STATISTICS

Subtitle

10/12/18

TOPICS IN STATISTICS

Probability Theory:

- Events and their Probabilities
- Rules of Probability
- Conditional Probability and Independence
- Distribution of a Random Variable
- Moment Generating functions Central
- Limit Theorem
- Expectation
- Variance
- Measures of Central tendency
- Measures of Dispersion
- Skewness and Kurtosis

Hypothesis, Parametric and Non-Parametric Tests

- Sample and Population
- Formulate the Hypothesis
- Select an Appropriate Test
- Choose level of Significance
- Calculate Test Statistics
- Determine the Probability
- Compare the Probability and Make Decision
- One Sample T-Test
- Two Independent Samples Tests
- Paired T-test
- Proportional Test
- Non Parametric One Sample Test
- Chi Square Test
- Z Test
- F Test

Linear Regression

- Assumptions
- Hypothesis
- Variable and Model Significance
- Ordinary Least Squares Notion
- Regression Table
- Anova Table
- Multicollinearity
- Heteroscedasticity
- Model Specification
- L1 & L2 Regularization

Logistic Regression

- Assumptions
- Reason for the Logit Transform
- Logit Transformation
- Hypothesis
- Variable and Model Significance
- Maximum Likelihood Concept
- Log Odds and Interpretation
- Regression Table
- Null Vs Residual Deviance
- Chi Squre Test
- ROC Curve
- Model Specification

Case for Prediction Probe

- Same steps as Linear regression Case
- Model Parameter Significance Evaluation
- Drawing the ROC Curve
- Estimating the Classification Model Hit Ratio
- Isolating the Classifier for Optimum Results

STATISTICS – PROBABILITY THEORY

DESCRIPTIVE AND INFERENTIAL STATISTICS

Statistics can be broken into two basic types:

- Descriptive Statistics

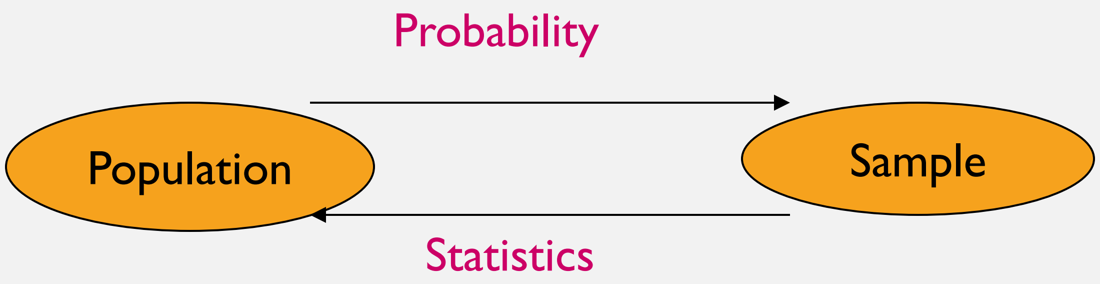
  What is this?

- Inferential Statistics

  Methods that making decisions or predictions about a population based on sampled data.

WHY LEARN PROBABILITY?

- Nothing in life is certain. In everything we do, we gauge the chances of successful outcomes, from business to medicine to the weather
- A probability provides a quantitative description of the chances or likelihoods associated with various outcomes
- It provides a bridge between descriptive and inferential statistics



INTRODUCTION

People use the term probability many times each day. For example, physician says that a patient has a 50-50 chance of surviving a certain operation. Another physician may say that she is 95% certain that a patient has a particular disease

PROBABILISTIC VS STATISTICAL REASONING

- Suppose we know exactly the proportions of car makes in India. Then I can find the probability that the first car I see in the street is a Honda. This is probabilistic reasoning as we know the population and predict the sample
- Now suppose that we do not know the proportions of car makes in California, but would like to estimate them. I observe a random sample of cars in the street and then I have an estimate of the proportions of the population. This is statistical reasoning

BASIC CONCEPTS

- An **experiment** is the process by which an observation (or measurement) is obtained.
- An **event** is an outcome of an experiment, usually denoted by a capital letter.
  - The basic element to which probability is applied
  - When an experiment is performed, a particular event either happens, or it doesn't!

EXPERIMENTS AND EVENTS

- **Experiment: Record an age**
  - A: person is 30 years old
  - B: person is older than 65
- **Experiment: Toss a die**
  - A: observe an odd number
  - B: observe a number greater than 2

BASIC CONCEPTS

- Two events are **mutually exclusive** if, when one event occurs, the other cannot, and vice versa.

- **Experiment: Toss a die**
  - –A: observe an odd number
  - –B: observe a number greater than 2
  - –C: observe a 6
  - –D: observe a 3

Not Mutually Exclusive

Mutually Exclusive

B and C?
B and D?

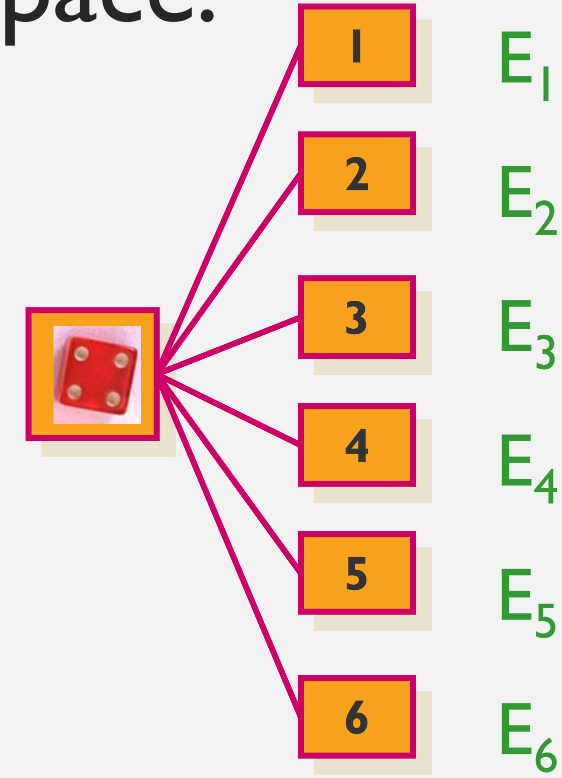
BASIC CONCEPTS

- An event that cannot be decomposed is called a **simple event.**
- Denoted by E with a subscript. E.g, E_1
- Each simple event will be assigned a probability, measuring "how often" it occurs.
- The set of all simple events of an experiment is called the **sample space, S.**

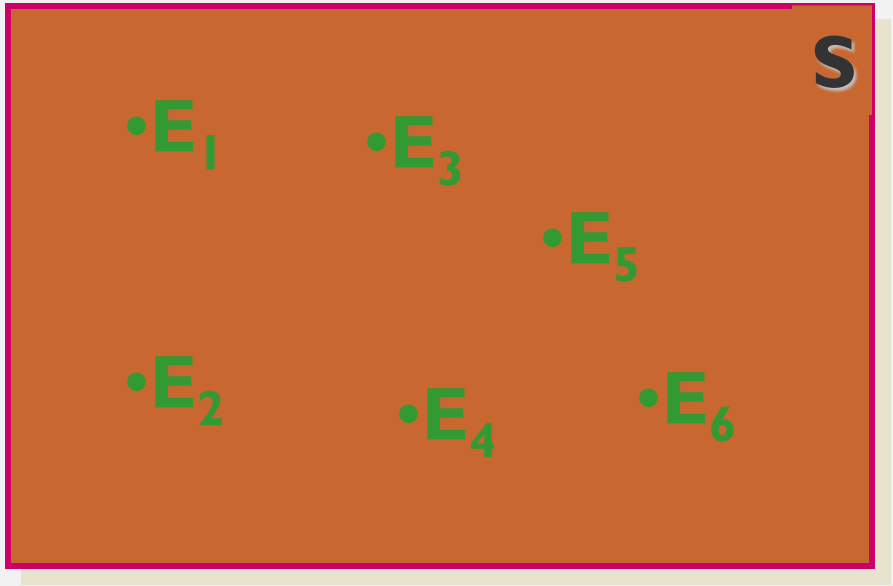
EXAMPLE

- **The die toss:** Simple events:

Sample space:



$$S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$$



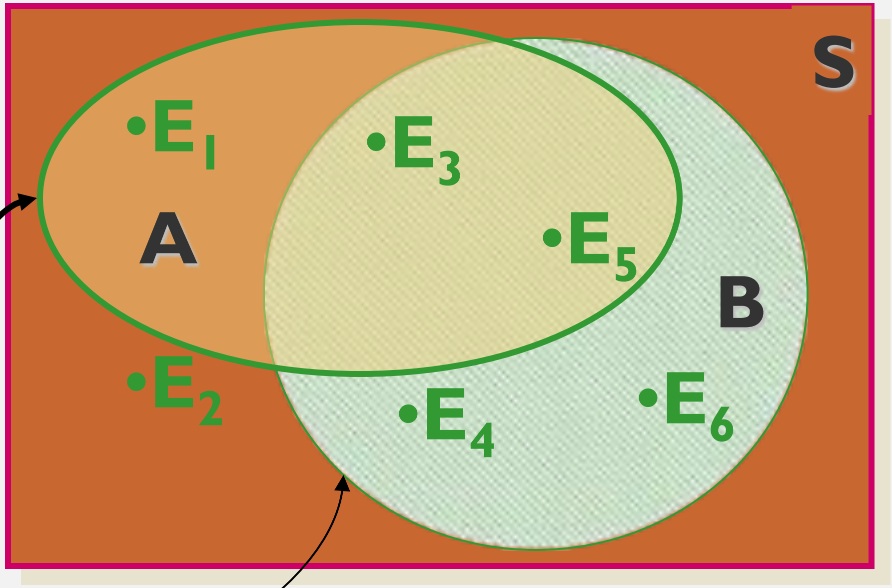
BASIC CONCEPTS

- An **event** is a collection of one or more **simple events.**

- **The die toss:**
  - –A: an odd number
  - –B: a number > 2

$A = \{E_1, E_3, E_5\}$

$B = \{E_3, E_4, E_5, E_6\}$



THE PROBABILITY OF AN EVENT

- The probability of an event A measures "how often" A will occur. We write **P(A).**
- Suppose that an experiment is performed n times. The relative frequency for an event A is

$$\frac{\text{Number of times A occurs}}{n} = \frac{f}{n}$$

- If we let n get infinitely large,

$$\boxed{P(A) = \lim_{n\to\infty} \frac{f}{n}}$$

THE PROBABILITY OF AN EVENT

- P(A) must be between 0 and 1.
    - If event A can never occur, P(A) = 0. If event A always occurs when the experiment is performed, P(A) =1.
- The sum of the probabilities for all simple events in S equals 1.

- The **probability of an event A** is found by adding the probabilities of all the simple events contained in A.

FINDING PROBABILITIES

- Probabilities can be found using
  - Estimates from empirical studies
  - Common sense estimates based on equally likely events.

- **Examples:**
  - –Toss a fair coin.

P(Head) = 1/2

– Suppose that 10% of the U.S. population has red hair. Then for a person selected at random, P(while hair) = .10

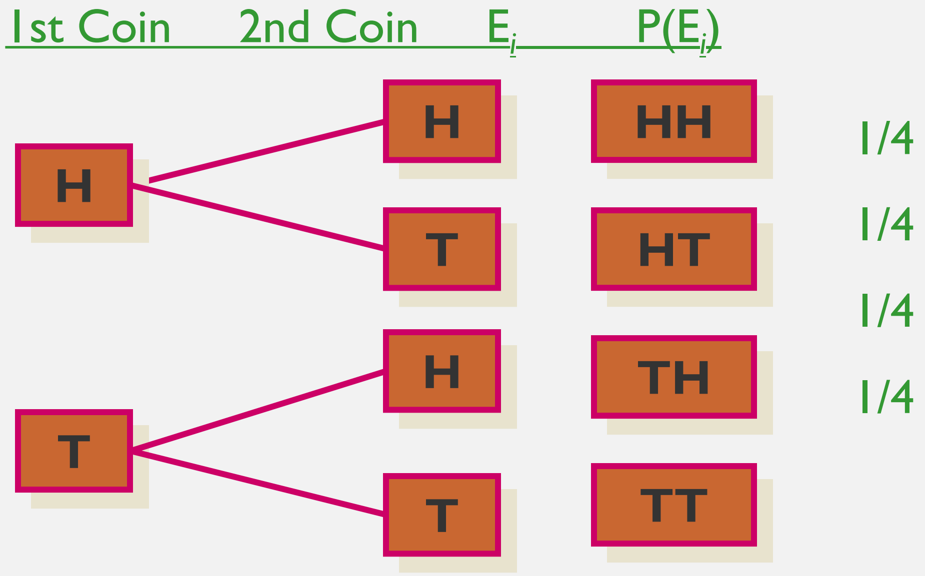
USING SIMPLE EVENTS

- The probability of an event A is equal to the sum of the probabilities of the simple events contained in A
- If the simple events in an experiment are **equally likely,** you can calculate

$$P(A) = \frac{n_A}{N} = \frac{\text{number of simple events in A}}{\text{total number of simple events}}$$

EXAMPLE 1

Toss a fair coin twice. What is the probability of observing at least one head?



P(at least 1 head)

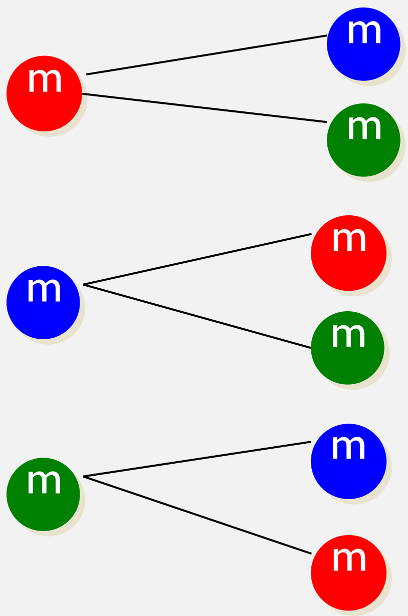
$= P(E_1) + P(E_2) + P(E_3)$

$= 1/4 + 1/4 + 1/4 = 3/4$

EXAMPLE 2

A bowl contains three Marbles , one red, one blue and one green. A child selects two M&Ms at random. What is the probability that at least one is red?

1st M&M	2nd M&M	E_i	$P(E_i)$
		RB	1/6
		RG	1/6
		BR	1/6
		BG	1/6
		GB	1/6
		GR	1/6



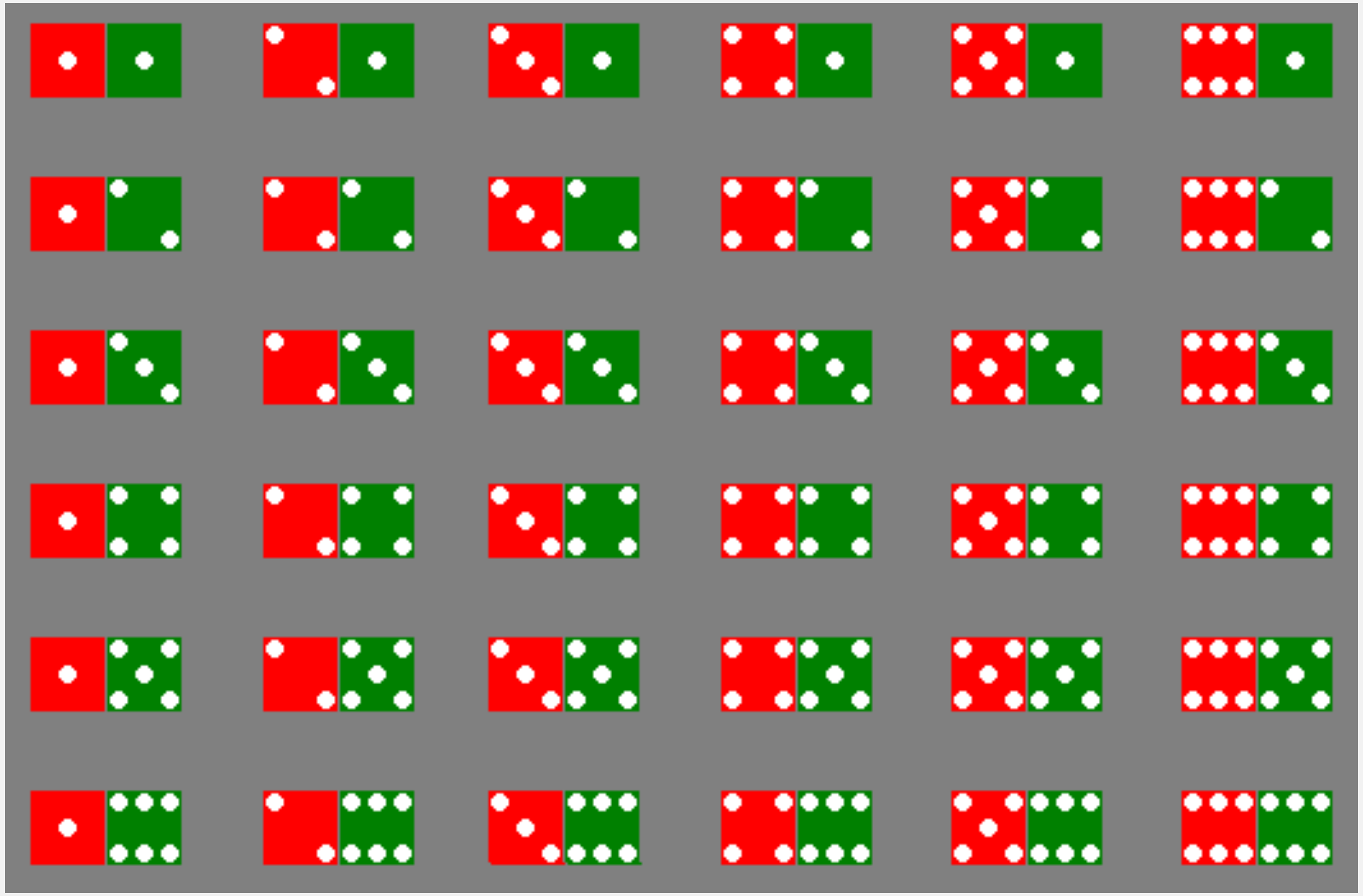
P(at least 1 red)

= P(RB) + P(BR)+ P(RG) + P(GR)

= 4/6 = 2/3

EXAMPLE 3

The sample space of throwing a pair of dice is



EXAMPLE 3

Event	Simple events	Probability
Dice add to 3	(1,2),(2,1)	2/36
Dice add to 6	(1,5),(2,4),(3,3), (4,2),(5,1)	5/36
Red die show 1	(1,1),(1,2),(1,3), (1,4),(1,5),(1,6)	6/36
Green die show 1	(1,1),(2,1),(3,1), (4,1),(5,1),(6,1)	6/36

COUNTING RULES

- Sample space of throwing 3 dice has 216 entries, sample space of throwing 4 dice has 1296 entries, …
- At some point, we have to stop listing and start thinking …
- We need some counting rules

THE *MN* RULE

- If an experiment is performed in two stages, with ***m*** ways to accomplish the first stage and ***n*** ways to accomplish the second stage, then there are ***mn*** ways to accomplish the experiment.
- This rule is easily extended to ***k*** stages, with the number of ways equal to

$$n_1 n_2 n_3 \dots n_k$$

Example: Toss two coins. The total number of simple events is:

$$2 \times 2 = 4$$

EXAMPLES

Example: Toss three coins. The total number of simple events is:

$2 \times 2 \times 2 = 8$

Example: Toss two dice. The total number of simple events is:

$6 \times 6 = 36$

Example: Toss three dice. The total number of simple events is:

$6 \times 6 \times 6 = 216$

Example: Two marbles are drawn from a dish containing two red and two blue ones. The total number of simple events is:

$4 \times 3 = 12$

PERMUTATIONS

- The number of ways you can arrange ***n*** distinct objects, taking them ***r*** at a time is

$$P_r^n = \frac{n!}{(n-r)!}$$

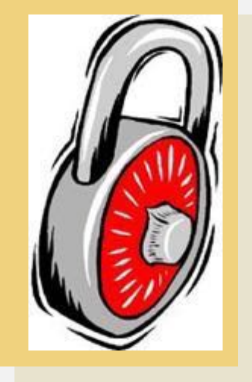
$$\text{where } n! = n(n-1)(n-2)\ldots(2)(1) \text{ and } 0! \equiv 1.$$

Example: How many 3-digit lock combinations can we make from the numbers 1, 2, 3, and 4?

The order of the choice is important!

$$P_3^4 = \frac{4!}{1!} = 4(3)(2) = 24$$

EXAMPLES



Example: A lock consists of five parts and can be assembled in any order. A quality control engineer wants to test each order for efficiency of assembly. How many orders are there?

The order of the choice is important!

$$P_5^5 = \frac{5!}{0!} = 5(4)(3)(2)(1) = 120$$

COMBINATIONS

The number of distinct combinations of ***n*** distinct objects that can be formed, taking them ***r*** at a time is

$$C_r^n = \frac{n!}{r!(n-r)!}$$

Example: Three members of a 5-person committee must be chosen to form a subcommittee. How many different subcommittees could be formed?

The order of the choice is not important!

$$C_3^5 = \frac{5!}{3!(5-3)!} = \frac{5(4)(3)(2)1}{3(2)(1)(2)1} = \frac{5(4)}{(2)1} = 10$$