Probability and Random Variables

STA 2300

Chapter 5

Random Samples

- Two random samples from the same population will likely result in different statistics
  - A result of the **variability** that results from choosing at **random**.

Random Processes

- Exhibit **chance behavior**:
  - Individual outcomes are uncertain, but there is a regular distribution in a large number of repetitions.
  - **Unpredictable in the short-run**, but produce a predictable pattern in the long-run.
- Experiments and observational studies are types of random processes.
  - Will refer to both generically as **experiments**

Outcomes

- The most basic possible results from random processes.
- Examples:
  - Rolling a 1 on a die.
  - Flipping a head on a coin.
  - Drawing the 2 of diamonds from a deck of cards.

Events

- Outcomes or a collection of outcomes that share some property of interest.
- Examples:
  - Rolling an odd number.
  - Flipping exactly two heads on four coin tosses.
  - Drawing a heart from a deck of cards.

The Sample Space *S*

- The set of ***all*** possible outcomes of a random process.
- Events are *subsets* of the sample space *S*.

Example

- Describe the sample space S:
- A new business is started. After two years, it is either still in business or it has closed.

Example

- Describe the sample space S:
- A rust prevention treatment is applied to a new car. The response variable is the length of time before rust begins to develop on the vehicle.

Example

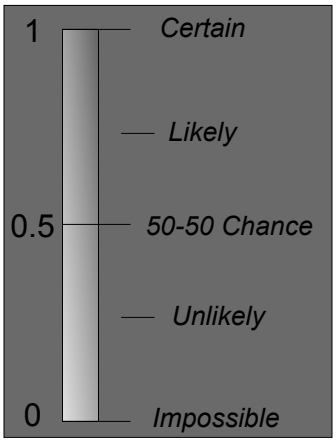
- Describe the sample space S:
- A quality inspector examines four portable CD players and rates each as either "acceptable" or "unacceptable." You record the sequence of ratings.

Example

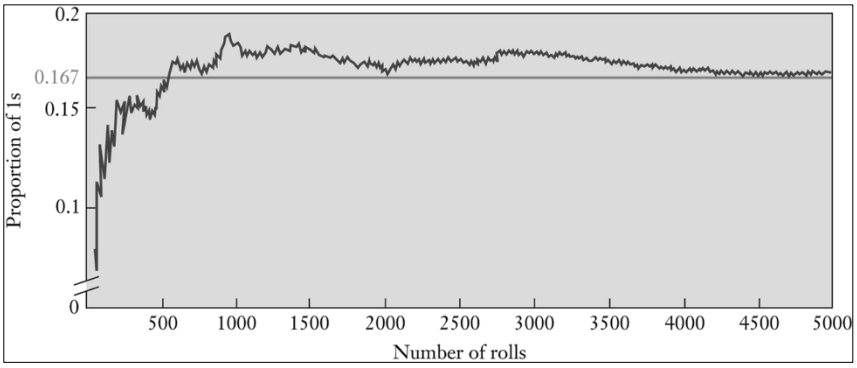
- Describe the sample space S:
- A quality inspector examines four portable CD players and rates each as either "acceptable" or "unacceptable." You record the number of units rated "acceptable."

Probability

- Reflects the proportion of times an event will occur in a very long series of trials.
- Expressed as P(event).
- Always a number between 0 and 1.



Example: Rolling a Die



Calculating Probabilities

- Empirical Approach
- Theoretical Approach

Empirical Approach

1. Repeat or observe a process many times and count the number of times event A occurs.
2. Estimate *P(A)*:

$$P(A) = \frac{\#\ times\ A\ occurs}{total\ \#\ observations}$$

Example: Rolling a Die

- You roll a die 5000 times.
- The number of times you roll a 5 is 830.
- Based on this information, calculate P(roll a 5).
- Is it likely that this die is ***fair***?
  - Fair meaning that each number is equally likely to be rolled.

Theoretical Approach

- In order to develop theoretical models, we need **set theory** to help us establish rules.

Complements

- The **complement** of event *A* is the set of all outcomes in *S* that are not contained in *A*.
- The event that A does ***not*** occur
- Denoted by *"not A"*.

Unions

- The **union** of *A* and *B* is the event consisting of the outcomes that are ***either*** *in A* ***or*** *B* ***or*** *in both*.
- Denoted by *"A or B"*

Intersections

- The **intersection** of events *A* and *B* is the event consisting of all outcomes that are in ***both A and B***.
- Denoted by ***"A & B"***

The Null Set

- The event consisting of no outcomes.
- Also called the null event.
- Denoted by $\varnothing$
- If $A \& B = \varnothing$, the events are said to be **disjoint** or **mutually exclusive**.

Probability Rules

- **Rule 1**: The probability P(A) of any event A satisfies $0 \le P(A) \le 1$.
  - If A can never occur, then P(A) = 0.
  - If A is certain to occur, then P(A) = 1.
- **Rule 2**: If *S* is the sample space, then P(*S*)=1.
  - All possible outcomes must have probabilities that sum to one.

Probability Rules

- **Rule 3**: For *disjoint* events A and B. The probability of either event occurring is the sum of the individual probabilities.
  - P(A or B) = P(A) + P(B)
- Events A and B are **disjoint** if:
  - P(A and B) = 0

Probability Rules

- **Rule 4**: The complement of any event A, A^c, is the event that A does not occur.
  - $P(A) + P(A^c) = 1$
  - $P(A^c) = 1 - P(A)$

Probability Rules

- **Rule 6**: For *any* events A and B, the probability of either event occurring is the following:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

Note

- The addition rules can be extended *beyond* two events.
- **Example**: For disjoint events A, B, and C:

$$P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C)$$

Example: Wealth

- A = Event that a selected household is wealthy (income > $100,000)
- B = Event that the householder completed college
- P(A) = 0.13
- P(B) = 0.25
- P(A and B) = 0.08

Example: Wealth

- Draw a Venn diagram of the sample space.
- P(A or B) = ?
- Shade the event C = The household is educated, but not wealthy.
- P(C) = ?

Example: Economy

- How would you rate the economic conditions today?

What is the probability of a positive ranking?

Outcome	Probability
Excellent	0.03
Good	0.12
Fair	0.45
Poor	0.39
No opinion	0.01

Determining Probabilities

- Let E_1, E_2, E_3, … be the outcomes in sample space ***S***.
- The probability of compound event A is the sum of the $P(E_i)$'s for all E_i's in A:

$$P(A) = \sum_{all\ E_i's\ in\ A} P(E_i)$$

Equal-likelihood Model

- Let N be the total number of outcomes in ***S***.
- The probability of any outcome is $\frac{1}{N}$
- Let *f* be the number of outcomes in event A.

$$P(A) = \frac{f}{N}$$

Probability Models

- Consists of two parts:
  - The sample space S
  - A way of assigning probabilities to events
- Common ways of depicting a model:
  - Probability distribution table
  - Probability histogram
  - Density curve

Example: Dice

- What is the probability of getting doubles when rolling two fair dice?

Random Variables

- A **random variable** is a variable whose value is a numerical outcome of a random process.
- The **probability distribution** of a random variable X tells us what values X can take and how to assign probabilities to those values.

Discrete Probability Models

- The sample space S is **discrete**.
  - All possible values can be listed
- To assign probabilities:
  - List the probabilities of all outcomes.
    - These MUST add to 1.
  - The probability of any event is the sum of the probabilities of the outcome making up the event.

Discrete Probability Models

Model Expressed in a Table

Value of X	x_1	x_2	x_3	…	x_k
Probability	p_1	p_2	p_3	…	p_k

Properties

1. For every p_i, $0 \le p_i \le 1$
2. $p_1 + p_2 + p_3 + \ldots + p_k = 1$

Continuous Probability Models

- Assign probabilities as areas under a *density curve*.
- The area under the curve and within any range of values is the probability of an outcome in that range.
  - ALL continuous probability models assign probability 0 to every **individual** outcome!

Example: AC Units

Weekly Air Conditioning Units Ordered

Units ordered X	0	1	2	3	4	5
Probability	0.05	0.15	0.27	0.33	0.13	0.07

Is this a valid probability model?

Construct a probability histogram

$P(X < 2) = ???$

$P(X \leq 2) = ???$

$P(X > 2) = ???$

$P(2 \leq X \leq 4) = ???$

Mean of a Discrete RV

- X is a discrete RV with a set of possible values *D*
- The **population mean,** denoted by μ_X or μ, is given by:

$$\mu = \mu_X = \sum_{x \in D} x \cdot P(X = x)$$

- Also called the **expected value of X**

Example: 2 Coin Tosses

Let X = # of tails in 2 tosses. What is μ?

x	0	1	2
P(X=x)	1/4	1/2	1/4
Product	0	1/2	1/2

μ = 0 + 0.5 + 0.5 = 1

Example: AC Units

Weekly Air Conditioning Units Ordered

Units ordered X	0	1	2	3	4	5
Probability	0.05	0.15	0.27	0.33	0.13	0.07

Calculate the expected number of units ordered weekly.

Variance of X

- X is a discrete RV with a set of possible values *D* and mean μ.
- The **variance** of X, denoted by σ^2, is given by:

$$\sigma^2 = \sigma_X^2 = \sum_{x \in D} (x - \mu)^2 \cdot P(X = x)$$

- Also called the **population variance**.

Example: 2 Coin Tosses

Let X = # of tails in 2 tosses. What is σ^2?

x	0	1	2
x-μ	-1	0	1

Recall E(X)= 1

$(x-\mu)^2$	1	0	1
P(X=x)	1/4	1/2	1/4
Product	1/4	0	1/4

σ^2= 1/4+0+1/4=1/2

Alternative Formula for σ^2

- The following formula serves as an alternative to the definition for σ^2

$$\sigma^2 = \sigma_X^2 = \left[\sum_{x \in D} x^2 \cdot P(X = x)\right] - \mu^2$$

- σ is the **population standard deviation**.

Example

Weekly Air Conditioning Units Ordered

Units ordered X	0	1	2	3	4	5
Probability	0.05	0.15	0.27	0.33	0.13	0.07

Calculate the standard deviation of X.