

4.9: Expected Value as an Integral

In the introductory section, we defined expected value separately for discrete, continuous, and mixed distributions, using density functions. In the section on additional properties, we showed how these definitions can be unified, by first defining expected value for nonnegative random variables in terms of the right-tail distribution function. However, by far the best and most elegant definition of expected value is as an integral with respect to the underlying probability measure. This definition and a review of the properties of expected value are the goals of this section. No proofs are necessary (you will be happy to know), since all of the results follow from the general theory of integration. However, to understand the exposition, you will need to review the advanced sections on the integral with respect to a positive measure and the properties of the integral. If you are a new student of probability, or are not interested in the measure-theoretic detail of the subject, you can safely skip this section.

Definitions

As usual, our starting point is a random experiment modeled by a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. So Ω is the set of outcomes, \mathcal{F} is the σ -algebra of events, and \mathbb{P} is the probability measure on the sample space (Ω, \mathcal{F}) .

Recall that a random variable X for the experiment is simply a measurable function from (Ω, \mathcal{F}) into another measurable space (S, \mathcal{S}) . When $S \subseteq \mathbb{R}^n$, we assume that S is Lebesgue measurable, and we take \mathcal{S} to be the σ -algebra of Lebesgue measurable subsets of S . As noted above, here is the measure-theoretic definition:

If X is a real-valued random variable on the probability space, the expected value of X is defined as the integral of X with respect to \mathbb{P} , assuming that the integral exists:

$$\mathbb{E}(X) = \int_{\Omega} X d\mathbb{P} \quad (4.9.1)$$

Let's review how the integral is defined in stages, but now using the notation of probability theory.

Let S denote the support set of X , so that S is a measurable subset of \mathbb{R} .

1. If S is finite, then $\mathbb{E}(X) = \sum_{x \in S} x \mathbb{P}(X = x)$.
2. If $S \subseteq [0, \infty)$, then $\mathbb{E}(X) = \sup \{ \mathbb{E}(Y) : Y \text{ has finite range and } 0 \leq Y \leq X \}$
3. For general $S \subseteq \mathbb{R}$, $\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-)$ as long as the right side is not of the form $\infty - \infty$, and where X^+ and X^- denote the positive and negative parts of X .
4. If $A \in \mathcal{F}$, then $\mathbb{E}(X; A) = \mathbb{E}(X \mathbf{1}_A)$, assuming that the expected value on the right exists.

Thus, as with integrals generally, an expected value can exist as a number in \mathbb{R} (in which case X is *integrable*), can exist as ∞ or $-\infty$, or can fail to exist. In reference to part (a), a random variable with a finite set of values in \mathbb{R} is a *simple function* in the terminology of general integration. In reference to part (b), note that the expected value of a nonnegative random variable always exists in $[0, \infty]$. In reference to part (c), $\mathbb{E}(X)$ exists if and only if either $\mathbb{E}(X^+) < \infty$ or $\mathbb{E}(X^-) < \infty$.

Our next goal is to restate the basic theorems and properties of integrals, but in the notation of probability. Unless otherwise noted, all random variables are assumed to be real-valued.

Basic Properties

The Linear Properties

Perhaps the most important and basic properties are the *linear properties*. Part (a) is the *additive property* and part (b) is the *scaling property*.

Suppose that X and Y are random variables whose expected values exist, and that $c \in \mathbb{R}$. Then

1. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ as long as the right side is not of the form $\infty - \infty$.
2. $\mathbb{E}(cX) = c\mathbb{E}(X)$

Thus, part (a) holds if at least one of the expected values on the right is finite, or if both are ∞ , or if both are $-\infty$. What is ruled out are the two cases where one expected value is ∞ and the other is $-\infty$, and this is what is meant by the indeterminate form $\infty - \infty$.

Equality and Order

Our next set of properties deal with equality and order. First, the expected value of a random variable over a null set is 0.

If X is a random variable and A is an event with $\mathbb{P}(A) = 0$. Then $\mathbb{E}(X; A) = 0$.

Random variables that are equivalent have the same expected value

If X is a random variable whose expected value exists, and Y is a random variable with $\mathbb{P}(X = Y) = 1$, then $\mathbb{E}(X) = \mathbb{E}(Y)$.

Our next result is the *positive property* of expected value.

Suppose that X is a random variable and $\mathbb{P}(X \geq 0) = 1$. Then

1. $\mathbb{E}(X) \geq 0$
2. $\mathbb{E}(X) = 0$ if and only if $\mathbb{P}(X = 0) = 1$.

So, if X is a nonnegative random variable then $\mathbb{E}(X) > 0$ if and only if $\mathbb{P}(X > 0) > 0$. The next result is the *increasing property* of expected value, perhaps the most important property after linearity.

Suppose that X, Y are random variables whose expected values exist, and that $\mathbb{P}(X \leq Y) = 1$. Then

1. $\mathbb{E}(X) \leq \mathbb{E}(Y)$
2. Except in the case that both expected values are ∞ or both $-\infty$, $\mathbb{E}(X) = \mathbb{E}(Y)$ if and only if $\mathbb{P}(X = Y) = 1$.

So if $X \leq Y$ with probability 1 then, except in the two cases mentioned, $\mathbb{E}(X) < \mathbb{E}(Y)$ if and only if $\mathbb{P}(X < Y) > 0$. The next result is the *absolute value inequality*.

Suppose that X is a random variable whose expected value exists. Then

1. $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$
2. If $\mathbb{E}(X)$ is finite, then equality holds in (a) if and only if $\mathbb{P}(X \geq 0) = 1$ or $\mathbb{P}(X \leq 0) = 1$.

Change of Variables and Density Functions

The Change of Variables Theorem

Suppose now that X is a general random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, taking values in a measurable space (S, \mathcal{S}) . Recall that the *probability distribution* of X is the probability measure P on (S, \mathcal{S}) given by $P(A) = \mathbb{P}(X \in A)$ for $A \in \mathcal{S}$. This is a special case of a new positive measure induced by a given positive measure and a measurable function. If $g: S \rightarrow \mathbb{R}$ is measurable, then $g(X)$ is a real-valued random variable. The following result shows how to compute the expected value of $g(X)$ as an integral with respect to the distribution of X , and is known as the *change of variables theorem*.

If $g: S \rightarrow \mathbb{R}$ is measurable then, assuming that the expected value exists,

$$\mathbb{E}[g(X)] = \int_S g(x) dP(x) \quad (4.9.2)$$

So, using the original definition and the change of variables theorem, and giving the variables explicitly for emphasis, we have

$$\mathbb{E}[g(X)] = \int_{\Omega} g[X(\omega)] d\mathbb{P}(\omega) = \int_S g(x) dP(x) \quad (4.9.3)$$

The Radon-Nikodym Theorem

Suppose now μ is a positive measure on (S, \mathcal{S}) , and that the distribution of X is absolutely continuous with respect to μ . Recall that this means that $\mu(A) = 0$ implies $P(A) = \mathbb{P}(X \in A) = 0$ for $A \in \mathcal{S}$. By the *Radon-Nikodym theorem*, named for Johann Radon and Otto Nikodym, X has a *probability density function* f with respect to μ . That is,

$$P(A) = \mathbb{P}(X \in A) = \int_A f d\mu, \quad A \in \mathcal{S} \quad (4.9.4)$$

In this case, we can write the expected value of $g(X)$ as an integral with respect to the probability density function.

If $g: S \rightarrow \mathbb{R}$ is measurable then, assuming that the expected value exists,

$$\mathbb{E}[g(X)] = \int_S g f d\mu \quad (4.9.5)$$

Again, giving the variables explicitly for emphasis, we have the following chain of integrals:

$$\mathbb{E}[g(X)] = \int_{\Omega} g[X(\omega)] d\mathbb{P}(\omega) = \int_S g(x) dP(x) = \int_S g(x) f(x) d\mu(x) \quad (4.9.6)$$

There are two critically important special cases.

Discrete Distributions

Suppose first that $(S, \mathcal{S}, \#)$ is a *discrete measure space*, so that S is countable, $\mathcal{S} = \mathcal{P}(S)$ is the collection of all subsets of S , and $\#$ is counting measure on (S, \mathcal{S}) . Thus, X has a discrete distribution on S , and this distribution is always absolutely continuous with respect to $\#$. Specifically, $\#(A) = 0$ if and only if $A = \emptyset$ and of course $\mathbb{P}(X \in \emptyset) = 0$. The probability density function f of X with respect to $\#$, as we know, is simply $f(x) = \mathbb{P}(X = x)$ for $x \in S$. Moreover, integrals with respect to $\#$ are sums, so

$$\mathbb{E}[g(X)] = \sum_{x \in S} g(x) f(x) \quad (4.9.7)$$

assuming that the expected value exists. Existence in this case means that either the sum of the positive terms is finite or the sum of the negative terms is finite, so that the sum makes sense (and in particular does not depend on the order in which the terms are added). Specializing further, if X itself is real-valued and $g = 1$ we have

$$\mathbb{E}(X) = \sum_{x \in S} x f(x) \quad (4.9.8)$$

which was our original definition of expected value in the discrete case.

Continuous Distributions

For the second special case, suppose that $(S, \mathcal{S}, \lambda_n)$ is a *Euclidean measure space*, so that S is a Lebesgue measurable subset of \mathbb{R}^n for some $n \in \mathbb{N}_+$, \mathcal{S} is the σ -algebra of Lebesgue measurable subsets of S , and λ_n is Lebesgue measure on (S, \mathcal{S}) . The distribution of X is absolutely continuous with respect to λ_n if $\lambda_n(A) = 0$ implies $\mathbb{P}(X \in A) = 0$ for $A \in \mathcal{S}$. If this is the case, then a probability density function f of X has its usual meaning. Thus,

$$\mathbb{E}[g(X)] = \int_S g(x)f(x) d\lambda_n(x) \quad (4.9.9)$$

assuming that the expected value exists. When g is a typically nice function, this integral reduces to an ordinary n -dimensional Riemann integral of calculus. Specializing further, if X is itself real-valued and $g = 1$ then

$$\mathbb{E}(X) = \int_S x f(x) dx \quad (4.9.10)$$

which was our original definition of expected value in the continuous case.

Interchange Properties

In this subsection, we review properties that allow the interchange of expected value and other operations: limits of sequences, infinite sums, and integrals. We assume again that the random variables are real-valued unless otherwise specified.

Limits

Our first set of convergence results deals with the interchange of expected value and limits. We start with the expected value version of *Fatou's lemma*, named in honor of Pierre Fatou. Its usefulness stems from the fact that no assumptions are placed on the random variables, except that they be nonnegative.

Suppose that X_n is a nonnegative random variable for $n \in \mathbb{N}_+$. Then

$$\mathbb{E}\left(\liminf_{n \rightarrow \infty} X_n\right) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(X_n) \quad (4.9.11)$$

Our next set of results gives conditions for the interchange of expected value and limits.

Suppose that X_n is a random variable for each $n \in \mathbb{N}_+$. then

$$\mathbb{E}\left(\lim_{n \rightarrow \infty} X_n\right) = \lim_{n \rightarrow \infty} \mathbb{E}(X_n) \quad (4.9.12)$$

in each of the following cases:

1. X_n is nonnegative for each $n \in \mathbb{N}_+$ and X_n is increasing in n .
2. $\mathbb{E}(X_n)$ exists for each $n \in \mathbb{N}_+$, $\mathbb{E}(X_1) > -\infty$, and X_n is increasing in n .
3. $\mathbb{E}(X_n)$ exists for each $n \in \mathbb{N}_+$, $\mathbb{E}(X_1) < \infty$, and X_n is decreasing in n .
4. $\lim_{n \rightarrow \infty} X_n$ exists, and $|X_n| \leq Y$ for $n \in \mathbb{N}$ where Y is a nonnegative random variable with $\mathbb{E}(Y) < \infty$.
5. $\lim_{n \rightarrow \infty} X_n$ exists, and $|X_n| \leq c$ for $n \in \mathbb{N}$ where c is a positive constant.

Statements about the random variables in the theorem above (nonnegative, increasing, existence of limit, etc.) need only hold with probability 1. Part (a) is the *monotone convergence theorem*, one of the most important convergence results and in a sense, essential to the definition of the integral in the first place. Parts (b) and (c) are slight generalizations of the monotone convergence theorem. In parts (a), (b), and (c), note that $\lim_{n \rightarrow \infty} X_n$ exists (with probability 1), although the limit may be ∞ in parts (a) and (b) and $-\infty$ in part (c) (with positive probability). Part (d) is the *dominated convergence theorem*, another of the most important convergence results. It's sometimes also known as *Lebesgue's dominated convergence theorem* in honor of Henri Lebesgue. Part (e) is a corollary of the dominated convergence theorem, and is known as the *bounded convergence theorem*.

Infinite Series

Our next results involve the interchange of expected value and an infinite sum, so these results generalize the basic additivity property of expected value.

Suppose that X_n is a random variable for $n \in \mathbb{N}_+$. Then

$$\mathbb{E}\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} \mathbb{E}(X_n) \quad (4.9.13)$$

in each of the following cases:

1. X_n is nonnegative for each $n \in \mathbb{N}_+$.
2. $\mathbb{E}(\sum_{n=1}^{\infty} |X_n|) < \infty$

Part (a) is a consequence of the monotone convergence theorem, and part (b) is a consequence of the dominated convergence theorem. In (b), note that $\sum_{n=1}^{\infty} |X_n| < \infty$ and hence $\sum_{n=1}^{\infty} X_n$ is absolutely convergent with probability 1. Our next result is the additivity of the expected value over a countably infinite collection of disjoint events.

Suppose that X is a random variable whose expected value exists, and that $\{A_n : n \in \mathbb{N}_+\}$ is a disjoint collection events. Let $A = \bigcup_{n=1}^{\infty} A_n$. Then

$$\mathbb{E}(X; A) = \sum_{n=1}^{\infty} \mathbb{E}(X; A_n) \quad (4.9.14)$$

Of course, the previous theorem applies in particular if X is nonnegative.

Integrals

Suppose that (T, \mathcal{T}, μ) is a σ -finite measure space, and that X_t is a real-valued random variable for each $t \in T$. Thus we can think of $\{X_t : t \in T\}$ is a stochastic process indexed by T . We assume that $(\omega, t) \mapsto X_t(\omega)$ is measurable, as a function from the product space $(\Omega \times T, \mathcal{F} \otimes \mathcal{T})$ into \mathbb{R} . Our next result involves the interchange of expected value and integral, and is a consequence of Fubini's theorem, named for Guido Fubini.

Under the assumptions above,

$$\mathbb{E} \left[\int_T X_t d\mu(t) \right] = \int_T \mathbb{E}(X_t) d\mu(t) \quad (4.9.15)$$

in each of the following cases:

1. X_t is nonnegative for each $t \in T$.
2. $\int_T \mathbb{E}(|X_t|) d\mu(t) < \infty$

Fubini's theorem actually states that the two iterated integrals above equal the joint integral

$$\int_{\Omega \times T} X_t(\omega) d(\mathbb{P} \otimes \mu)(\omega, t) \quad (4.9.16)$$

where of course, $\mathbb{P} \otimes \mu$ is the product measure on $(\Omega \times T, \mathcal{F} \otimes \mathcal{T})$. However, our interest is usually in evaluating the iterated integral above on the left in terms of the iterated integral on the right. Part (a) is the expected value version of *Tonelli's theorem*, named for Leonida Tonelli.

Examples and Exercises

You may have worked some of the computational exercises before, but try to see them in a new light, in terms of the general theory of integration.

The Cauchy Distribution

Recall that the *Cauchy distribution*, named for Augustin Cauchy, is a continuous distribution with probability density function f given by

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R} \quad (4.9.17)$$

The Cauchy distribution is studied in more generality in the chapter on Special Distributions.

Suppose that X has the Cauchy distribution.

1. Show that $\mathbb{E}(X)$ does not exist.
2. Find $\mathbb{E}(X^2)$

Answer

1. $\mathbb{E}(X^+) = \mathbb{E}(X^-) = \infty$
2. ∞

Open the Cauchy Experiment and keep the default parameters. Run the experiment 1000 times and note the behavior of the sample mean.

The Pareto Distribution

Recall that the *Pareto distribution*, named for Vilfredo Pareto, is a continuous distribution with probability density function f given by

$$f(x) = \frac{a}{x^{a+1}}, \quad x \in [1, \infty) \quad (4.9.18)$$

where $a > 0$ is the *shape parameter*. The Pareto distribution is studied in more generality in the chapter on Special Distributions.

Suppose that X has the Pareto distribution with shape parameter a . Find $\mathbb{E}(X)$ in the following cases:

1. $0 < a \leq 1$
2. $a > 1$

Answer

1. ∞
2. $\frac{a}{a-1}$

Open the special distribution simulator and select the Pareto distribution. Vary the shape parameter and note the shape of the probability density function and the location of the mean. For various values of the parameter, run the experiment 1000 times and compare the sample mean with the distribution mean.

Suppose that X has the Pareto distribution with shape parameter a . Find $E(1/X^n)$ for $n \in \mathbb{N}_+$.

Answer

$$\frac{a}{a+n}$$

Special Results for Nonnegative Variables

For a nonnegative variable, the moments can be obtained from integrals of the right-tail distribution function.

If X is a nonnegative random variable then

$$\mathbb{E}(X^n) = \int_0^\infty nx^{n-1}\mathbb{P}(X > x) dx \quad (4.9.19)$$

Proof

By [Fubini's theorem](#) we can interchange an expected value and integral when the integrand is nonnegative. Hence

$$\int_0^\infty nx^{n-1}\mathbb{P}(X > x) dx = \int_0^\infty nx^{n-1}\mathbb{E}[\mathbf{1}(X > x)] dx = \mathbb{E}\left(\int_0^\infty nx^{n-1}\mathbf{1}(X > x) dx\right) = \mathbb{E}\left(\int_0^X nx^{n-1} dx\right) = \mathbb{E}(X^n) \quad (4.9.20)$$

When $n = 1$ we have $\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X > x) dx$. We saw this result before in the section on additional properties of expected value, but now we can understand the proof in terms of Fubini's theorem.

For a random variable taking nonnegative integer values, the moments can be computed from sums involving the right-tail distribution function.

Suppose that X has a discrete distribution, taking values in \mathbb{N} . Then

$$\mathbb{E}(X^n) = \sum_{k=1}^\infty [k^n - (k-1)^n] \mathbb{P}(X \geq k) \quad (4.9.21)$$

Proof

By the [theorem above](#), we can interchange expected value and infinite series when the terms are nonnegative. Hence

$$\begin{aligned} \sum_{k=1}^\infty [k^n - (k-1)^n] \mathbb{P}(X \geq k) &= \sum_{k=1}^\infty [k^n - (k-1)^n] \mathbb{E}[\mathbf{1}(X \geq k)] = \mathbb{E}\left(\sum_{k=1}^\infty [k^n - (k-1)^n] \mathbf{1}(X \geq k)\right) \\ &= \mathbb{E}\left(\sum_{k=1}^X [k^n - (k-1)^n]\right) = \mathbb{E}(X^n) \end{aligned} \quad (4.9.22)$$

When $n = 1$ we have $\mathbb{E}(X) = \sum_{k=1}^\infty \mathbb{P}(X \geq k)$. We saw this result before in the section on additional properties of expected value, but now we can understand the proof in terms of the interchange of sum and expected value.

This page titled [4.9: Expected Value as an Integral](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist \(Random Services\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.