Moments

25M

The ***expected value***, or ***mean***, of a random variable is its average value. In general, the expected value of a random variable X can be calculated using one of the two formulas below:

$$\begin{aligned}\mathrm{E}[X] &= \sum_{\text{all } x} x \cdot p(x) && \text{(discrete)} \\ &= \int_{-\infty}^{\infty} x \cdot f(x)\,\mathrm{d}x && \text{(continuous)}\end{aligned}$$

The formula for $\mathrm{E}[X]$ can be generalized by substituting the random variable X for some function $g(X)$. Then, X would be a special case of $g(X)$. The expected value formulas now become:

$$\begin{aligned}\mathrm{E}[g(X)] &= \sum_{\text{all } x} g(x) \cdot p(x) && \text{(discrete)} \\ &= \int_{-\infty}^{\infty} g(x) \cdot f(x)\,\mathrm{d}x && \text{(continuous)} \qquad \text{(S2.1.4.1)}\end{aligned}$$

There is an alternative method to calculate the expected values. This method only works if the random variable X is **non-negative**. For a function $g(X)$ where $g(0) = 0$, we have

$$\mathrm{E}[g(X)] = \int_{0}^{\infty} g'(x) \cdot S(x)\,\mathrm{d}x \qquad \text{(S2.1.4.2)}$$

Note that the lower limit of the integral should **always** be 0, regardless of the domain of X. The derivation of this method is provided in the appendix at the end of this section.

Coach's Remarks

$(\mathrm{S2.1.4.2})$ above should only be used for **continuous** variables. Under certain conditions, the same formula also works for discrete variables. However, to keep it simple, **avoid** this method for discrete variables.

Expected values have three simple properties that are worth remembering. For a constant c,

1. $\mathrm{E}[c] = c$
2. $\mathrm{E}[c \cdot g(X)] = c \cdot \mathrm{E}[g(X)]$
3. $\mathrm{E}[g_1(X) + g_2(X) + \ldots + g_k(X)] = \mathrm{E}[g_1(X)] + \mathrm{E}[g_2(X)] + \ldots + \mathrm{E}[g_k($

Certain expected values are also known as ***moments***. Generally, moments can be categorized into two types: raw moments and central moments.

Raw Moments

The k^{th} ***raw moment*** of X, or the k^{th} ***moment*** of X for short, is defined as

$$\mu'_k = \mathrm{E}\left[X^k\right]$$

Note that the 1^{st} raw moment of X is the mean. For the sake of simplicity, the mean is usually denoted as μ.

$$\mu'_1 = \mathrm{E}[X] = \mu$$

Central Moments

The k^{th} ***central moment*** of X is defined as

$$\mu_k = \mathrm{E}\left[(X-\mu)^k\right]$$

where μ is the mean of X.

Coach's Remarks

It is important to distinguish between the notations for raw moments, central moments, and the mean.

- μ'_k (with prime symbol and subscript k): k^{th} raw moment
- μ_k (with subscript k): k^{th} central moment
- μ: mean, or alternative notation for the 1st raw moment

Time for a pop quiz! What is μ_1, the 1st central moment? As it turns out, it **always** equals 0.

$$\begin{aligned}\mu_1 &= \mathrm{E}\left[(X-\mu)^1\right] \\ &= \mathrm{E}[X]-\mu \\ &= \mu-\mu \\ &= 0\end{aligned}$$

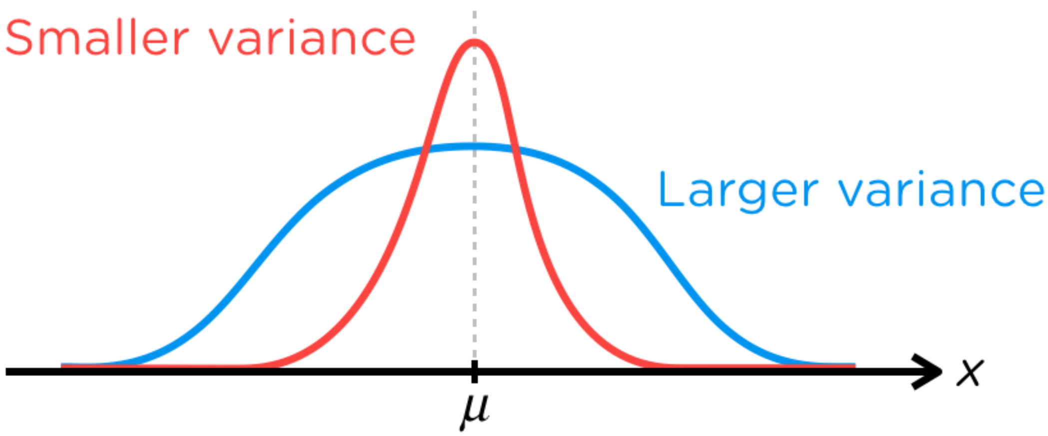
Variance

The most well-known central moment is the 2nd central moment, otherwise known as the ***variance***. Notation-wise, the variance of X is $\mathrm{Var}[X]$, or σ^2.

$$\mu_2 = \mathrm{E}\left[(X-\mu)^2\right] = \mathrm{Var}[X] = \sigma^2$$

Variance measures how much the observations deviate from their mean. Mathematically, variance is the expected squared difference between the random variable and its mean.

- A **larger** variance means the observations are more **dispersed**, or more spread out from the mean.
- A **smaller** variance means the observations are **closer** to the mean.



Let's expand the expression for the 2nd central moment of X.

$$\begin{aligned}\mathrm{Var}[X] &= \mathrm{E}\left[(X-\mu)^2\right] \\ &= \mathrm{E}\left[X^2 - 2X\mu + \mu^2\right] \\ &= \mathrm{E}\left[X^2\right] - 2\mathrm{E}[X]\mu + \mu^2 \\ &= \mathrm{E}\left[X^2\right] - 2\mu^2 + \mu^2 \\ &= \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2\end{aligned}$$

This gives us an alternative formula to calculate the variance. Since the 1st and 2nd raw moments are often easier to calculate, the formula above is our default variance formula. Further generalizing the formula using a function $g(X)$, we have

$$\mathrm{Var}[g(X)] = \mathrm{E}\left[g(X)^2\right] - \mathrm{E}[g(X)]^2 \qquad \textbf{(S2.1.4.3)}$$

Variance has four important properties that are useful for this exam. For constants a, b, and c,

1. $\mathrm{Var}[c] = 0$
2. $\mathrm{Var}[X + c] = \mathrm{Var}[X]$
3. $\mathrm{Var}[cX] = c^2\mathrm{Var}[X]$
4. $\mathrm{Var}[aX + bY] = a^2\mathrm{Var}[X] + b^2\mathrm{Var}[Y] + 2ab\mathrm{Cov}[X,\ Y]$, where $\mathrm{Cov}[X,\ Y] = \mathrm{E}[XY] - \mathrm{E}[X] \cdot \mathrm{E}[Y]$

The square root of the variance is the ***standard deviation***, which is often denoted by σ. In the case of multiple random variables, subscripts will be added to σ to distinguish between the random variables.

$$\sigma = \sqrt{\mathrm{Var}[X]} \qquad \text{(S2.1.4.4)}$$

The ***coefficient of variation*** is the ratio of the standard deviation to the mean. In other words, it calculates the standard deviation per unit of mean.

$$CV = \frac{\sigma}{\mu} \qquad \text{(S2.1.4.5)}$$

Skewness

The 3rd central moment can be used to calculate the ***skewness*** of a distribution, which equals the 3rd central moment divided by the standard deviation cubed.

$$\mathrm{Skewness} = \frac{\mu_3}{\sigma^3} \qquad \text{(S2.1.4.6)}$$

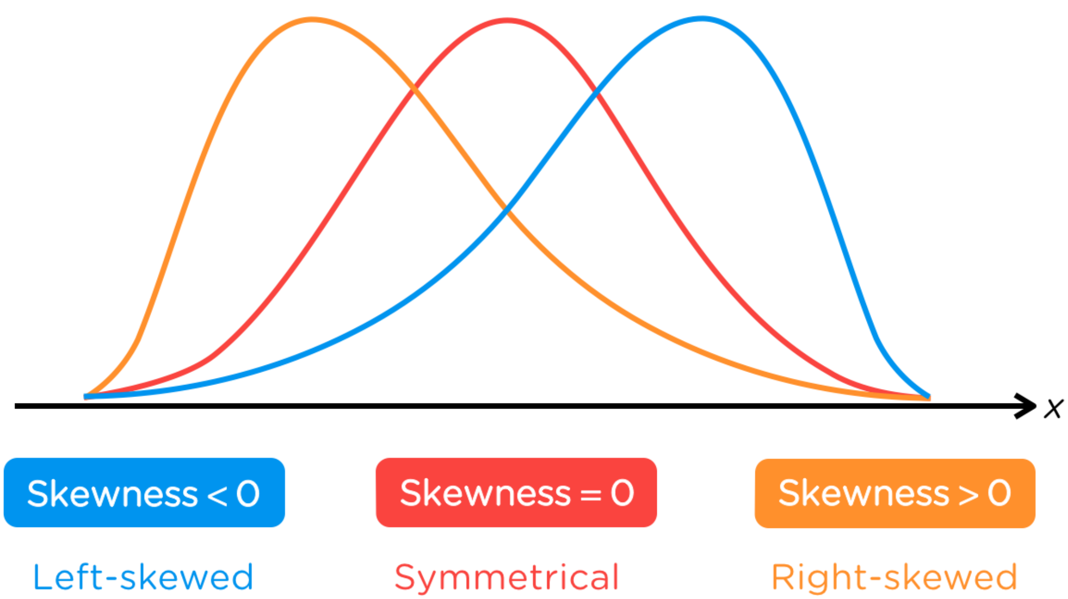
It is usually more practical to compute the numerator, μ_3, by expanding the expression for the 3rd central moment, similar to how we derived our default variance formula above.

$$\mu_3 = \mu_3' - 3\mu_2'\mu + 2\mu^3$$

The complete expansion is provided in the appendix at the end of this section. You should not memorize this equation. Instead, learn how to express a central moment in terms of raw moments.

Skewness measures how symmetric a distribution is relative to its mean.

- **Zero skewness** means the distribution is **perfectly symmetrical**. One popular example is the normal distribution.
- **Positive skewness** means the distribution is **right-skewed**, suggesting a longer right-tail. This implies smaller values are more likely to occur than larger values.
- **Negative skewness** means the distribution is **left-skewed**, suggesting a longer left-tail. This implies larger values are more likely to occur than smaller values.



Coach's Remarks

It is easy to confuse positive and negative skewness. Here is one way to help you remember.

- If a distribution has positive skewness, the endpoint of its longer tail will point towards the positive direction of the x-axis, i.e. point to the **right**. Therefore, "positively-skewed" also means "right-skewed".
- If a distribution has negative skewness, the endpoint of its longer tail will point towards the negative direction of the x-axis, i.e. point to the **left**. Therefore, "negatively-skewed" also means "left-skewed".

Kurtosis

The ***kurtosis*** of a distribution is the 4th central moment divided by the standard deviation to the 4th power.

$$\text{Kurtosis} = \frac{\mu_4}{\sigma^4} \tag{S2.1.4.7}$$

The numerator, μ_4, can be expressed using raw moments:

$$\mu_4 = \mu_4' - 4\mu_3'\mu + 6\mu_2'\mu^2 - 3\mu^4$$

The complete expansion is provided in the appendix at the end of this section.

Kurtosis measures the flatness of a distribution. It suggests the likelihood for a distribution to produce extreme values, or outliers. Here are a few kurtosis facts:

- The normal distribution has a kurtosis of 3.
- A distribution with kurtosis greater than 3 is more likely to produce outliers than the normal distribution, and vice versa.

Coach's Remarks

While it might not be directly tested, it is good to know that the coefficient of variation, skewness, and kurtosis are all **scale invariant**. In other words, if a random variable is multiplied by a positive factor, these three quantities will remain unchanged.

Example S2.1.4.1

A random variable X has the following density function:

$$f(x) = \frac{x}{8}, \qquad 0 \leq x \leq 4$$

Calculate

1. the coefficient of variation of X.
2. the skewness of X.
3. the kurtosis of X.

Solution to (1)

Calculate the coefficient of variation as follows:

$$\begin{aligned}
\mathrm{E}[X] &= \int_0^4 x \cdot \frac{x}{8} \, \mathrm{d}x \\
&= \left[\frac{x^3}{3(8)} \right]_0^4 \\
&= \frac{8}{3}
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}\left[X^2\right] &= \int_0^4 x^2 \cdot \frac{x}{8} \, \mathrm{d}x \\
&= \left[\frac{x^4}{4(8)} \right]_0^4 \\
&= 8
\end{aligned}$$

$$\mathrm{Var}[X] = 8 - \left(\frac{8}{3}\right)^2 = \frac{8}{9}$$

$$\begin{aligned} CV &= \frac{\sqrt{\mathrm{Var}[X]}}{\mathrm{E}[X]} \\ &= \frac{\sqrt{8/9}}{8/3} \\ &= \mathbf{0.3536} \end{aligned}$$

Solution to (2)

All quantities needed to calculate the skewness have been calculated above, except the 3rd central moment.

$$\begin{aligned} \mathrm{E}[X^3] &= \int_0^4 x^3 \cdot \frac{x}{8}\,\mathrm{d}x \\ &= \left[\frac{x^5}{5(8)}\right]_0^4 \\ &= 25.6 \end{aligned}$$

$$\begin{aligned} \mu_3 &= \mu_3' - 3\mu_2'\mu + 2\mu^3 \\ &= 25.6 - 3(8)\left(\frac{8}{3}\right) + 2\left(\frac{8}{3}\right)^3 \\ &= -0.4741 \end{aligned}$$

$$\begin{aligned}\text{Skewness} &= \frac{-0.4741}{\left(\sqrt{8/9}\right)^3} \\ &= \mathbf{-0.5657}\end{aligned}$$

Solution to (3)

To determine the kurtosis, calculate the 4th moment. All other required quantities have been calculated above.

$$\begin{aligned}\mathrm{E}\left[X^4\right] &= \int_0^4 x^4 \cdot \frac{x}{8}\,\mathrm{d}x \\ &= \left[\frac{x^6}{6(8)}\right]_0^4 \\ &= 85.333\end{aligned}$$

$$\begin{aligned}\mu_4 &= \mu_4' - 4\mu_3'\mu + 6\mu_2'\mu^2 - 3\mu^4 \\ &= 85.333 - 4(25.6)\left(\frac{8}{3}\right) + 6(8)\left(\frac{8}{3}\right)^2 - 3\left(\frac{8}{3}\right)^4 \\ &= 1.8963\end{aligned}$$

$$\begin{aligned}\text{Kurtosis} &= \frac{1.8963}{\left(\sqrt{8/9}\right)^4} \\ &= \mathbf{2.4}\end{aligned}$$

Example S2.1.4.2

The annual claim size of an insured has the following probability distribution:

Claim size	Probability
100	0.3
200	0.2
250	0.5

Calculate the skewness of annual claim size.

Solution

Let X represent the annual claim size.

$$\text{Skewness} = \frac{\mu_3}{\sigma^3}$$

$$\begin{aligned}\mu &= \mathrm{E}[X] \\ &= 0.3(100) + 0.2(200) + 0.5(250) \\ &= 195\end{aligned}$$

$$\begin{aligned}\sigma^2 &= \mathrm{E}\left[(X-195)^2\right] \\ &= 0.3(100-195)^2 + 0.2(200-195)^2 + 0.5(250-195)^2 \\ &= 4{,}225\end{aligned}$$

$$\begin{aligned}\mu_3 &= \mathrm{E}\left[(X-195)^3\right] \\ &= 0.3(100-195)^3 + 0.2(200-195)^3 + 0.5(250-195)^3 \\ &= -174{,}000\end{aligned}$$

$$\begin{aligned}\text{Skewness} &= \frac{\mu_3}{\sigma^3} \\ &= \frac{-174{,}000}{\left(\sqrt{4{,}225}\right)^3} \\ &= \mathbf{-0.6336}\end{aligned}$$

Coach's Remarks

Notice that in this example, it is easier to calculate the 3rd central moment directly rather than in terms of raw moments. This is because the X is a discrete random variable with only 3 possible values.

In addition, the final answer of -0.6336 makes sense because a negative skewness implies a left-skewed distribution, which suggests that larger values of the random variable are more likely to occur. This agrees with