Math Review

William Brasic

The University of Arizona

Summation Operator

Definition 1: Summation Operator

The summation operator

$$\sum_{i=1}^{n}$$

is a convenient way of expressing summations.

- The $i = 1$ means we start from index $i = 1$.
- The n means we sum until $i = n$.
- $\sum_{i \in A}$ means we only sum over the indices contained in A.

Properties of the Summation Operator

Property 1: Summation Operator

1. $\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n$.

Properties of the Summation Operator

Property 1: Summation Operator

1. $\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n$.
2. $\sum_{i=1}^{n} c = \underbrace{c + c + \ldots + c}_{n \text{ times}} = n \cdot c$.

Properties of the Summation Operator

Property 1: Summation Operator

1. $\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n$.
2. $\sum_{i=1}^{n} c = \underbrace{c + c + \ldots + c}_{n \text{ times}} = n \cdot c$.
3. $\sum_{i=1}^{n} c x_i = c \sum_{i=1}^{n} x_i$.

Properties of the Summation Operator

Property 1: Summation Operator

1. $\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n$.
2. $\sum_{i=1}^{n} c = \underbrace{c + c + \ldots + c}_{n \text{ times}} = n \cdot c$.
3. $\sum_{i=1}^{n} c x_i = c \sum_{i=1}^{n} x_i$.
4. $\sum_{i=1}^{n} (a x_i + b y_i) = a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} y_i$.

Properties of the Summation Operator

Property 1: Summation Operator

1. $\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n$.

2. $\sum_{i=1}^{n} c = \underbrace{c + c + \ldots + c}_{n \text{ times}} = n \cdot c$.

3. $\sum_{i=1}^{n} c x_i = c \sum_{i=1}^{n} x_i$.

4. $\sum_{i=1}^{n} (a x_i + b y_i) = a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} y_i$.

5. $\sum_{i=1}^{n} \frac{x_i}{y_i} \neq \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i}$.

Linear Functions

Definition 2: Linear Function

A variable y is a linear function of a single variable x if

$$y = \beta_0 + \beta_1 x.$$

- β_0 is the intercept.
- β_1 is the slope.

Non-Linear Functions

Definition 3: Non-Linear Function

A variable y is a non-linear function of a single variable x if

$$y = f(x).$$

- $f(x)$ could be x^2, $\sqrt{x}$, e^x, etc.

The Natural Logarithm Function

Definition 4: The Natural Logarithm Function

The natural logarithm is a function defined as $y = \ln(x)$.

The Natural Logarithm

Property 2: Natural Logarithm

1. $\ln(x)$ is only defined for $x > 0$.

The Natural Logarithm

Property 2: Natural Logarithm

1. $\ln(x)$ is only defined for $x > 0$.
2. $\ln(x) < 0,$ for $0 < x < 1$.

The Natural Logarithm

Property 2: Natural Logarithm

1. $\ln(x)$ is only defined for $x > 0$.
2. $\ln(x) < 0$, for $0 < x < 1$.
3. $\ln(1) = 0$.

The Natural Logarithm

Property 2: Natural Logarithm

1. $\ln(x)$ is only defined for $x > 0$.
2. $\ln(x) < 0,$ for $0 < x < 1$.
3. $\ln(1) = 0$.
4. $\ln(x) > 0,$ for $x > 1$.

The Natural Logarithm

Property 2: Natural Logarithm

1. $\ln(x)$ is only defined for $x > 0$.
2. $\ln(x) < 0,$ for $0 < x < 1$.
3. $\ln(1) = 0$.
4. $\ln(x) > 0,$ for $x > 1$.
5. $\ln\left(\prod_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} \ln(x_i),$ for $x_1, \ldots, x_n > 0$.

The Natural Logarithm

Property 2: Natural Logarithm

1. $\ln(x)$ is only defined for $x > 0$.
2. $\ln(x) < 0,$ for $0 < x < 1$.
3. $\ln(1) = 0$.
4. $\ln(x) > 0,$ for $x > 1$.
5. $\ln\left(\prod_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} \ln(x_i),$ for $x_1, \ldots, x_n > 0$.
6. $\ln\left(\frac{x_1}{x_2}\right) = \ln(x_1) - \ln(x_2),$ for $x_1, x_2 > 0$.

The Natural Logarithm

Property 2: Natural Logarithm

1. $\ln(x)$ is only defined for $x > 0$.
2. $\ln(x) < 0$, for $0 < x < 1$.
3. $\ln(1) = 0$.
4. $\ln(x) > 0$, for $x > 1$.
5. $\ln\left(\prod_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} \ln(x_i)$, for $x_1, \ldots, x_n > 0$.
6. $\ln\left(\frac{x_1}{x_2}\right) = \ln(x_1) - \ln(x_2)$, for $x_1, x_2 > 0$.
7. $\ln(x^c) = c\ln(x)$, for $x > 0$ and for any number c.

The Natural Logarithm

Example 1: Example of Property 2.5

Consider three variables x_1, x_2, and x_3 all greater than zero. Then,

$$\begin{aligned}\ln\left(\prod_{i=1}^{3} x_i\right) &= \ln\left(x_1 * x_2 * x_3\right) \\ &= \ln(x_1) + \ln(x_2) + \ln(x_3) \\ &= \sum_{i=1}^{3} \ln(x_i).\end{aligned}$$

Differential Calculus

Definition 5: Derivative

The derivative represents how a function changes as its inputs change. Formally, it represents the rate of change or the slope of a function at a particular point.

- We denote the derivative of the function f with respect to x as $\frac{df(x)}{dx}$.

Differential Calculus

Property 3: Differential Calculus

Suppose f is a function and x is a variable.

1. If f is not a function of x, then $\frac{df(x)}{dx} = 0$.

Differential Calculus

Property 3: Differential Calculus

Suppose f is a function and x is a variable.

1. If f is not a function of x, then $\frac{df(x)}{dx} = 0$.
2. If $f(x) = x$, then $\frac{df(x)}{dx} = 1$.

Differential Calculus

Property 3: Differential Calculus

Suppose f is a function and x is a variable.

1. If f is not a function of x, then $\frac{df(x)}{dx} = 0$.
2. If $f(x) = x$, then $\frac{df(x)}{dx} = 1$.
3. If $f(x) = x^a$, then $\frac{df(x)}{dx} = ax^{a-1}$.

Differential Calculus

Property 3: Differential Calculus

Suppose f is a function and x is a variable.

1. If f is not a function of x, then $\frac{df(x)}{dx} = 0$.
2. If $f(x) = x$, then $\frac{df(x)}{dx} = 1$.
3. If $f(x) = x^a$, then $\frac{df(x)}{dx} = ax^{a-1}$.
4. If $f(x) = \ln(x)$, then $\frac{df(x)}{dx} = \frac{1}{x}$.

Differential Calculus

Property 3: Differential Calculus

Suppose f is a function and x is a variable.

1. If f is not a function of x, then $\frac{df(x)}{dx} = 0$.
2. If $f(x) = x$, then $\frac{df(x)}{dx} = 1$.
3. If $f(x) = x^a$, then $\frac{df(x)}{dx} = ax^{a-1}$.
4. If $f(x) = \ln(x)$, then $\frac{df(x)}{dx} = \frac{1}{x}$.
5. If $f(x) = c + x$ for any constant c, then $\frac{df(x)}{dx} = 1$.

Differential Calculus

Property 3: Differential Calculus

Suppose f is a function and x is a variable.

1. If f is not a function of x, then $\frac{df(x)}{dx} = 0$.
2. If $f(x) = x$, then $\frac{df(x)}{dx} = 1$.
3. If $f(x) = x^a$, then $\frac{df(x)}{dx} = ax^{a-1}$.
4. If $f(x) = \ln(x)$, then $\frac{df(x)}{dx} = \frac{1}{x}$.
5. If $f(x) = c + x$ for any constant c, then $\frac{df(x)}{dx} = 1$.
6. If $f(x) = cx$ for any constant c, then $\frac{df(x)}{dx} = c$.

Partial Differentiation

Definition 6: Partial Derivative

The partial derivative is a derivative taken with respect to one variable while holding the other variables constant. It measures the rate of change of a function with respect to one of its variables in a multivariable function.

- We denote the partial derivative of the multivariable function f with respect to one of its arguments x_1 as $\frac{\partial f(x_1,\ldots,x_n)}{\partial x_1}$.

Partial Differentiation

Question 1: Partial Differentiation

Suppose $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$.

$$\frac{\partial y}{\partial x_2} = ?$$

Partial Differentiation

Question 1: Partial Differentiation

Suppose $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$.

$$\frac{\partial y}{\partial x_2} = ?$$

Answer to Question 1

$$\frac{\partial y}{\partial x_2} = \beta_2 + \beta_3 x_1.$$

First Order Condition

Definition 7: First Order Condition

The first order condition for maximizing or minimizing a function is that the partial derivatives of the function with respect to all variables must be equal to zero. Mathematically, for a function $f(x_1, x_2, \ldots, x_n)$, this means:

$$\frac{\partial f}{\partial x_1} = 0, \ \frac{\partial f}{\partial x_2} = 0, \ \ldots, \ \frac{\partial f}{\partial x_n} = 0.$$

- This condition is necessary for finding local maximum/minimum points of the function.
- Upon setting these derivatives equal to zero, we then solve for the variable in question to determine its function maximizing/minimizing value.

Why Do We Need Statistics in Econometrics?

Question 3: Why Do We Need Statistics in Econometrics?

What purpose does statistics serve in econometric analysis?

Why Do We Need Statistics in Econometrics?

Question 3: Why Do We Need Statistics in Econometrics?

What purpose does statistics serve in econometric analysis?

Answer to Question 3

We want to obtain a point estimate of a parameter and statistics provides estimation methods. After obtaining a point estimate, we can obtain confidence intervals that allow us to conduct hypothesis tests to determine the significance of that parameter.

Sample

Definition 8: Sample

A sample $x_1, \ldots, x_n = \{x_i\}_{i=1}^n$ of n observations is a subset of a population used to represent the entire group as a whole.

- A sample is used to make inferences about the population.
- A well-chosen sample should accurately reflect the characteristics of the population (the entire pool of observations we can select a sample from).
- A sample is said to be random if each observation has an equal probability of being selected.

Descriptive Statistics

Definition 9: Descriptive Statistics

Descriptive statistics involve summarizing and organizing our sample data so it can be easily understood.

- Common descriptive statistics include mean, median, variance, and standard deviation.

Sample Mean

Definition 10: Sample Mean

The sample mean (average) of a sample $\{x_i\}_{i=1}^n$ is

$$\overline{X} = \frac{1}{n}\sum_{i=1}^n x_i = n^{-1}\sum_{i=1}^n x_i.$$

- The mean is a measure of central tendency.
- $\overline{X}$ attempts to estimate the population mean μ.
- The mean is sensitive to outliers, which can skew the result.

Sample Variance

Definition 11: Sample Variance

The sample variance $\widehat{\sigma}^2$ of a sample $\{x_i\}_{i=1}^n$ is defined as

$$\widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2$$

and represents how dispersed our data is.

- $\widehat{\sigma}^2$ attempts to estimate the population variance σ^2.

Sample Standard Deviation

Definition 12: Standard Deviation

The sample standard deviation $\widehat{\sigma}$ of a sample $\{x_i\}_{i=1}^{n}$ is defined as

$$\widehat{\sigma} = \sqrt{\widehat{\sigma}^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}$$

and is a standardized version of how dispersed our data is.

- $\widehat{\sigma}$ attempts to estimate the population standard deviation σ.

Sample Covariance

Definition 13: Sample Covariance

The sample covariance of a sample $\{(x_i, y_i)\}_{i=1}^{n}$ is

$$\widehat{\sigma}_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{X})(y_i - \overline{Y}) \, .$$

Sample Correlation

Definition 14: Sample Correlation

The sample correlation of a sample $\{(x_i, y_i)\}_{i=1}^n$ is

$$\begin{aligned}\widehat{\rho}_{xy} &= \frac{\frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{X})(y_i - \overline{Y})}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n \left(x_i - \overline{X}\right)^2}\sqrt{\frac{1}{n-1}\sum_{i=1}^n \left(y_i - \overline{Y}\right)^2}} \\ &= \frac{\widehat{\sigma}_{xy}}{\widehat{\sigma}_x \widehat{\sigma}_y}.\end{aligned}$$

- This number is bounded between -1 and 1 so we can determine how strong a relationship is, unlike the covariance.

Experiments

Definition 15: Experiment

An experiment is a process by which an observation is made.

Example 2: Six Sided Dice Experiment

Rolling a six sided dice is an experiment and the observation made is the number the dice lands on.

Samples Spaces

Definition 16: Sample Space

The sample space S associated with an experiment is the set consisting of all possible sample points.

Example 3: Six Sided Dice Sample Space

The sample space of a six sided dice consists of the outcomes $1,\ 2,\ 3,\ 4,\ 5,$ and 6.

Events

Definition 17: Event

An event A in a sample space is a collection of sample points - that is, any subset of the sample space.

Example 4: Six Sided Dice Events

Some events of the six sided dice rolling experiment include:

1. The dice landing on the number 5 ($A = \{5\}$).
2. The dice landing on an odd number ($A = \{1, 3, 5\}$).
3. The dice landing on 1 or 6 ($A = \{1, 6\}$).

Probability

Definition 18: Probability

The term probability of an event A, denoted by $\mathbb{P}(A)$, is a measure of one's belief in the occurrence of a future event.

Probability

Property 4: Probability

If S is our sample space consisting of pairwise mutually exclusive events (no two events can happen at the same time) $A_1, A_2, \ldots, A_n$ in S, then

1. $\mathbb{P}(A) \geq 0$.
2. $\mathbb{P}(S) = 1$.
3. $\mathbb{P}(A_1 \cup \ldots \cup A_n) = \sum_{i=1}^{n} \mathbb{P}(A_i)$.

Probability

Question 4: Probability

What is the probability of rolling a six sided dice and landing on an odd number?

Probability

Question 4: Probability

What is the probability of rolling a six sided dice and landing on an odd number?

Answer to Question 4

The probability of any outcome of the six sided dice experiment is $\frac{1}{6}$. We are interested in the events $A_1 = \{1\}$, $A_2 = \{3\}$, and $A_3 = \{5\}$. Since each of these events is pairwise mutually exclusive, we have that

$$\mathbb{P}(A_1 \cup A_2 \cup A_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

Probability Theorems

Theorem 1: Multiplicative Law of Probability

The probability of A and ($\cap$) B occurring is defined as

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B|A) = \mathbb{P}(B) \cdot \mathbb{P}(A|B).$$

Theorem 2: Complement Law of Probability

If A is an event and A^c (A complement) is the event that A does not occur, then

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

Conditional Probability

Definition 19: Conditional Probability

The conditional probability of an event A given an event B has occurred is defined as

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Example 5: Conditional Probability

Denote the event A as rolling a six sided dice and landing on a 1. Denote the event B as rolling a six sided dice and landing on an odd number. Then,

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{2}{6} = \frac{1}{2}.$$

Independent Events

Definition 20: Independent Events

Two events are independent if the occurrence of one event does not affect the occurrence of the other. Mathematically, events A and B are independent if any of the following holds:

1. $\mathbb{P}(A|B) = \mathbb{P}(A)$.
2. $\mathbb{P}(B|A) = \mathbb{P}(B)$.
3. $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

Example 6: Independent Events

Denote the event A as rolling a six sided dice and landing on a 5. Denote the event B as flipping a coin and landing on tails. Then,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}.$$

Random Variables

Definition 21: Random Variable

A random variable is a variable that takes on numerical values determined by the outcome of an experiment.

- We typically denote random variables with capital letters such as X and Y.
- Examples of random variables include a person's height and a student's GPA.

Discrete Random Variables

Definition 22: Discrete Random Variable

A random variable X is said to be discrete if it can assume only a finite or countably infinite number of distinct values.

Probability Density Function (PDF)

Definition 23: Probability Density Function (PDF)

The probability density function f_X for a random variable X provides $f_X(x) = \mathbb{P}(X = x)$ for all x.

Probability Density Function (PDF)

Example 7: Probability Density Function (PDF)

If we let X equal 0 when the flip of a coin lands on tails while equaling 1 when the flip of a coin lands on heads, then the PDF of X can be completely characterized as

$$f_X(1) = \mathbb{P}(X = 1) = \frac{1}{2}$$
$$f_X(0) = \mathbb{P}(X = 0) = \frac{1}{2}.$$

- X is an example of a Bernoulli random variable.

Cumulative Distribution Function (CDF)

Definition 24: Cumulative Distribution Function (CDF)

The cumulative distribution function (CDF) of a random variable X, denoted by F_X, is such that $F_X(x) = \mathbb{P}(X \leq x)$ for all x.

- If X has the CDF F_X, then

$$\mathbb{P}(X > x) = 1 - \mathbb{P}(X \leq x) = 1 - F_X(x).$$

Cumulative Distribution Function (CDF)

Question 5: Cumulative Distribution Function (CDF)

If X is the random variable denoting the outcome of a dice roll, what is $\mathbb{P}(X \leq 5)$?

Cumulative Distribution Function (CDF)

Question 5: Cumulative Distribution Function (CDF)

If X is the random variable denoting the outcome of a dice roll, what is $\mathbb{P}(X \leq 5)$?

Answer to Question 5

$$\begin{aligned} F_X(5) &= \mathbb{P}(X \leq 5) \\ &= 1 - \mathbb{P}(X > 5) \\ &= 1 - \mathbb{P}(X = 6) \\ &= 1 - \frac{1}{6} \\ &= \frac{5}{6}. \end{aligned}$$

Continuous Random Variables

Definition 25: Continuous Random Variable

A random variable X is said to be continuous if it can assume an infinite number of values.

- A good example of this is a person's weight.

Expected Value

Definition 26: Expected Value

If X is a discrete random variable with outcomes $x_1, \ldots, x_n$ and PDF f_X, then the expected value of X is

$$\mathbb{E}[X] = \sum_{i=1}^{n} x_i f_X(x_i).$$

If X is a continuous random variable with PDF f_X and can take on any real number, then the expected value of X is

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx.$$

- We often denote the expected value of a random variable by μ.

Expected Value

Example 8: Expected Value

If X is a random variable denoting the outcome of a dice roll, then

$$\begin{aligned}\mathbb{E}[X] &= \sum_{i=1}^{n} x_i f_X(x_i) \\ &= \sum_{i=1}^{6} x_i f_X(x_i) \\ &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} \\ &= \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) \\ &= \frac{1}{6} \cdot 21 \\ &= 3.5.\end{aligned}$$

Expected Value

Property 5: Expected Value

If X and Y are any random variables and a and b are constants,

1. $\mathbb{E}[a] = a$.

Expected Value

Property 5: Expected Value

If X and Y are any random variables and a and b are constants,

1. $\mathbb{E}[a] = a$.
2. $\mathbb{E}[aX] = a\mathbb{E}[X]$.

Expected Value

Property 5: Expected Value

If X and Y are any random variables and a and b are constants,

1. $\mathbb{E}[a] = a$.
2. $\mathbb{E}[aX] = a\mathbb{E}[X]$.
3. $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.

Expected Value

Property 5: Expected Value

If X and Y are any random variables and a and b are constants,

1. $\mathbb{E}[a] = a$.
2. $\mathbb{E}[aX] = a\mathbb{E}[X]$.
3. $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.
4. $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ when X and Y are independent.

Conditional Expectation

Definition 27: Conditional Expectation

If X and Y are random variables, then the conditional expectation of Y given X is $\mathbb{E}[Y|X]$.

- Given we have information on X, what can we say about Y.

Conditional Expectation Properties

Property 6: Conditional Expectation

If X and Y are any random variables and g_1 and g_2 are functions of X,

1. $\mathbb{E}[g(X)|X] = g(X)$.

Conditional Expectation Properties

Property 6: Conditional Expectation

If X and Y are any random variables and g_1 and g_2 are functions of X,

1. $\mathbb{E}[g(X)|X] = g(X)$.
2. $\mathbb{E}[g_1(X)Y + g_2(X)|X] = g_1(X)\mathbb{E}[Y|X] + g_2(X)$.

Conditional Expectation Properties

Property 6: Conditional Expectation

If X and Y are any random variables and g_1 and g_2 are functions of X,

1. $\mathbb{E}[g(X)|X] = g(X)$.
2. $\mathbb{E}[g_1(X)Y + g_2(X)|X] = g_1(X)\mathbb{E}[Y|X] + g_2(X)$.
3. $\mathbb{E}[Y|X] = \mathbb{E}[Y]$ when X and Y are independent.

Conditional Expectation Properties

Property 6: Conditional Expectation

If X and Y are any random variables and g_1 and g_2 are functions of X,

1. $\mathbb{E}[g(X)|X] = g(X)$.
2. $\mathbb{E}[g_1(X)Y + g_2(X)|X] = g_1(X)\mathbb{E}[Y|X] + g_2(X)$.
3. $\mathbb{E}[Y|X] = \mathbb{E}[Y]$ when X and Y are independent.
4. $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]]$.

Co-Variability of Random Variables

Definition 28: Covariance

The covariance of two random variables X and Y is given by

$$\text{Cov}[X, Y] = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

- Covariance is a measure of how two variables are related or vary with one another.
- If $\text{Cov}[X, Y] > 0$, we say X and Y are positively related.
- If $\text{Cov}[X, Y] < 0$, we say X and Y are negatively related.

Independence of Random Variables

Definition 29: Independence of Random Variables

If X and Y are independent, then $\text{Cov}[X, Y] = 0$.

Variability of Random Variables

Definition 30: Variance of a Random Variable

The variance of a random variable X is given by

$$\begin{aligned}\mathbb{V}[X] &= \mathbb{E}\left[(X-\mu)^2\right] \\ &= \mathbb{E}\left[X^2\right] - \mu^2 \\ &= \sigma^2.\end{aligned}$$

Definition 31: Standard Deviation of a Random Variable

The standard deviation of a random variable X is given by

$$\sigma = \sqrt{\mathbb{V}[X]}.$$

Variability of Random Variables

Property 7: Variance

If X and Y are random variables and a and b are constants, then

1. $\mathbb{V}[a] = \mathbb{V}[b] = 0$.

Variability of Random Variables

Property 7: Variance

If X and Y are random variables and a and b are constants, then

1. $\mathbb{V}[a] = \mathbb{V}[b] = 0$.
2. $\mathbb{V}[aX + b] = \mathbb{V}[aX] + \mathbb{V}[b] = a^2\mathbb{V}[X]$.

Variability of Random Variables

Property 7: Variance

If X and Y are random variables and a and b are constants, then

1. $\mathbb{V}[a] = \mathbb{V}[b] = 0$.
2. $\mathbb{V}[aX + b] = \mathbb{V}[aX] + \mathbb{V}[b] = a^2\mathbb{V}[X]$.
3. If X and Y are not independent, then
   $\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y] + 2\text{Cov}[X, Y]$.

Variability of Random Variables

Property 7: Variance

If X and Y are random variables and a and b are constants, then

1. $\mathbb{V}[a] = \mathbb{V}[b] = 0$.
2. $\mathbb{V}[aX + b] = \mathbb{V}[aX] + \mathbb{V}[b] = a^2\mathbb{V}[X]$.
3. If X and Y are not independent, then $\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y] + 2\text{Cov}[X, Y]$.
4. $\mathbb{V}[Y] = \mathbb{E}[\mathbb{V}[Y|X]] + \mathbb{V}[\mathbb{E}[Y|X]]$.

Correlation

Definition 32: Correlation Between Random Variables

If X and Y are two random variables, then the correlation between X and Y is defined as

$$\text{Corr}[X, Y] = \frac{\text{Cov}[X, Y]}{\sigma_X \sigma_Y}.$$

- The correlation between any two random variables is bounded between -1 and 1.

Correlation Versus Covariance

Question 6: Correlation Versus Covariance

What is the difference between correlation and covariance?

Correlation Versus Covariance

Question 6: Correlation Versus Covariance

What is the difference between correlation and covariance?

Answer to Question 6

The correlation is bounded between -1 and 1 so by using it we can obtain both the magnitude (small or large) and direction (positive or negative) of a relationship between two variables while the covariance only gives a direction.

Distributions

Definition 33: Distribution

A distribution describes how the values of a random variable are spread or distributed. It provides the probabilities of occurrence of different possible outcomes in an experiment. Distributions can be represented using probability density functions.

Normal Distribution

Definition 34: Normal Distribution

A normal distribution, also known as a Gaussian distribution, is a continuous probability distribution characterized by its bell-shaped curve. The probability density function (PDF) of a normally distributed random variable X with mean μ and variance σ^2 is given by:

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

- The normal distribution is symmetric around its mean.
- If X is a normally distributed random variable with mean μ and variance σ^2, we write $X \sim \mathbb{N}(\mu, \sigma^2)$.

Normal Distribution

Property 8: Normal Distribution

If X and Y are independent normal random variables with mean μ_X and μ_Y, respectively, and variance σ_X^2 and σ_Y^2, then

1. $X + Y \sim \mathbb{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.
2. $cX + b \sim \mathbb{N}(c\mu_X + b, c^2\sigma_X^2)$ for any constants c and b.

Standardizing a Random Variable

Definition 35: Standardizing a Random Variable

If X is a random variable with mean μ and variance σ^2, then its standardized version is

$$Z = \frac{X - \mu}{\sigma}$$

with mean $\mathbb{E}[Z] = 0$ and variance $\mathbb{V}[Z] = 1$.

- If X is normally distributed, then Z follows the standard normal distribution (i.e., $Z \sim \mathbb{N}(0, 1)$).

Independent and Identically Distributed (i.i.d.)

Definition 36: Independent and Identically Distributed (i.i.d.)

A sample $\{x_i\}_{i=1}^{n}$ is said to be independent and identically distributed (i.i.d.) if each random variable in the sample:

1. Is independent: The occurrence of any one variable does not affect the others.
2. Is identically distributed: All variables follow the same probability distribution.

Central Limit Theorem

Theorem 3: Central Limit Theorem

If $X_1, X_2, \ldots, X_n$ are i.i.d. random variables with mean μ and variance σ^2, then as n approaches infinity,

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \xrightarrow{d} \mathbb{N}(0, 1).$$

- This means that $\overline{X}$ converges toward a normal distribution with mean μ and variance $\frac{\sigma^2}{n}$ for large n.

Statistical Inference

Definition 37: Terms in Statistical Inference

1. A parameter is the true value of what we are estimating.
2. An estimator is a random variable that attempts to estimate the parameter.
3. An estimate is the value produced by the estimator.
4. A sampling distribution is the distribution of our estimator.

Statistical Inference

Example 9: Example of Terms in Statistical Inference

- The mean of a sample $\{x_i\}_{i=1}^n$ given by $\overline{X} = n^{-1}\sum_{i=1}^{n} x_i$ is an estimator.
- The number given by this sample is an estimate of the parameter μ.
- If we were to repeatedly draw samples and compute $\overline{X}$ for each sample, we would form the sampling distribution for $\overline{X}$.

Unbiased Estimators

Definition 38: Unbiased Estimator

An estimator $\widehat{\beta}$ of the parameter β is unbiased if

$$\mathbb{E}\left[\widehat{\beta}\right] = \beta.$$

- On average, our estimator gives us the correct answer.

Consistency

Definition 39: Consistency

An estimator $\widehat{\beta}$ of the parameter β is consistent if

$$\mathbb{P}\left(\left|\widehat{\beta} - \beta\right| > \epsilon\right) \to 0$$

as $n \to \infty$ for any $\epsilon > 0$.

- As we get a larger sample, our estimator converges toward the truth.
- We often write $\widehat{\beta} \xrightarrow{p} \beta$.

Efficient

Definition 40: Efficient

An unbiased estimator $\widehat{\beta}$ of the parameter β is efficient in the class of unbiased estimators if

$$\mathbb{V}\left[\widehat{\beta}\right] \leq \mathbb{V}\left[\widetilde{\beta}\right]$$

for any other unbiased estimator $\widetilde{\beta}$ of β.

- If we were to gather multiple samples, estimate $\widehat{\beta}$ for each sample, form a sampling distribution, and compute the variance of the sampling distribution, $\widehat{\beta}$ would have the smallest such variance.

Thank You!