation of  $X$  is less than or equal to that of  $Y$ :

$$\text{If } X \leq Y, \text{ then } \mathbf{E}[X] \leq \mathbf{E}[Y].$$

In particular, if we set  $Y$  to  $|X|$  we know  $X \leq |X|$  and  $-X \leq |X|$ . Therefore, we know  $\mathbf{E}[X] \leq \mathbf{E}[|X|]$  and  $\mathbf{E}[-X] \leq \mathbf{E}[|X|]$ .

$\leq E[Y]$ . From the linearity of expectation we know  $-E[X] \leq E[Y]$ . Therefore, the absolute value of expectation of a random variable is less than or equal to the expectation of its absolute value:

$$|E[X]| \leq E[|X|]$$

This is a special case of Jensen's inequality.

### Non-multiplicativity

If one considers the joint probability density function of  $X$  and  $Y$ , say  $j(x,y)$ , then the expectation of  $XY$  is

$$E[XY] = \iint xy j(x, y) dx dy.$$

In general, the expected value operator is not multiplicative, i.e.  $E[XY]$  is not necessarily equal to  $E[X] \cdot E[Y]$ . In fact, the amount by which multiplicativity fails is called the covariance:

$$\text{Cov}(X, Y) = E[XY] - E[X] E[Y].$$

Thus multiplicativity holds precisely when  $\text{Cov}(X, Y) = 0$ , in which case  $X$  and  $Y$  are said to be uncorrelated (independent variables are a notable case of uncorrelated variables).

Now if  $X$  and  $Y$  are independent, then by definition  $j(x,y) = f(x)g(y)$  where  $f$  and  $g$  are the marginal PDFs for  $X$  and  $Y$ . Then

$$\begin{aligned} E[XY] &= \iint xy j(x, y) dx dy = \iint xy f(x)g(y) dy dx \\ &= \left[ \int x f(x) dx \right] \left[ \int y g(y) dy \right] = E[X] E[Y] \end{aligned}$$

and  $\text{Cov}(X, Y) = 0$ .

Observe that independence of  $X$  and  $Y$  is required only to write  $j(x, y) = f(x)g(y)$ , and this is required to establish the second equality above. The third equality follows from a basic application of the Fubini-Tonelli theorem.

### Functional non-invariance

In general, with the exception of linear functions, the expectation operator and functions of random variables do not commute; that is

$$E[g(X)] = \int_{\Omega} g(X) dP \neq g(E[X]).$$

A notable inequality concerning this topic is Jensen's inequality, involving expected values of convex (or concave) functions.

### Relation to characteristic function

The probability density function of a scalar random variable  $X$  is related to its characteristic function  $\varphi_X$  by the inversion formula:

$$f_X(x) = \frac{1}{2\pi} \int_{\mathbf{R}} e^{-itx} \varphi_X(t) dt.$$

We can use this inversion formula in expected value of a function  $g(X)$  to obtain

$$\mathbf{E}[g(X)] = \frac{1}{2\pi} \int \int g(x) e^{-itx} \varphi_X(t) dt dx$$

Changing the order of integration, we get

$$\mathbf{E}[g(X)] = \frac{1}{2\pi} \int G(t) \varphi_X(t) dt,$$

where

$$G(t) = \int g(x) e^{-itx} dx$$

is the Fourier transform of  $g(x)$ . This can also be seen as a direct consequence of Plancherel theorem.

## Uses and applications

It is possible to construct an expected value equal to the probability of an event by taking the expectation of an indicator function that is one if the event has occurred and zero otherwise. This relationship can be used to translate properties of expected values into properties of probabilities, e.g. using the law of large numbers to justify estimating probabilities by frequencies.

The expected values of the powers of  $X$  are called the moments of  $X$ ; the moments about the mean of  $X$  are expected values of powers of  $X - \mathbf{E}[X]$ . The moments of some random variables can be used to specify their distributions, via their moment generating functions.

To empirically estimate the expected value of a random variable, one repeatedly measures observations of the variable and computes the arithmetic mean of the results. If the expected value exists, this procedure estimates the true expected value in an unbiased manner and has the property of minimizing the sum of the squares of the residuals (the sum of the squared differences between the observations and the estimate). The law of large numbers demonstrates (under fairly mild conditions) that, as the size of the sample gets larger, the variance of this estimate gets smaller.

This property is often exploited in a wide variety of applications, including general problems of statistical estimation and machine learning, to estimate (probabilistic) quantities of interest via Monte Carlo methods, since most quantities of interest can be written in terms of expectation, e.g.

$\mathbf{P}(X \in \mathcal{A}) = \mathbf{E}[I_{\mathcal{A}}(X)]$  where  $I_{\mathcal{A}}(X)$  is the indicator function for set  $\mathcal{A}$ , i.e.

$X \in \mathcal{A} \rightarrow I_{\mathcal{A}}(X) = 1, X \notin \mathcal{A} \rightarrow I_{\mathcal{A}}(X) = 0.$

In classical mechanics, the center of mass is an analogous concept to expectation. For example, suppose  $X$  is a discrete random variable with values  $x_i$  and corresponding probabilities  $p_i$ . Now consider a weightless rod on which are placed weights, at locations  $x_i$  along the rod and having masses  $p_i$  (whose sum is one). The point at which the rod balances is  $\mathbf{E}[X]$ .

Expected values can also be used to compute the variance, by means of the computational formula for the variance



$$\mathbf{Var}(X) = \mathbf{E}[X^2] - (\mathbf{E}[X])^2.$$

A very important application of the expectation value is in the field of quantum mechanics. The expectation value of a quantum mechanical operator  $\hat{A}$  operating on a quantum state vector  $|\psi\rangle$  is written as  $\langle\hat{A}\rangle = \langle\psi|A|\psi\rangle$ . The uncertainty in  $\hat{A}$  can be calculated using the formula  $(\Delta A)^2 = \langle\hat{A}^2\rangle - \langle\hat{A}\rangle^2$ .

## Expectation of matrices

If  $X$  is an  $m \times n$  matrix, then the expected value of the matrix is defined as the matrix of expected values:

$$\mathbf{E}[X] = \mathbf{E} \left[ \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,n} \end{pmatrix} \right] = \begin{pmatrix} \mathbf{E}[x_{1,1}] & \mathbf{E}[x_{1,2}] & \cdots & \mathbf{E}[x_{1,n}] \\ \mathbf{E}[x_{2,1}] & \mathbf{E}[x_{2,2}] & \cdots & \mathbf{E}[x_{2,n}] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{E}[x_{m,1}] & \mathbf{E}[x_{m,2}] & \cdots & \mathbf{E}[x_{m,n}] \end{pmatrix}.$$

This is utilized in covariance matrices.

## Formulas for special cases

### Discrete distribution taking only non-negative integer values

When a random variable takes only values in  $\{0, 1, 2, 3, \dots\}$  we can use the following formula for computing its expectation (even when the expectation is infinite):

$$\mathbf{E}[X] = \sum_{i=1}^{\infty} P(X \geq i).$$

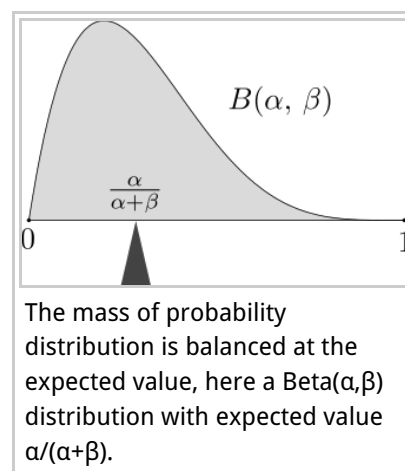
Proof.

$$\sum_{i=1}^{\infty} P(X \geq i) = \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} P(X = j).$$

Interchanging the order of summation, we have

$$\begin{aligned} \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} P(X = j) &= \sum_{j=1}^{\infty} \sum_{i=1}^j P(X = j) \\ &= \sum_{j=1}^{\infty} j P(X = j) \\ &= \mathbf{E}[X]. \end{aligned}$$

This result can be a useful computational shortcut. For example, suppose we toss a coin where the probability of heads is  $p$ . How many tosses can we expect until the first heads (not including the heads



itself)? Let  $X$  be this number. We are counting only the tails and not the heads which ends the experiment; in particular, we can have  $X = 0$ . The expectation of  $X$  may be computed by

$\sum_{i=1}^{\infty} (1-p)^i = \frac{1}{p} - 1$ . This is because, when the first  $i$  tosses yield tails, the number of tosses is at least  $i$ . The last equality used the formula for a geometric progression,

$$\sum_{i=1}^{\infty} r^i = \sum_{i=0}^{\infty} r^i - 1 = \frac{1}{1-r} - 1, \text{ where } r = 1-p < 1.$$

## Continuous distribution taking non-negative values

Analogously with the discrete case above, when a continuous random variable  $X$  takes only non-negative values, we can use the following formula for computing its expectation (even when the expectation is infinite):

$$\mathbf{E}[X] = \int_0^{\infty} P(X \geq x) \, dx$$

Proof: It is first assumed that  $X$  has a density  $f_X(x)$ . We present two techniques:

- Using integration by parts (a special case of Section 1.4 above):

$$\mathbf{E}[X] = \int_0^{\infty} (-x)(-f_X(x)) \, dx = \left[ -x(1-F(x)) \right]_0^{\infty} + \int_0^{\infty} (1-F(x)) \, dx$$

and the bracket vanishes because (see Cumulative distribution function#Derived functions)

$$1 - F(x) = o\left(\frac{1}{x}\right) \text{ as } x \rightarrow \infty.$$

- Using an interchange in order of integration:

$$\begin{aligned} \int_0^{\infty} P(X \geq x) \, dx &= \int_0^{\infty} \int_x^{\infty} f_X(t) \, dt \, dx \\ &= \int_0^{\infty} \int_0^t f_X(t) \, dx \, dt = \int_0^{\infty} t f_X(t) \, dt = \mathbf{E}[X] \end{aligned}$$

In case no density exists, it is seen that

$$\mathbf{E}[X] = \int_0^{\infty} \int_0^x dt \, dF(x) = \int_0^{\infty} \int_t^{\infty} dF(x) dt = \int_0^{\infty} (1-F(t)) \, dt.$$

If both negative and non-negative values are taken for  $X$ , then its expectation is

$$\mathbf{E}[X] = \int_0^{\infty} P(X \geq x) \, dx - \int_{-\infty}^0 P(X \leq x) \, dx$$

## History

The idea of the expected value originated in the middle of the 17th century from the study of the so-called problem of points, which seeks to divide the stakes *in a fair way* between two players who

have to end their game before it's properly finished. This problem had been debated for centuries, and many conflicting proposals and solutions had been suggested over the years, when it was posed in 1654 to Blaise Pascal by French writer and amateur mathematician Chevalier de Méré. de Méré claimed that this problem couldn't be solved and that it showed just how flawed mathematics was when it came to its application to the real world. Pascal, being a mathematician, was provoked and determined to solve the problem once and for all. He began to discuss the problem in a now famous series of letters to Pierre de Fermat. Soon enough they both independently came up with a solution. They solved the problem in different computational ways but their results were identical because their computations were based on the same fundamental principle. The principle is that the value of a future gain should be directly proportional to the chance of getting it. This principle seemed to have come naturally to both of them. They were very pleased by the fact that they had found essentially the same solution and this in turn made them absolutely convinced they had solved the problem conclusively. However, they did not publish their findings. They only informed a small circle of mutual scientific friends in Paris about it.<sup>[9]</sup>

Three years later, in 1657, a Dutch mathematician Christiaan Huygens, who had just visited Paris, published a treatise (see Huygens (1657)) "*De ratiociniis in ludo aleæ*" on probability theory. In this book he considered the problem of points and presented a solution based on the same principle as the solutions of Pascal and Fermat. Huygens also extended the concept of expectation by adding rules for how to calculate expectations in more complicated situations than the original problem (e.g., for three or more players). In this sense this book can be seen as the first successful attempt of laying down the foundations of the theory of probability.

In the foreword to his book, Huygens wrote: "It should be said, also, that for some time some of the best mathematicians of France have occupied themselves with this kind of calculus so that no one should attribute to me the honour of the first invention. This does not belong to me. But these savants, although they put each other to the test by proposing to each other many questions difficult to solve, have hidden their methods. I have had therefore to examine and go deeply for myself into this matter by beginning with the elements, and it is impossible for me for this reason to affirm that I have even started from the same principle. But finally I have found that my answers in many cases do not differ from theirs." (cited by Edwards (2002)). Thus, Huygens learned about de Méré's Problem in 1655 during his visit to France; later on in 1656 from his correspondence with Carcavi he learned that his method was essentially the same as Pascal's; so that before his book went to press in 1657 he knew about Pascal's priority in this subject.

Neither Pascal nor Huygens used the term "expectation" in its modern sense. In particular, Huygens writes: "That my Chance or Expectation to win any thing is worth just such a Sum, as wou'd procure me in the same Chance and Expectation at a fair Lay. ... If I expect a or b, and have an equal Chance of gaining them, my Expectation is worth  $\frac{a+b}{2}$ ." More than a hundred years later, in 1814, Pierre-Simon Laplace published his tract "*Théorie analytique des probabilités*", where the concept of expected value was defined explicitly:

... this advantage in the theory of chance is the product of the sum hoped for by the probability of obtaining it; it is the partial sum which ought to result when we do not wish to run the risks of the event in supposing that the division is made proportional to the probabilities. This division is the only equitable one when all strange circumstances are eliminated; because an equal degree of probability gives an equal right for the sum hoped for. We will call this advantage *mathematical hope*.

The use of the letter *E* to denote expected value goes back to W.A. Whitworth in 1901,<sup>[10]</sup> who used a script *E*. The symbol has become popular since for English writers it meant "Expectation", for Germans "Erwartungswert", for Spanish "Esperanza matemática" and for French "Espérance mathématique".<sup>[11]</sup>

## See also

- Center of mass
- Central tendency
- Chebyshev's inequality (an inequality on location and scale parameters)
- Conditional expectation
- Expected value is also a key concept in economics, finance, and many other subjects
- The general term expectation
- Expectation value (quantum mechanics)
- Law of total expectation –the expected value of the conditional expected value of  $X$  given  $Y$  is the same as the expected value of  $X$ .
- Moment (mathematics)
- Nonlinear expectation a generalization of the expected value
- Wald's equation for calculating the expected value of a random number of random variables

## Notes

- Sheldon M Ross (2007). "§2.4 Expectation of a random variable". *Introduction to probability models* (9th ed.). Academic Press. p. 38 ff. ISBN 0-12-598062-0.
- Richard W Hamming (1991). "§2.5 Random variables, mean and the expected value". *The art of probability for scientists and engineers*. Addison–Wesley. p. 64 ff. ISBN 0-201-40686-1.
- Richard W Hamming (1991). "Example 8.7–1 The Cauchy distribution". *The art of probability for scientists and engineers*. Addison-Wesley. p. 290 ff. ISBN 0-201-40686-1. "Sampling from the Cauchy distribution and averaging gets you nowhere — one sample has the same distribution as the average of 1000 samples!"
- Gordon, Lawrence; Loeb, Martin (November 2002). "The Economics of Information Security Investment". *ACM Transactions on Information and System Security*. 5 (4): 438–457. doi:10.1145/581271.581274.
- Expectation Value*, retrieved October 2013 Check date values in: |access-date= (help)
- Papoulis, A. (1984), *Probability, Random Variables, and Stochastic Processes*, New York: McGraw–Hill, pp. 139–152
- Sheldon M Ross. "Chapter 3: Conditional probability and conditional expectation". *cited work*. p. 97 ff. ISBN 0-12-598062-0.
- Sheldon M Ross. "§3.4: Computing expectations by conditioning". *cited work*. p. 105 ff. ISBN 0-12-598062-0.
- "Ore, Pascal and the Invention of Probability Theory". *The American Mathematical Monthly*. 67 (5): 409–419. 1960. doi:10.2307/2309286.
- Whitworth, W.A. (1901) *Choice and Chance with One Thousand Exercises*. Fifth edition. Deighton Bell, Cambridge. [Reprinted by Hafner Publishing Co., New York, 1959.]
- "Earliest uses of symbols in probability and statistics".

## Literature

- Edwards, A.W.F (2002). *Pascal's arithmetical triangle: the story of a mathematical idea* (2nd ed.). JHU Press. ISBN 0-8018-6946-3.
- Huygens, Christiaan (1657). *De ratiociniis in ludo aleæ* (English translation, published in 1714:).

Retrieved from "https://en.wikipedia.org/w/index.php?title=Expected\_value&oldid=738279080"

Categories: Theory of probability distributions | Gambling terminology

- 
- This page was last modified on 8 September 2016, at 00:32.

- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.