

Chapter 6A: Moments, Skewness, and Kurtosis

MATH/STAT 394: Probability I

Arman Jahangiri

University of Washington
Department of Mathematics

Summer 2026

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

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.

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.

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.

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).

Suppose salaries are:

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

Then:

median = 46,

while

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

Definitions

The n th population moment:

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

The n th central population moment:

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

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

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

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

- ▶ In this example, values far to the left (from μ the center) act "better" compared to the values far to the right.
- ▶ 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:

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

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:

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

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,

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

Hence

Proposition

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

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

provided the moment exists.

Proof.

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

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 $X \sim \text{Unif}(\mu - a, \mu + a)$, then

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

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 $X \sim \text{Bin}(n, \frac{1}{2})$, 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 $\frac{n}{2}$.

Definition

The skewness of X is

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

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

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

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

Common Discrete Distributions

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

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)	$\frac{\alpha-1}{\alpha+\beta-2}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$\frac{2(\beta-\alpha)\sqrt{\alpha+\beta+1}}{(\alpha+\beta+2)\sqrt{\alpha\beta}}$	*

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

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

Suppose

$$X_1, \dots, X_n$$

are i.i.d. random variables.

Definition

The k th 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.

Definition

A random variable X has a *point-mass distribution at c* if

$$\mathbb{P}(X = c) = 1.$$

Equivalently, $X = c$ almost surely. We may write

$$X \sim \delta_c,$$

where δ_c denotes the point mass at c .

Its PMF is

$$p_X(x) = \mathbb{P}(X = x) = \begin{cases} 1, & x = c, \\ 0, & x \neq c. \end{cases}$$

Its CDF is

$$F_X(x) = \mathbb{P}(X \leq x) = \begin{cases} 0, & x < c, \\ 1, & x \geq c. \end{cases}$$

Thus, the CDF has a jump of size 1 at $x = c$:

$$F_X(c) - F_X(c^-) = \mathbb{P}(X = c) = 1.$$

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

Definition: i.i.d. Random Variables

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

$$X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} F,$$

if:

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

Terminology: If $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} F$, 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

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

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

$$\forall i \in \{1, \dots, n\}, X_i \sim N(\mu, \sigma^2), \text{ (identical distribution)} \quad \text{and} \quad \forall i, j \in \{1, \dots, n\}, X_i \perp X_j \text{ (independence)}$$

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:

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

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 (after observing $X_1, \dots, X_n$) 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.

Remark 1: X_i 's are random variables (non deterministic) in theory. Once we perform the random experiment, they become known to us. We would then call those numbers an observation of our random variables. If we perform the experiment again, we will see another observation (possibly different numbers). For example, let the population be *Seattle's residents* and define $X = \text{income}$. Then, drawing a random sample $X_1, \dots, X_n$ results in a few numbers (observations) such as 30,000, 40,000, $\dots$, 150,000, 24,000. These are the values that the random variables $X_1, \dots, X_n$ have taken in our random experiment. So we would write $x_1 = 30,000, x_2 = 40,000, \dots, x_{n-1} = 150,000, x_n = 24,000$.

Remark 2: 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

$$X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} F, \quad \mathbb{E}(X_j) = \mu, \quad \text{Var}(X_j) = \sigma^2.$$

Definitions

The *sample mean* and *sample variance* are

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

Mean and Variance of the Sample Mean and Sample Variance

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

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

where

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

Expected Value

By linearity of expectation,

$$\mathbb{E}(\bar{X}_n) = \mathbb{E}\left(\frac{1}{n} \sum_{j=1}^n X_j\right) = \frac{1}{n} \sum_{j=1}^n \mathbb{E}(X_j) = \frac{1}{n} \sum_{j=1}^n \mu = \mu.$$

Thus, $\bar{X}_n$ is an unbiased estimator of μ .

Variance

Since $X_1, \dots, X_n$ are independent,

$$\text{Var}(\bar{X}_n) = \text{Var}\left(\frac{1}{n} \sum_{j=1}^n X_j\right) = \frac{1}{n^2} \sum_{j=1}^n \text{Var}(X_j) = \frac{1}{n^2} \sum_{j=1}^n \sigma^2 = \frac{\sigma^2}{n}.$$

A Useful Identity (for the Sample Variance, before the proofs)

To study S_n^2 , first rewrite its numerator.

Sum-of-Squares Identity

$$\sum_{j=1}^n (X_j - \bar{X}_n)^2 = \sum_{j=1}^n (X_j - \mu)^2 - n(\bar{X}_n - \mu)^2$$

Proof. Write

$$X_j - \bar{X}_n = (X_j - \mu) - (\bar{X}_n - \mu).$$

Therefore,

$$\begin{aligned} \sum_{j=1}^n (X_j - \bar{X}_n)^2 &= \sum_{j=1}^n [(X_j - \mu) - (\bar{X}_n - \mu)]^2 \\ &= \sum_{j=1}^n (X_j - \mu)^2 - 2(\bar{X}_n - \mu) \sum_{j=1}^n (X_j - \mu) \\ &\quad + n(\bar{X}_n - \mu)^2. \end{aligned}$$

But $\sum_{j=1}^n (X_j - \mu) = n(\bar{X}_n - \mu)$. Hence,

$$\sum_{j=1}^n (X_j - \bar{X}_n)^2 = \sum_{j=1}^n (X_j - \mu)^2 - n(\bar{X}_n - \mu)^2.$$

Proofs (Sample Variance)

Using the identity from the previous slide, $(n-1)S_n^2 = \sum_{j=1}^n (X_j - \mu)^2 - n(\bar{X}_n - \mu)^2$. Take expectations:

$$(n-1)\mathbb{E}(S_n^2) = \sum_{j=1}^n \mathbb{E}[(X_j - \mu)^2] - n\mathbb{E}[(\bar{X}_n - \mu)^2].$$

Since $\mathbb{E}[(X_j - \mu)^2] = \sigma^2$ and $\mathbb{E}[(\bar{X}_n - \mu)^2] = \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$, we obtain

$$(n-1)\mathbb{E}(S_n^2) = n\sigma^2 - n\left(\frac{\sigma^2}{n}\right) = n\sigma^2 - \sigma^2 = (n-1)\sigma^2.$$

Therefore,

$$\boxed{\mathbb{E}(S_n^2) = \sigma^2.}$$

Why Do We Divide by $n-1$?

The denominator $n-1$, rather than n , corrects for the downward bias caused by replacing the unknown population mean μ with the sample mean $\bar{X}_n$.

Proofs (Sample Variance)

Using the identity from the previous slide,

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

Take expectations:

$$(n-1)\mathbb{E}(S_n^2) = \sum_{j=1}^n \mathbb{E}[(X_j - \mu)^2] - n\mathbb{E}[(\bar{X}_n - \mu)^2].$$

Since

$$\mathbb{E}[(X_j - \mu)^2] = \sigma^2$$

and

$$\mathbb{E}[(\bar{X}_n - \mu)^2] = \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n},$$

we obtain

$$\begin{aligned}(n-1)\mathbb{E}(S_n^2) &= n\sigma^2 - n\left(\frac{\sigma^2}{n}\right) \\ &= n\sigma^2 - \sigma^2 \\ &= (n-1)\sigma^2.\end{aligned}$$

Therefore,

$$\boxed{\mathbb{E}(S_n^2) = \sigma^2.}$$

Proof Setup: Variance of the Sample Variance

Define the centered random variables

$$Y_j = X_j - \mu.$$

Then

$$\mathbb{E}(Y_j) = 0, \quad \mathbb{E}(Y_j^2) = \sigma^2, \quad \mathbb{E}(Y_j^4) = \mu_4.$$

Also define

$$A = \sum_{j=1}^n Y_j^2, \quad B = \sum_{j=1}^n Y_j.$$

Because

$$\bar{X}_n - \mu = \frac{B}{n},$$

the sum-of-squares identity gives

$$(n-1)S_n^2 = A - \frac{B^2}{n}.$$

Therefore,

$$S_n^2 = \frac{1}{n-1} \left(A - \frac{B^2}{n} \right).$$

Thus,

$$\text{Var}(S_n^2) = \frac{1}{(n-1)^2} \text{Var} \left(A - \frac{B^2}{n} \right).$$

Equivalently,

$$\text{Var}(S_n^2) = \frac{1}{(n-1)^2} \left[\mathbb{E} \left(A - \frac{B^2}{n} \right)^2 - ((n-1)\sigma^2)^2 \right].$$

Required Calculations

We need three quantities:

$$\mathbb{E}(A^2), \quad \mathbb{E}(AB^2), \quad \mathbb{E}(B^4).$$

First Calculation

Since

$$A^2 = \left(\sum_{j=1}^n Y_j^2 \right)^2 = \sum_{j=1}^n Y_j^4 + 2 \sum_{i < j} Y_i^2 Y_j^2,$$

independence gives

$$\mathbb{E}(A^2) = n\mu_4 + 2 \binom{n}{2} \sigma^4 = n\mu_4 + n(n-1)\sigma^4.$$

Second Calculation

Because $B^2 = A + 2 \sum_{i < j} Y_i Y_j$, we have $AB^2 = A^2 + 2A \sum_{i < j} Y_i Y_j$. Every term in the second part has at least one independent centered factor appearing to the first power, so its expectation is zero. Hence,

$$\mathbb{E}(AB^2) = \mathbb{E}(A^2) = n\mu_4 + n(n-1)\sigma^4.$$

The Fourth Moment of the Centered Sum

Recall that

$$B = \sum_{j=1}^n Y_j, \quad \mathbb{E}(Y_j) = 0.$$

When expanding B^4 , only terms in which every index appears at least twice have nonzero expectation. Thus, the surviving terms are:

$$\sum_{j=1}^n Y_j^4 \quad \text{and} \quad 6 \sum_{i < j} Y_i^2 Y_j^2.$$

Therefore,

$$B^4 = \sum_{j=1}^n Y_j^4 + 6 \sum_{i < j} Y_i^2 Y_j^2 + \text{terms having expectation zero.}$$

Taking expectations,

$$\mathbb{E}(B^4) = n\mu_4 + 6 \binom{n}{2} \sigma^4 = n\mu_4 + 3n(n-1)\sigma^4.$$

Important Point

No assumption of symmetry is needed. Terms such as $\mathbb{E}(Y_i^3 Y_j)$ vanish because

$$\mathbb{E}(Y_i^3 Y_j) = \mathbb{E}(Y_i^3) \mathbb{E}(Y_j) = 0.$$

Proof: Variance of the Sample Variance

Recall that

$$(n-1)S_n^2 = A - \frac{B^2}{n}.$$

Therefore,

$$\begin{aligned}\mathbb{E}\left[\left((n-1)S_n^2\right)^2\right] &= \mathbb{E}\left[\left(A - \frac{B^2}{n}\right)^2\right] \\ &= \mathbb{E}(A^2) - \frac{2}{n}\mathbb{E}(AB^2) + \frac{1}{n^2}\mathbb{E}(B^4).\end{aligned}$$

Substituting the previous results,

$$\begin{aligned}\mathbb{E}\left[\left((n-1)S_n^2\right)^2\right] &= \left(1 - \frac{2}{n}\right) \left[n\mu_4 + n(n-1)\sigma^4\right] \\ &\quad + \frac{1}{n^2} \left[n\mu_4 + 3n(n-1)\sigma^4\right] \\ &= \frac{(n-1)^2}{n} \mu_4 + \frac{(n-1)(n^2 - 2n + 3)}{n} \sigma^4.\end{aligned}$$

Since $\mathbb{E}[(n-1)S_n^2] = (n-1)\sigma^2$, we obtain

$$\begin{aligned}\text{Var}((n-1)S_n^2) &= \mathbb{E}\left[\left((n-1)S_n^2\right)^2\right] - \left(\mathbb{E}[(n-1)S_n^2]\right)^2 \\ &= \mathbb{E}\left[\left((n-1)S_n^2\right)^2\right] - (n-1)^2\sigma^4 \\ &= \frac{(n-1)^2}{n} \mu_4 - \frac{(n-1)(n-3)}{n} \sigma^4.\end{aligned}$$

Dividing by $(n-1)^2$,

$$\boxed{\text{Var}(S_n^2) = \frac{1}{n} \left(\mu_4 - \frac{n-3}{n} \sigma^4 \right)}.$$

What Do These Results Tell Us?

Sample Mean

$$\mathbb{E}(\bar{X}_n) = \mu, \quad \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0.$$

Thus, $\bar{X}_n$ is unbiased, and its variability decreases at rate $1/n$. Equivalently,

$$\text{SD}(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}.$$

Sample Variance

$$\mathbb{E}(S_n^2) = \sigma^2,$$

Thus, S_n^2 is an unbiased estimator of the population variance.

Its variance is

$$\text{Var}(S_n^2) = \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \sigma^4 \right) \xrightarrow{n \rightarrow \infty} 0.$$

Therefore, the accuracy of S_n^2 depends not only on σ^2 , but also on the fourth central moment μ_4 , which measures the influence of large deviations and heavy tails.

Summary: Population Quantities

- ▶ Mean, median, and mode describe different notions of center.
- ▶ The mean is sensitive to extreme values, while the median is more robust.
- ▶ Population moments summarize different features:

$$\mathbb{E}(X), \quad \text{Var}(X), \quad \text{Skew}(X), \quad \text{Kurt}(X).$$

- ▶ Skewness measures asymmetry, while kurtosis describes the influence of extreme deviations and tail behavior.
- ▶ X is symmetric about μ if

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

and all existing odd central moments are zero.

Important

Mean and variance alone do not completely determine the shape of a distribution.

Summary: Sample Quantities

Suppose

$$X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} F, \quad \mathbb{E}(X_j) = \mu, \quad \text{Var}(X_j) = \sigma^2.$$

- ▶ Sample moments are used to estimate population moments.
- ▶ The sample mean and sample variance are

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

- ▶ Both are unbiased:

$$\mathbb{E}(\bar{X}_n) = \mu, \quad \mathbb{E}(S_n^2) = \sigma^2.$$

- ▶ Their variances decrease to zero as $n \rightarrow \infty$:

$$\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}, \quad \text{Var}(S_n^2) = \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \sigma^4 \right).$$

Main idea

Larger random samples produce more stable estimates of population quantities.

Temporary page!

$\text{\LaTeX}$ was unable to guess the total number of pages correctly. As there was some unprocessed data have been added to the final page this extra page has been added to receive it.
If you rerun the document (without altering it) this surplus page will go away, because $\text{\LaTeX}$ now knows the number of pages to expect for this document.