

[Unit 5: Averages, Law of Large Numbers, and Central Limit](#)[Course](#) > [Theorem](#)

5.8 Expectation of a continuous

> [5.1 Reading](#) > random variable

5.8 Expectation of a continuous random variable

Unit 5: Averages

Adapted from Blitzstein-Hwang Chapters 4, 5, and 10.

The definition of expectation for continuous r.v.s is analogous to the definition for discrete r.v.s; we just replace the sum with an integral and the PMF with the PDF.

DEFINITION 5.8.1 (EXPECTATION OF A CONTINUOUS R.V.).

The *expected value* (also called the *expectation* or *mean*) of a continuous r.v. X with PDF f is

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

As in the discrete case, the expectation of a continuous r.v. may or may not exist. When discussing expectations, it would be very tedious to have to add "(if it exists)" after every mention of an expectation not yet shown to exist, so we will often leave this implicit.

The integral is taken over the entire real line, but if the support of X is not the entire real line we can just integrate over the support.

Linearity of expectation holds for continuous r.v.s, just as it did for discrete r.v.s. LOTUS also holds for continuous r.v.s, replacing the sum with an integral and the PMF with the PDF:

THEOREM 5.8.2 (LOTUS, CONTINUOUS).

If X is a continuous r.v. with PDF f and g is a function from \mathbb{R} to \mathbb{R} , then

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

Example 5.8.3 (Mean and variance of a Uniform r.v.).

Let's derive the mean and variance of $U \sim \text{Unif}(a, b)$. The expectation is extremely intuitive: the PDF is constant, so its balancing point should be the midpoint of (a, b) . This is exactly what we find by using the definition of expectation for continuous r.v.s:

$$E(U) = \int_a^b x \cdot \frac{1}{b-a} dx = \frac{1}{b-a} \left(\frac{b^2}{2} - \frac{a^2}{2} \right) = \frac{a+b}{2}.$$

For the variance, we first find $E(U^2)$ using the continuous version of LOTUS:

$$E(U^2) = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{3} \cdot \frac{b^3 - a^3}{b-a}.$$

Then

$$\text{Var}(U) = E(U^2) - (EU)^2 = \frac{1}{3} \cdot \frac{b^3 - a^3}{b-a} - \left(\frac{a+b}{2} \right)^2,$$

which reduces, after factoring $b^3 - a^3 = (b-a)(a^2 + ab + b^2)$ and simplifying, to

$$\text{Var}(U) = \frac{(b-a)^2}{12}.$$

Example 5.8.4 (Mean and variance of a Normal r.v.).

Next let's derive the mean and variance of a Normal r.v., showing that a $\mathcal{N}(\mu, \sigma^2)$ r.v. does indeed have mean μ and variance σ^2 . To start, let's consider the standard Normal. By symmetry, its mean must be 0. We can also see this symmetry by looking at the definition of $E(Z)$:

$$E(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} ze^{-z^2/2} dz,$$

and since $g(z) = ze^{-z^2/2}$ is an odd function, the area under g from $-\infty$ to 0 cancels the area under g from 0 to ∞ . Therefore $E(Z) = 0$. In fact, the same argument shows that $E(Z^n) = 0$ for any odd positive number n . For the variance, we can use LOTUS.

$$\begin{aligned} \text{Var}(Z) &= E(Z^2) - (EZ)^2 = E(Z^2) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-z^2/2} dz \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} z^2 e^{-z^2/2} dz \end{aligned}$$

The last step uses the fact that $z^2 e^{-z^2/2}$ is an even function. Now we use integration by parts with $u = z$ and $dv = ze^{-z^2/2} dz$, so $du = dz$ and $v = -e^{-z^2/2}$.

$$\begin{aligned}\text{Var}(Z) &= \frac{2}{\sqrt{2\pi}} \left(-ze^{-z^2/2} \Big|_0^\infty + \int_0^\infty e^{-z^2/2} dz \right) \\ &= \frac{2}{\sqrt{2\pi}} \left(0 + \frac{\sqrt{2\pi}}{2} \right) \\ &= 1.\end{aligned}$$

The first term of the integration by parts equals 0 because $e^{-z^2/2}$ decays much faster than z grows, and the second term is $\sqrt{2\pi}/2$ because it's half of the total area under $e^{-z^2/2}$, which we've already proved is $\sqrt{2\pi}$. So the standard Normal distribution has mean 0 and variance 1.

For $X \sim \mathcal{N}(\mu, \sigma^2)$, we can write $X = \mu + \sigma Z$ with $Z \sim \mathcal{N}(0, 1)$, and then

$$\begin{aligned}E(X) &= \mu + \sigma \cdot 0 = \mu, \\ \text{Var}(X) &= \sigma^2 \text{Var}(Z) = \sigma^2.\end{aligned}$$

Example 5.8.5 (Mean and variance of an Exponential r.v.).

To obtain the mean and variance of an Exponential r.v., let's start by finding the mean and variance of $X \sim \text{Expo}(1)$:

$$E(X) = \int_0^\infty x e^{-x} dx = 1,$$

and by LOTUS,

$$E(X^2) = \int_0^\infty x^2 e^{-x} dx = 2,$$

where the integrals were done using standard integration by parts calculations. Then

$$\text{Var}(X) = E(X^2) - (EX)^2 = 1.$$

Now let $Y = X/\lambda \sim \text{Expo}(\lambda)$. Then

$$\begin{aligned}E(Y) &= \frac{1}{\lambda} E(X) = \frac{1}{\lambda}, \\ \text{Var}(Y) &= \frac{1}{\lambda^2} \text{Var}(X) = \frac{1}{\lambda^2},\end{aligned}$$

Example 5.8.6 (Blissville and Blotchville).

Fred lives in Blissville, where buses always arrive exactly on time, with the time between successive buses fixed at 10 minutes. Having lost his watch, he arrives at the bus stop at a uniformly random time on a certain day (assume that buses run 24 hours a day, every day, and that the time that Fred arrives is independent of the bus arrival process).

- (a) What is the distribution of how long Fred has to wait for the next bus? What is the average time that Fred has to wait?
- (b) Given that the bus has not yet arrived after 6 minutes, what is the probability that Fred will have to wait at least 3 more minutes?
- (c) Fred moves to Blotchville, a city with inferior urban planning and where buses are much more erratic. Now, when any bus arrives, the time until the next bus arrives is an Exponential random variable with mean 10 minutes. Fred arrives at the bus stop at a random time, not knowing how long ago the previous bus came. What is the distribution of Fred's waiting time for the next bus? What is the average time that Fred has to wait?
- (d) When Fred complains to a friend how much worse transportation is in Blotchville, the friend says: "Stop whining so much! You arrive at a uniform instant between the previous bus arrival and the next bus arrival. The average length of that interval between buses is 10 minutes, but since you are equally likely to arrive at any time in that interval, your average waiting time is only 5 minutes." Fred disagrees, both from experience and from solving Part (c) while waiting for the bus. Explain what is wrong with the friend's reasoning.

Solution

(a) The distribution is Uniform on $(0, 10)$, so the mean is **5** minutes.

(b) Let T be the waiting time. Then

$$P(T \geq 6 + 3 | T > 6) = \frac{P(T \geq 9, T > 6)}{P(T > 6)} = \frac{P(T \geq 9)}{P(T > 6)} = \frac{1/10}{4/10} = \frac{1}{4}.$$

In particular, Fred's waiting time in Blissville is not memoryless; conditional on having waited 6 minutes already, there's only a **1/4** chance that he'll have to wait another 3 minutes, whereas if he had just showed up, there would be a $P(T \geq 3) = 7/10$ chance of having to wait at least 3 minutes.

(c) By the memoryless property, the distribution is Exponential with parameter **1/10** (and mean 10 minutes) regardless of when Fred arrives; how much longer the next bus will take to arrive is independent of how long ago the previous bus arrived. The average time that Fred has to wait is 10 minutes.

(d) Fred's friend is making the mistake, explained in [Warning 5.1.3](#), of replacing a random variable (the time between buses) by its expectation (10 minutes), thereby ignoring the variability in interarrival times. The *average* length of a time interval between two buses is 10 minutes, but Fred is not equally likely to arrive at any of these intervals: Fred is more likely to arrive during a long interval between buses than to arrive during a short interval between buses. For example, if one interval between buses is 50 minutes and another interval is 5 minutes, then Fred is 10 times more likely to arrive during the 50-minute interval. This phenomenon is known as *length-biasing*, and it comes up in many real-life situations. For example, asking randomly chosen mothers how many children they have yields a different distribution from asking randomly chosen people how many siblings they have, including themselves. Asking students the sizes of their classes and averaging those results may give a much higher value than taking a list of classes and averaging the sizes of each (this is called the *class size paradox*).

You can explore the bus paradox further in the interactives in this unit.

Learn About Verified Certificates

