Chapter 6A: Moments, Skewness, and Kurtosis

MATH/STAT 394: Probability I

Arman Jahangiri

University of Washington Department of Mathematics

Summer 2026

Learning Goals

By the end of this lecture, students should be able to:

- define and interpret the mean, median, and mode;
- define moments, standardized moments, and standardized moments;
- define skewness and kurtosis;
- understand symmetry of distributions;
- define sample moments;
- compute the mean and variance of the sample mean.

Summaries of a Distribution

Why Summaries Matter

A probability distribution contains a large amount of information.

Often we want simpler numerical summaries:

- Where is the center?
- How spread out is the distribution?
- Is it symmetric?
- Are extreme values common?

Main idea

Moments provide numerical summaries of distributions.

Measures of Standardized Tendency

Three common summaries of the center are:

- Mean (expected value)
- Median
- Mode

These summarize different aspects of a distribution.

Important

Different notions of “center” can give very different values.

Median

Definition

A number m is a median of X if

$$\mathbb{P}(X \leq m) \geq \frac{1}{2}$$

and

$$\mathbb{P}(X \geq m) \geq \frac{1}{2}.$$

Intuitively:

- at least half the distribution lies to the left;
- at least half lies to the right.

Medians are less sensitive to extreme values than means.

Mode

Definition

A value c is called a *mode* of a random variable X if it maximizes:

$$\begin{cases} p(c), & \text{for a discrete random variable with PMF } p(x), \\ f(c), & \text{for a continuous random variable with PDF } f(x). \end{cases}$$

Equivalently,

$$\begin{cases} p(c) \geq p(x), & \forall x, \\ f(c) \geq f(x), & \forall x. \end{cases}$$

Interpretation

The mode is the most likely value (discrete case) or the point of highest density (continuous case).

Mean vs Median

Suppose salaries are:

$$40,\ 42,\ 45,\ 46,\ 48,\ 50,\ 1000.$$

Then:

$$\text{median} = 46,$$

while

$$\text{mean} \approx 181.6.$$

Observation

The mean is heavily affected by extreme values (outliers), while the median is robust.

Best Prediction Depends on Loss Function

Suppose we predict an unknown random variable X by a constant c.

Two common error measures:

$$\mathbb{E}(X - c)^2$$

(mean squared error)

and

$$\mathbb{E}|X - c|$$

(mean absolute error).

Key fact

The mean minimizes mean squared error.
A median minimizes mean absolute error.

Mean and Variance Are Not Enough

Two distributions can have:

same mean, same variance,

yet look very different.

Possible differences:

- asymmetry;
- heavy tails;
- sharp peaks;
- multiple modes.

Conclusion

Higher moments help describe these additional features.

Moments and Symmetry

Kinds of Population Moments

Definitions

The *n*th population moment:

$$\mu_n' = \mathbb{E}(X^n).$$

The *n*th central population moment:

$$\mu_n = E\left[(X-\mu)^n\right]$$

The *n*th standardized population moment:

$$\tilde{\mu}_n = \frac{\mu_n}{\sigma^n} = \mathbb{E}\left(\left(\frac{X-\mu}{\sigma}\right)^n\right),$$

So, we have the following relations

- **Mean** is the 1st population moment $\mu_1' = \mu$, determines center
- **Variance** is the 2nd standardized population moment) $\mu_2 = \sigma^2$, determines spread
- **Skewness** is the 3rd standardized population moment $\frac{\mu_3}{\sigma^3}$, determines asymmetry
- **Kurtosis** is the 4th standardized population moment $\frac{\mu_4}{\sigma^4}$, determines tail behavior

Symmetry About a Point

Definition

A random variable X is symmetric about μ if

$$X - \mu \stackrel{d}{=} \mu - X = -(X - \mu),$$

where $\stackrel{d}{=}$ denotes equality in distribution.

Let

$$Y = X - \mu.$$

Then the CDF of Y is

$$F_Y(y) = P(Y \leq y) = P(X - \mu \leq y) = P(X \leq y + \mu) = F_X(y + \mu).$$

If X has PDF f_X, then differentiating gives

$$f_Y(y) = f_X(y + \mu).$$

Similarly, for $-Y = \mu - X$,

$$f_{-Y}(y) = f_Y(-y) = f_X(\mu - y).$$

Symmetry About a Point (Continue)

Therefore, X is symmetric about μ when

$$f_X(\mu + y) = f_X(\mu - y) \qquad \text{for all } y.$$

- Values equally far to the left and right of μ have the same density.
- The graph of f_X is a mirror image across the vertical line $x = \mu$.
- If the mean exists, then the center of symmetry is the mean:

$$E[X] = \mu.$$

- More generally, all odd central moments that exist are zero:

$$E\left[(X - \mu)^{2k+1}\right] = 0.$$

Symmetry and Odd Central Moments

Suppose X is symmetric about μ. Then

$$f_X(\mu + t) = f_X(\mu - t).$$

Consider an odd central moment, say the $(2k+1)$st:

$$\mu_{2k+1} = \mathbb{E}\left[(X-\mu)^{2k+1}\right] = \int_{-\infty}^{\infty} (x-\mu)^{2k+1} f_X(x)\, dx.$$

Make the change of variable

$$t = x - \mu.$$

Then $x = \mu + t$, so

$$\mu_{2k+1} = \int_{-\infty}^{\infty} t^{2k+1} f_X(\mu + t)\ dt.$$

Now define

$$g(t) = t^{2k+1} f_X(\mu + t).$$

Then

$$\begin{aligned} g(-t) &= \underbrace{(-t)^{2k+1}}_{=-t^{2k+1}} \underbrace{f_X(\mu - t)}_{=f_X(\mu+t)} \\ &= -g(t). \end{aligned}$$

Thus g is an **odd function**. Therefore,

$$\underbrace{\int_{-\infty}^{\infty} g(t)\, dt}_{\substack{\text{integral of an odd function}\\ \text{over a symmetric interval}}} = 0.$$

Hence

A second way of proving it...

Proposition

Let X be a random variable symmetric about μ. Then, for every integer $k \geq 0$,

$$\mathbb{E}\left[(X-\mu)^{2k+1}\right] = 0,$$

provided the moment exists.

Proof.

Let $Y = X - \mu$. By symmetry, $Y \stackrel{d}{=} -Y$.. Therefore, $\mathbb{E}\left[Y^{2k+1}\right] = \mathbb{E}\left[(-Y)^{2k+1}\right]$. But since $2k+1$ is odd, $(-Y)^{2k+1} = -Y^{2k+1}$. Hence, $\mathbb{E}\left[Y^{2k+1}\right] = -\mathbb{E}\left[Y^{2k+1}\right]$. So, $2\mathbb{E}\left[Y^{2k+1}\right] = 0$, and therefore $\mathbb{E}\left[Y^{2k+1}\right] = 0$. Since $Y = X - \mu$, $\mathbb{E}\left[(X-\mu)^{2k+1}\right] = 0$. $\square$

Examples of Symmetry

A distribution is symmetric about μ when points equally far from μ have the same probability density:

$$f_X(\mu + t) = f_X(\mu - t).$$

- **Normal distribution:** If $X \sim N(\mu, \sigma^2)$, then

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

  For any t,

$$f_X(\mu + t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-t^2/(2\sigma^2)} = f_X(\mu - t).$$

  Thus, the normal distribution is symmetric about μ.

- **Uniform distribution on a symmetric interval:** If $X \sim \mathrm{Unif}(\mu - a, \mu + a)$, then

$$f_X(x) = \begin{cases} \dfrac{1}{2a}, & \mu - a \le x \le \mu + a, \\ 0, & \text{otherwise.} \end{cases}$$

  Hence, for every t, $f_X(\mu + t) = f_X(\mu - t)$, since either both points lie inside the interval and have density $\frac{1}{2a}$, or both lie outside and have density 0.

- **Binomial distribution with $p = \frac{1}{2}$:** If $X \sim \mathrm{Bin}(n, \frac{1}{2})$, then

$$P(X = k) = \binom{n}{k}\left(\frac{1}{2}\right)^n = \binom{n}{n-k}\left(\frac{1}{2}\right)^n = P(X = n - k).$$

  Thus, the distribution is symmetric about $\frac{n}{2}$.

Skewness

Definition

The skewness of X is

$$\boxed{\operatorname{Skew}(X) = \mathbb{E}\left(\left(\frac{X-\mu}{\sigma}\right)^3\right).}$$

Interpretation:

- positive skewness: long right tail;
- negative skewness: long left tail;
- zero skewness: often indicates symmetry.

Examples of Skewness

- Exponential distributions are right-skewed.
- Income distributions are often right-skewed.
- Normal distributions have skewness 0.

Important

Zero skewness does not necessarily imply symmetry.

Definition

The kurtosis of X is

$$\boxed{\text{Kurt}(X) = \mathbb{E}\left(\left(\frac{X-\mu}{\sigma}\right)^4\right).}$$

- Kurtosis measures tail heaviness and extremeness.
- Kurt($N(\mu, \sigma^2)$)=3. If $Kurt(X) >< 3$ then f_X has heavier/lighter tails compared to normal.

Common Discrete Distributions

Distribution	Mean	Median	Mode	Variance	Skewness	Kurtosis
Bernoulli(p)	p	$\begin{cases} 0, & p < 1/2 \\ 1/2, & p = 1/2 \\ 1, & p > 1/2 \end{cases}$	$\begin{cases} 0, & p < 1/2 \\ 0, 1, & p = 1/2 \\ 1, & p > 1/2 \end{cases}$	$p(1-p)$	$\dfrac{1-2p}{\sqrt{p(1-p)}}$	$\dfrac{1-6p(1-p)}{p(1-p)}+3$
Binomial(n, p)	np	$\approx np$	$\lfloor (n+1)p \rfloor$	$np(1-p)$	$\dfrac{1-2p}{\sqrt{np(1-p)}}$	$\dfrac{1-6p(1-p)}{np(1-p)}+3$
Discrete Uniform$(1, \ldots, n)$	$\dfrac{n+1}{2}$	$\dfrac{n+1}{2}$	none	$\dfrac{n^2-1}{12}$	0	$\dfrac{9}{5} \cdot \dfrac{n^2+1}{n^2-1}$
Geometric(p)	$\dfrac{1}{p}$	$\left\lceil \dfrac{-1}{\log_2(1-p)} \right\rceil$	1	$\dfrac{1-p}{p^2}$	$\dfrac{2-p}{\sqrt{1-p}}$	$6 + \dfrac{p^2}{1-p} + 3$
Poisson(λ)	λ	$\approx \lfloor \lambda + \frac{1}{3} - \frac{0.02}{\lambda} \rfloor$	$\lfloor \lambda \rfloor - 1, \lfloor \lambda \rfloor$	λ	$\dfrac{1}{\sqrt{\lambda}}$	$\dfrac{1}{\lambda} + 3$

Click here to see a bigger table, which you can bring a copy on the exams too. It is on the instructor's course website.

Common Continuous Distributions

Distribution	Mean	Median	Mode	Variance	Skewness	Kurtosis
Uniform(a, b)	$\frac{a+b}{2}$	$\frac{a+b}{2}$	none	$\frac{(b-a)^2}{12}$	0	$\frac{9}{5}$
Normal(μ, σ^2)	μ	μ	μ	σ^2	0	3
Exponential(λ)	$\frac{1}{\lambda}$	$\frac{\log 2}{\lambda}$	0	$\frac{1}{\lambda^2}$	2	9
Gamma(α, λ)	$\frac{\alpha}{\lambda}$	(no closed form)	$\frac{\alpha-1}{\lambda} \quad (\alpha > 1)$	$\frac{\alpha}{\lambda^2}$	$\frac{2}{\sqrt{\alpha}}$	$\frac{6}{\alpha} + 3$
Beta(α, β)	$\frac{\alpha}{\alpha+\beta}$	(no closed form)	$\frac{\alpha-1}{\alpha+\beta-2}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$\frac{2(\beta-\alpha)\sqrt{\alpha+\beta+1}}{(\alpha+\beta+2)\sqrt{\alpha\beta}}$	*

* Kurt(Beta(α, β)) $= \frac{6\left[(\alpha-\beta)^2(\alpha+\beta+1)-\alpha\beta(\alpha+\beta+2)\right]}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)} + 3$

Click here to see a bigger table, which you can bring a copy on the exams too. It is on the instructor's course website.

Sample Moments

Sample Moments

Suppose

$$X_1, \ldots, X_n$$

are i.i.d. random variables.

Definition

The kth sample moment is

$$M_k = \frac{1}{n} \sum_{j=1}^{n} X_j^k.$$

- Sample moments estimate population moments.
- The first sample moment, $M_1 = \frac{1}{n} \sum_{j=1}^{n} X_j$, is an important quantity in Statistics, also known as the *Sample Mean*, denoted by $\bar{X}$. It estimates the first population moment, i.e., $\mu_1' = \mathbb{E}[X]$, the population mean.

i.i.d. Random Variables (Random Samples)

Definition: i.i.d. Random Variables

We say that the random variables $X_1, \ldots, X_n$ are *independent and identically distributed* (i.i.d.) with distribution F, and write

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} F,$$

if:

- $X_1, \ldots, X_n$ are independent;
- each X_j has distribution F.

Terminology: If $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} F$, then we also call a *random sample* of size n from the population distribution F. Quantities associated with F are called *population quantities*, while quantities computed from the random sample are called *sample quantities*.

Example

If

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2),$$

then for every $j = 1, \ldots, n$,

$$\mathbb{E}(X_j) = \mu, \qquad \mathrm{Var}(X_j) = \sigma^2.$$

Example of a Random Sample

Suppose the population consists of the GPAs of all University of Washington undergraduate students.
Let F denote the population distribution of undergraduate GPAs.
Now suppose we randomly select 12 undergraduate students and record their GPAs:

$$X_1, \dots, X_{12} \overset{\text{i.i.d.}}{\sim} F.$$

Here:

- each X_j represents one student's GPA;
- all X_j's have distribution F;
- the observations are assumed to be independent.

A possible sample is:

$$3.5,\ 3.7,\ 4.0,\ 2.9,\ 3.1,\ 3.8,\ 3.4,\ 2.7,\ 3.9,\ 3.2,\ 3.6,\ 3.3.$$

Important

In practice, we usually do not know the population distribution F or its parameters, such as the population mean μ and variance σ^2.
Instead, we observe a random sample

$$X_1, \dots, X_n$$

from the population and use the random sample to estimate unknown population quantities.

Sample Mean and Sample Variance

Let

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} F, \qquad \mathbb{E}(X_j) = \mu, \qquad \mathrm{Var}(X_j) = \sigma^2.$$

Definitions

The *sample mean* and *sample variance* are

$$\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j, \qquad S_n^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X}_n)^2.$$

Mean and Variance

$$\mathbb{E}(\bar{X}_n) = \mu, \qquad \mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}.$$

$$\mathbb{E}(S_n^2) = \sigma^2, \qquad \mathrm{Var}(S_n^2) = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right),$$

where

$$\mu_4 = \mathbb{E}\big[(X_j - \mu)^4\big].$$

Summary

- Mean, median, and mode describe center differently.
- Higher moments describe asymmetry and tail behavior.
- Skewness measures asymmetry.
- Kurtosis measures tail heaviness.
- Sample moments estimate population moments.
- The variance of the sample mean decreases like

$$\frac{1}{n}.$$

Next lecture

Moment generating functions.