## Inferential Statistics and Probability
a Holistic Approach

Chapter 4
Probability



1

## Probability

- Classical probability
  - Based on mathematical formulas
- Empirical probability
  - Based on the relative frequencies of historical data.
- Subjective probability
  - "one-shot" educated guess.

2

## Examples of Probability

- What is the probability of rolling a four on a 6-sided die?
- What percentage of De Anza students live in Cupertino?
- What is the chance that the Golden State Warriors will be NBA champions in 2018?

3

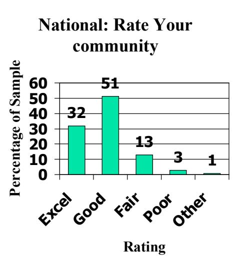
## Classical Probability

- Event
  - A result of an experiment
- Outcome
  - A result of the experiment that cannot be broken down into smaller events
- Sample Space
  - The set of all possible outcomes
- Probability Event Occurs
  - # of elements in Event / # Elements in Sample Space
- Example – flip two coins, find the probability of exactly 1 head.
  - {HH, HT, TH, TT}
  - P(1 head) = 2/4 = .5

4

## Empirical Probability

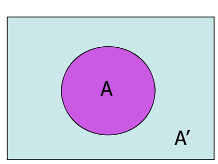
- Historical Data
- Relative Frequencies
- Example: What is the chance someone rates their community as good or better?
- **0.51 + 0.32 = 0.83**




5

## Rule of Complement

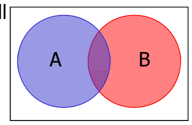
- Complement of an event
- The event does not occur
- A′ is the complement of A
- P(A) + P(A′) =1
- P(A) = 1 – P(A′)




6

## Additive Rule

- The UNION of two events A and B is that either A or B occur (or both). (All colored parts)
- The INTERSECTION of two events A and B is that both A and B will occur. (Purple Part only)
- Additive Rule:
  P(A or B) = P(A) + P(B) – P(A and B)



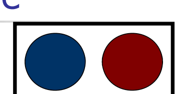

7

## Example

- In a group of students, 40% are taking Math, 20% are taking History.
- 10% of students are taking both Math and History.
- Find the Probability of a Student taking either Math or History or both.
- P(M or H) = 40% + 20% - 10% = 50%

8

## Mutually Exclusive



- Mutually Exclusive
- Both cannot occur
- If A and B are mutually exclusive, then
  - P(A or B) = P(A) + P(B)
- Example roll a die
  - A: Roll 2 or less B: Roll 5 or more
  - P(A)=2/6 P(B)=2/6
  - P(A or B) = P(A) + P(B) = 4/6

9

## Conditional Probability

- The probability of an event occuring GIVEN another event has already occurred.
- P(A|B) = P(A and B) / P(B)

- Example: Of all cell phone users in the US, 15% have a smart phone with AT&T. 25% of all cell phone users use AT&T. Given a selected cell phone user has AT&T, find the probability the user also has a smart phone.

- A=AT&T subscriber B=Smart Phone User
- P(A and B) = .15 P(A)=.25
- P(B|A) = .15/.25 = .60

10

## Contingency Tables

- Two data items can be displayed in a contingency table.
- Example: auto accident during year and DUI of driver.

| | Accident | No Accident | Total |
|---|---|---|---|
| **DUI** | 70 | 130 | 200 |
| **Non- DUI** | 30 | 770 | 800 |
| **Total** | 100 | 900 | 1000 |

11

## Contingency Tables

| | Accident | No Accident | Total |
|---|---|---|---|
| **DUI** | 70 | 130 | 200 |
| **Non- DUI** | 30 | 770 | 800 |
| **Total** | 100 | 900 | 1000 |

Given the Driver is DUI, find the Probability of an Accident.

A=Accident D=DUI

P(A and D) = .07 P(D) = .2

P(A|D) = .07/.2 = .35

12

## Marginal, Joint and Conditional Probability

- **Marginal Probability** means the probability of a single event occurring.
- **Joint Probability** means the probability of the union or intersection of multiple events occurring.
- **Conditional Probability** means the probability of an event occurring given that another event has already occurred.

13

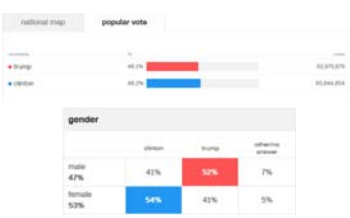
## Creating Contingency Tables

- You can create a hypothetical contingency table from reported cross tabulated data.
- First choose a convenient sample size (called a radix) like 10000.
- Then apply the reported marginal probabilities to the radix of one of the variables.
- Then apply the reported conditional probabilities to the total values of one of the other variable.
- Complete the table with arithmetic.

14

## Example

Create a two-way table from the cross tabulation of gender from the 2016 election results (from CNN)




15

## Example

First select a radix (sample size) of 10000

| | GENDER | | |
|---|---|---|---|
| VOTED FOR | Female | Male | Total |
| Trump | | | |
| Clinton | | | |
| Other | | | |
| Total | | | 10000 |

16

## Example

Then apply the marginal probabilities to the radix (53% female, 47% male)

| | GENDER | | |
|---|---|---|---|
| VOTED FOR | Female | Male | Total |
| Trump | | | |
| Clinton | | | |
| Other | | | |
| Total | 5300 | 4700 | 10000 |

17

## Example

Then apply the cross tabulated percentages for each gender. Make sure the numbers add up.

| | GENDER | | |
|---|---|---|---|
| VOTED FOR | Female | Male | Total |
| Trump | 2173 | 2444 | |
| Clinton | 2862 | 1927 | |
| Other | 265 | 329 | |
| Total | 5300 | 4700 | 10000 |

18

## Example

Finally, complete the table using arithmetic.

| | GENDER | | |
|---|---|---|---|
| VOTED FOR | Female | Male | Total |
| Trump | 2173 | 2444 | 4617 |
| Clinton | 2862 | 1927 | 4789 |
| Other | 265 | 329 | 594 |
| Total | 5300 | 4700 | 10000 |

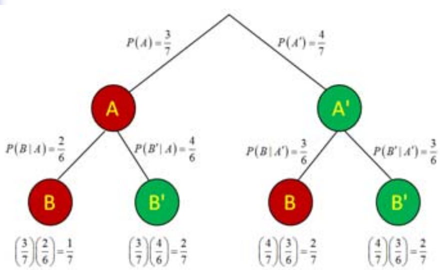
19

## Multiplicative Rule

- P(A and B) = P(A) x P(B|A)
- P(A and B) = P(B) x P(A|B)
- Example: A box contains 4 green balls and 3 red balls. Two balls are drawn. Find the probability of choosing two red balls.
- A=Red Ball on 1st draw   B=Red Ball on 2nd Draw
- P(A)=3/7   P(B|A)=2/6
- P(A and B) = (3/7)(2/6) = 1/7

20

## Multiplicative Rule – Tree Diagram




21

## Independence

- If A is not dependent on B, then they are INDEPENDENT events, and the following statements are true:
  - P(A|B)=P(A)
  - P(B|A)=P(B)
  - P(A and B) = P(A) x P(B)

22

## Example

| | Accident | No Accident | Total |
|---|---|---|---|
| **DUI** | 70 | 130 | 200 |
| **Non- DUI** | 30 | 770 | 800 |
| **Total** | 100 | 900 | 1000 |

A: Accident D:DUI Driver

P(A) = .10 P(A|D) = .35 (70/200)

Therefore A and D are **DEPENDENT** events as P(A) < P(A|D)

23

## Example

| | Accident | No Accident | Total |
|---|---|---|---|
| **Domestic Car** | 60 | 540 | 600 |
| **Import Car** | 40 | 360 | 400 |
| **Total** | 100 | 900 | 1000 |

A: Accident D:Domestic Car

P(A) = .10 P(A|D) = .10 (60/600)

Therefore A and D are **INDEPENDENT** events as P(A) = P(A|D)

Also P(A and D) = P(A)xP(D) = (.1)(.6) = .06

24

## Random Sample

- A random sample is where each member of the population has an equally likely chance of being chosen, and each member of the sample is INDEPENDENT of all other sampled data.

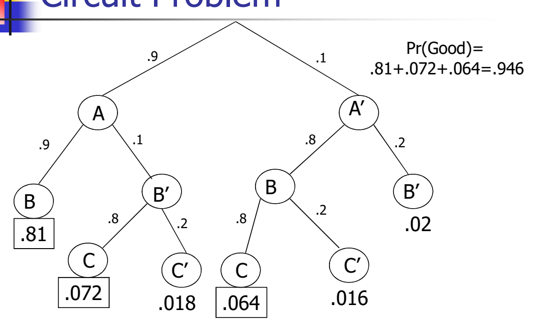
25

## Tree Diagram method

- Alternative Method of showing probability
- Example: Flip Three Coins
- Example: A Circuit has three switches. If at least two of the switches function, the Circuit will succeed. Each switch has a 10% failure rate if all are operating, and a 20% failure rate if one switch has already failed. Find the probability the circuit will succeed.

26

## Circuit Problem




27

## Switching the Conditionality

- Often there are questions where you desire to change the conditionality from one variable to the other variable
- First construct a tree diagram.
- Second, create a Contingency Table using a convenient radix (sample size)
- From the Contingency table it is easy to calculate all conditional probabilities.

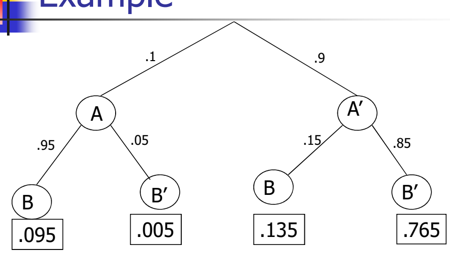
28

## Example

- 10% of prisoners in a Canadian prison are HIV positive.
- A test will correctly detect HIV 95% of the time, but will incorrectly "detect" HIV in non-infected prisoners 15% of the time (false positive).
- If a randomly selected prisoner tests positive, find the probability the prisoner is HIV+

29

## Example




A=Prisoner is HIV+
B=Test is Positive for HIV

30

## Example

| | HIV+ A | HIV- A' | Total |
|---|---|---|---|
| **Test+ B** | 950 | 1350 | 2300 |
| **Test- B'** | 50 | 7650 | 7700 |
| **Total** | 1000 | 9000 | 10000 |

$$P(A \mid B) = \frac{950}{2300} \approx .413$$

31