Chapter 1 Probability

1.1 Basic Concepts

1

Probability

A researcher claims that 10% of a large population have disease H.

A random sample of 100 people is taken from this population and examined.

If 20 people in this random sample have the disease, what does it mean? **How likely** would this happen if the researcher is right?

2

A Simple Example

- What's the probability of getting a **head** on the toss of a single fair coin? Use a scale from **0** (**no way**) to **1** (**sure thing**).
- **So toss a coin twice.** Do it! Did you get one head & one tail? What's it all mean?

3

Sample Space and Probability

- **Random Experiment**: (Probability Experiment) an experiment whose outcomes depend on chance.
- **Sample Space (S)**: collection of all possible outcomes in random experiment.
- **Event (E)**: a collection of outcomes

4

Sample Space and Event

- Sample Space:

S = {Head, Tail}
S = {Life span of a human} = {x | x≥0, x∈R}
S = {All individuals live in a community with or without having disease A}

- Event:

E = {Head}
E = {Life span of a human is less than 3 years}
E = {The individuals who live in a community that has disease A}

5

Sample Space (Two Dice)

Sample Space: Compound event

S = {(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),
(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),
(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),
(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),
(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),
(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)}

Simple event

How many outcomes are "**Sum is 7**"?

6

Chapter 1 Probability

1.2 Properties of Probability

7

Notations

$A = \varnothing \Rightarrow$ A is an empty set

$A \subset B \Rightarrow$ A is a subset of B

$A \supset B \Rightarrow$ B is a subset of A

$A \cup B \Rightarrow$ Union of A & B

$A \cap B \Rightarrow$ Intersection of A & B

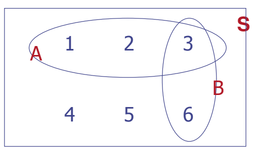
$A' \Rightarrow$ Complement of A

8

S = {1, 2, 3, 4, 5, 6}

A = {1, 2, 3}

B = {3, 6}



Venn Diagram

(with elements listed)

Intersection of events:

$A \cap B$ <=> A and B

Example: $A \cap B = \{3\}$

Union of events:

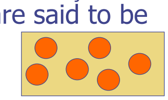
$A \cup B$ <=> A or B

Example: $A \cup B = \{1, 2, 3, 6\}$

9

Mutually Exclusive and Exhaustive

If $A_1, A_2, \ldots, A_k$ are events and $A_i \cap A_j = \varnothing$ for all $i \neq j$ then $A_1, A_2, \ldots, A_k$ are said to be **mutually exclusive** events.



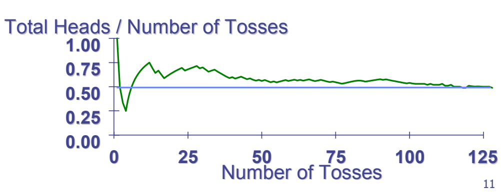
If $A_1, A_2, \ldots, A_k$ are events in the sample space S, then $A_1, A_2, \ldots, A_k$ **exhaustive events** if $A_1 \cup A_2 \cup \ldots \cup A_k = S$.

10

Definition of Probability

A rough definition: (frequentist definition)

Probability of a certain outcome to occur in a *random experiment* is the **proportion of times** that the this outcome would occur in a very long series of repetitions of the random experiment.



11

Determining Probability

How to determine probability?

- Empirical Probability
- Theoretical Probability
- (Subjective approach)

12

Empirical Probability Assignment

Empirical study: (Don't know if it is a balanced Coin?)

Outcome	Frequency
Head	512
Tail	488
Total	1000

13

Empirical Probability Assignment

Empirical probability assignment:

$$P(A) = \frac{\text{Number of times event } E \text{ occurs}}{\text{Number of times experiment is repeated}}$$

$$= \frac{n}{N(A)}$$

Probability of Head:

P(Head) = $\frac{512}{1000}$ **= .512 = 51.2%**

14

Empirical Probability Distribution

Empirical study:

Outcome	Frequency	Probability
Head	512	.512
Tail	488	.488
Total	1000	1.0

Empirical **Probability Distribution**

15

Theoretical Probability Assignment

Make a reasonable assumption:

What is the probability distribution in tossing a coin?

Assumption:

We have a balanced coin!

16

Theoretical Probability Assignment

Theoretical probability assignment:

$$P(A) = \frac{\text{Number of equally likely outcomes in event } A}{\text{Size of the sample space}}$$

$$= \frac{N(A)}{N(S)}$$

Probability of Head:

P(Head) = $\frac{1}{2}$ **= .5 = 50%**

17

Theoretical Probability Distribution

Empirical study:

Outcome	Probability
Head	.50
Tail	.50
Total	1.0

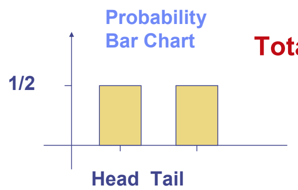
Empirical **Probability Distribution**

18

Probability Distribution (Probability Model)

If a balanced coin is tossed, *Head and Tail are equally likely to occur,*

P(Head) = .5 = 1/2   and   P(Tail) = .5 = 1/2



Total probability is 1.

19

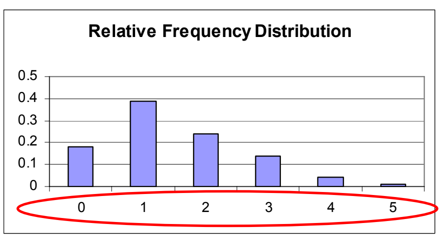
Relative Frequency and Probability Distributions

Number of times visited a doctor from a random sample of 300 individuals from a community

Class	Frequency	Relative Frequency	
0	54	.18	P(0) = .18
1	117	.39	P(1) = .39
2	72	.24	P(2) = .24
3	42	.14	P(3) = .14
4	12	.04	P(4) = .04
5	3	.01	P(5) = .01
Total	300	1.00	

20

Discrete Distribution



21

Relative Frequency and Probability

When selecting one individual at random from a population, the probability distribution and the relative frequency distribution are the same.

22

Probability for the Discrete Case

If an individual is randomly selected from this group 300, what is the probability that this person visited doctor 3 times?

P(3 times) = (42)/300
= .14 or 14%

Class	Frequency	Relative Frequency
0	54	.18
1	117	.39
2	72	.24
3	42	.14
4	12	.04
5	3	.01
Total	300	1.00

23

Discrete Distribution

If an individual is randomly selected from this group 300, what is the probability that this person visited doctor 4 or 5 times?

P(4 or 5 times) = P(4) + P(5)
= .04 + .01
= .05

Class	Frequency	Relative Frequency
0	54	.18
1	117	.39
2	72	.24
3	42	.14
4	12	.04
5	3	.01
Total	300	1.00

It would be an empirical probability distribution, if the sample of 300 individuals is utilized for understanding a large population.

24

Properties of Probability

Probability is a real-valued set function P that assigns to each event A in the sample space S a number P(A), called the probability of event A, such that the following properties are satisfied:

- $P(A) \geq 0$, probability is always a value between 0 and 1.
- $P(S) = 1$, total probability (all outcomes together) equals 1.
- If $A_1, A_2, \ldots, A_k$ are events and $A_i \cap A_j = \varnothing$ then $P(A_1 \cup A_2 \ldots \cup A_k) = P(A_1) + P(A_2) + \ldots + P(A_k)$

25

- $P(\varnothing) = 0$
- If $A \subset B$ then $P(A) \leq P(B)$
- For any event A, $P(A) \leq 1$

26

Probability Distribution (Probability Model)

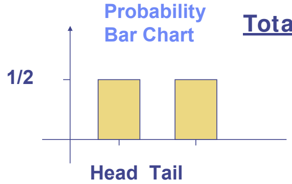
If a balanced coin is tossed, *Head* (H) *and Tail* (T) *are equally likely to occur,*

P(H) =.5 and **P(T) =.5**

P(all possible outcomes) = P(H) + P(T) = .5 + .5 = 1

Probability Bar Chart

Total probability is 1



27

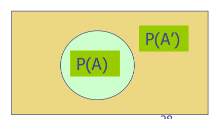
Complementation Rule

For any event A,

P(A does not occur) = 1 – P(A)

Complement of A = A′

* Some places use A^c or A′



28

Complementation Rule

If an unbalanced coin has a probability of 0.7 to turn up Head each time tossing this coin. What is the probability of not getting a Head for a random toss?

P(not getting Head) = 1 – 0.7
= 0.3

29

Complementation Rule

If the chance of a randomly selected individual living in community A to have disease H is .001, what is the probability that this person does not have disease H?

P(having disease H) = .001

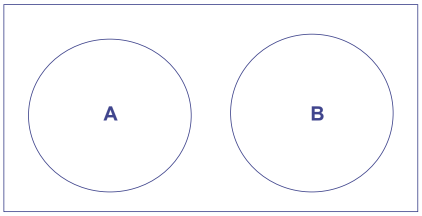
P(not having disease H)
= 1 – P(having disease H)
= 1 – 0.001
= **0.999**

30

Addition Rules for Probability

Mutually Exclusive Events

A and B are mutually exclusive (disjointed) events



Additive Rule: $P(A \cup B) = P(A) + P(B)$

31

Special Addition Rule

In an experiment of casting an unbalanced die, the event A and B are defined as the following:

A = an odds number

B = 6

and given P(A) = .61, P(B) = 0.1. What is the probability of getting an odds number or a six?

A and B are mutually exclusive.

$P(A \cup B) = P(A) + P(B) = .61 + .1 = .71$

32

Special Addition Rule

In A_1, A_2, ... , A_k are defined k mutually exclusive (m.e.) then

$$P(A_1 \cup A_2 \cup \ldots \cup A_k) = P(A_1) + P(A_2) + \ldots + P(A_k)$$

<u>Example</u>: When casting a balanced die, the probability of getting 1 or 2 or 3 is

$P(1 \cup 2 \cup 3) = P(1) + P(2) + P(3)$

$= 1/6 + 1/6 + 1/6 = 1/2$

33

<u>Example</u>: When casting a unbalanced die, such that P(1) = .2, P(2) = .1, P(3) =.3, the probability of getting 1 or 2 or 3 is

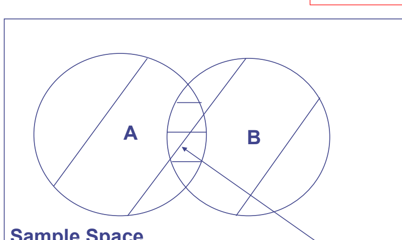
$P(1 \cup 2 \cup 3) = P(1) + P(2) + P(3)$

$= .2 + .1 + .3 = \mathbf{.6}$

34

General Addition Rule

Not M.E.!
$A \cap B \neq \emptyset$



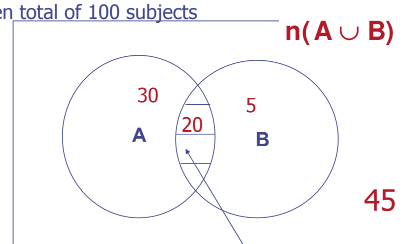
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

35

Venn Diagram (with counts)

Given total of 100 subjects

$n(A \cup B) = 55$



A=Smokers, n(A) = 50

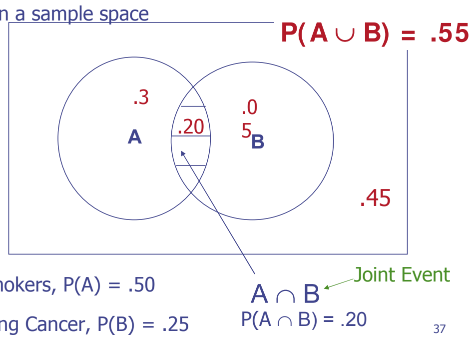
B=Lung Cancer, n(B) = 25

$A \cap B$ ← Joint Event

$n(A \cap B) = 20$

36

Venn Diagram (with relative frequencies)



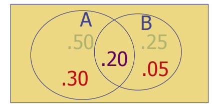
A=Smokers, P(A) = .50

B=Lung Cancer, P(B) = .25

37

Probability of Joint Events

In a study, an individual is randomly selected from a population, and the event A and B are defined as the followings:



A = the individual is a smoker.
B = the individual has lung cancer.

P(A) = .50, P(B) = .25
P(A and B) = P(A ∩ B)
= P(smoker and has lung cancer) = .20

What is the probability P(A or B) = P(A ∪ B) = ?

38

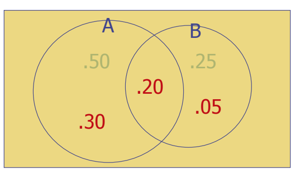
General Addition Rule

What is the probability that this randomly selected person is a smoker or has lung cancer? P(A or B) = ?

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

$= .50 + .25 - .20$

$= \mathbf{.55}$

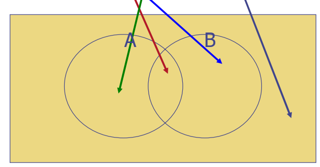


39

Contingency Table

	Cancer, B	No Cancer, B^c	Total
Smoke, A	20	30	50
Not Smoke, A^c	5	45	50
	25	75	100

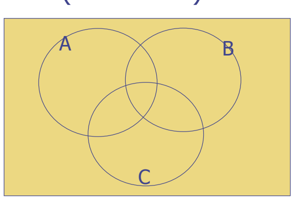
Venn Diagram



40

Extended General Addition Rule

$P(A \cup B \cup C) = P(A) + P(B) + P(C)$
$\quad - P(A \cap B) - P(B \cap C) - P(A \cap C)$
$\quad + P(A \cap B \cap C)$



41

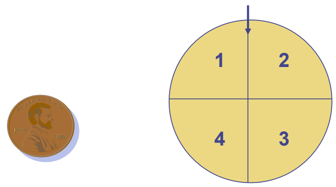
Chapter 1 Probability

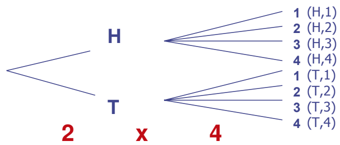
1.3 Methods of Enumeration

42

Sample Space = {(H,1),(H,2),(H,3),(H,4),(T,1),(T,2),(T,3),(T,4)}

Tree Diagram





2 x 4

43

Counting Rules

Multiplication Principle of Counting

In a sequence of n events in which the first one has k_1 possibilities and the second event has k_2 and the third has k_3, and so forth, the total possibilities of the sequence will be $k_1 \times k_2 \times k_3 \times \ldots \times k_n$

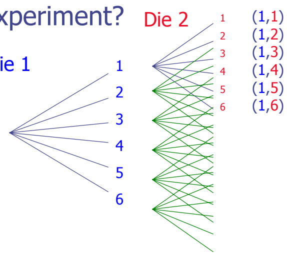
2 x 4 = 8 possible outcomes (coin & wheel)
6 X 6 = 36 possible outcomes (2 dice)

$k_1 \times k_2$

44

Tree Diagram

What is the sample space for casting two dice experiment?



6 x 6 = 36 outcomes

45

Example:

There are three shirts of different colors, two jackets of different styles and five pairs of pants in the closet. How many ways can you dress yourself with one shirt, one jacket and one pair of pants selected from the closet?

Sol: 3 x 2 x 5 = **30** ways

$k_1 \times k_2 \times k_3$

46

Permutations

B	A	S	E

How many different four-letter code words can be formed by using the four letters in the word "BASE" without repeating use of the same letter?

$k_1 = 4 \quad k_2 = 3 \quad k_3 = 2 \quad k_4 = 1$

$4 \cdot 3 \cdot 2 \cdot 1 =$ **24**

47

Factorial Formula

For any counting number n

n! = 1 x 2 x … x (n – 1) x n
0! = 1

Example: 3! = 1 x 2 x 3 = **6**

48

B A S E

How many different four-letter code words can be formed by using the four letters in the word "BASE" and the same letter can be used repeatedly?

$k_1 = 4 \quad k_2 = 4 \quad k_3 = 4 \quad k_4 = 4$

$4 \cdot 4 \cdot 4 \cdot 4 = \mathbf{256}$

49

Ordered Sampling B A S E

How many different four-letter code words can be formed by using the four letters selected from letters A through Z and the same letter can **not** be used repeatedly (sampling without replacement taking an ordered sample of size 4)?

$k_1 = 26 \quad k_2 = 25 \quad k_3 = 24 \quad k_4 = 23$

$26 \cdot 25 \cdot 24 \cdot 23 = \mathbf{358{,}800}$

50

Ordered Sampling B A S E

How many different four-letter code words can be formed by using the four letters selected from letters A through Z and the same letter can be used repeatedly (sampling with replacement taking an ordered sample of size 4)?

$k_1 = 26 \quad k_2 = 26 \quad k_3 = 26 \quad k_4 = 26$

$26 \cdot 26 \cdot 26 \cdot 26 = \mathbf{456{,}976}$

51

Permutation Rule

Permutation Rule:

The number of possible permutations of r objects from a collection of n distinct objects is

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

Order does count!

52

Permutations

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

How many ways can a four-digit code be formed by selecting 4 distinct digits from nine digits, 1 through 9, without repeating use of the same digit?

$${}_9P_4 = \frac{9!}{(9-4)!} = \frac{9!}{5!} = \frac{9 \cdot 8 \cdot 7 \cdot 6 \cdot \cancel{5!}}{\cancel{5!}}$$

$$= 9 \cdot 8 \cdot 7 \cdot 6 = \mathbf{3024}$$

53

Birthday Problem

In a group of randomly select 23 people, what is the probability that at least two people have the same birth date? (Assume there are 365 days in a year.)

P(at least two people have the same birth date)

Too hard !!!

= 1 – P(everybody has different birth date)

= 1 – [365x364x…x(365-23+1)] / 365^{23}

≈ .5

54

Combination Rule

Combination Rule:

The number of possible combinations of r objects from a collection of n distinct objects is

$$_nC_r = \frac{n!}{r!\cdot(n-r)!}$$

Order does not count!

Binomial Coefficient

55

Combinations

$$_nC_r = \frac{n!}{r!\cdot(n-r)!}$$

How many ways can a committee be formed by selecting 3 people from a group 10 candidates?

n = 10, r = 3

$$_{10}C_3 = \frac{10!}{3!\cdot(10-3)!} = \frac{10!}{3!\cdot 7!} = \frac{10\cdot 9\cdot 8\cdot \cancel{7!}}{3!\cdot \cancel{7!}}$$

$$= \frac{10\cdot 9\cdot 8}{3!} = \mathbf{120}$$

ABC, ACB, BAC, BCA, CAB, CBA are the same combination.

56

Combinations

$$_nC_r = \frac{n!}{r!\cdot(n-r)!}$$

How many ways can a combination of 4 distinct digits be selected from nine digits, 1 through 9?

$$_9C_4 = \frac{9!}{4!\cdot(9-4)!} = \frac{9!}{4!\cdot 5!} = \frac{9\cdot 8\cdot 7\cdot 6\cdot \cancel{5!}}{4!\cdot \cancel{5!}}$$

$$= \frac{9\cdot 8\cdot 7\cdot 6}{4\cdot 3\cdot 2\cdot 1}$$

$$= 126$$

57

Combinations

$$_nC_r = \frac{n!}{r!\cdot(n-r)!}$$

How many ways can 6 distinct numbers be selected from a set of 47 distinct numbers?

n = 47, r = 6

$$_{47}C_6 = \frac{47!}{6!\cdot 41!} = \frac{47\cdot 46\cdot 45\cdot 44\cdot 43\cdot 42\cdot 41!}{6!\cdot 41!}$$

$$= \frac{47\cdot 46\cdot 45\cdot 44\cdot 43\cdot 42}{6!}$$

$$= \mathbf{10{,}737{,}573}$$

58

Probability Using Counting (Theoretical Approach)

In a lottery game, one selects 6 distinct numbers from a set of 47 distinct numbers, what is the probability of winning the jackpot?

n = 47, r = 6

$_{47}C_6 = 47!/(6!\cdot 41!) = \mathbf{10{,}737{,}573}$

$$P(\text{Jackpot}) = \frac{1}{10{,}737{,}573}$$ (Theoretical Prob.)

59

Distinguishable Permutations

Let a set of n objects of two types, r of one type, and $n-r$ of other type. The number of distinguishable permutations of these n objects is $_nC_r$.

Example: How many possible outcomes could it be if a coin is tossed 8 times and getting 2 heads and 6 tails?

$$_nC_r = \binom{n}{r} = {}_8C_2 = \binom{8}{2} = \frac{8!}{2!\cdot 6!} = 28$$

How many ways can a World Series in baseball end?

60

Distinguishable Permutations

Let a set of n objects of s types, n_1 of one type, and n_2 of the 2nd type, … and, n_s the s-th type. The number of distinguishable permutations of these n objects is

$$\binom{n}{n_1, n_2, \ldots, n_s} = \frac{n!}{n_1!\, n_2! \ \ldots\ n_s!}$$

$\boxed{\text{Multinomial Coefficient}}$

61

Distinguishable Permutations

Example: For Christmas decoration, three different colors of light bulbs are used to line up is a row with 3 yellow, 5 green, and 6 red light bulbs. How many different ways can these light bulbs be arranged?

$$\binom{14}{3,5,6} = \frac{14\,!}{3!\cdot 5!\cdot 6!} = 168168$$

62

Chapter 1 Probability

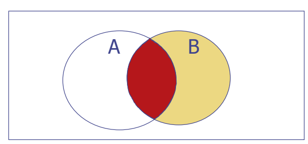
1.4 Conditional Probability

63

Conditional Probability

The conditional probability of event A to occur given event B has occurred (or given the condition B) is denoted as $P(A|B)$ and is, if $P(B)$ is not zero, $n(E)$ = # of equally likely outcomes in E,

$$P(A\,|\,B) = \frac{P(A \cap B)}{P(B)} \quad \text{or} \quad P(A\,|\,B) = \frac{n(A \cap B)}{n(B)}$$



64

Conditional Probability $P(A\,|\,B) = \frac{n(A \cap B)}{n(B)}$

	Cancer C	No Cancer C′	Total
Smoke S	20	30	50
Not Smoke S′	5	45	50
	25	75	100

P(C|S) = 20/50 = .4

P(C|S′) = 5/50 = .1

65

Conditional Probability $P(A\,|\,B) = \frac{P(A \cap B)}{P(B)}$

	Cancer C	No Cancer C′	Total
Smoke S	20 (.2)	30 (.3)	50 P(S) =(.5)
Not Smoke S′	5 (.05)	45 (.45)	50 P(S′) =(.5)
	25 P(C) =(.25)	75 P(C)=(.75)	100 (1.0)

P(C|S) = .2/.5 = .4

P(C|S′) = .05/.5 = .1

What is $\frac{P(C|S)}{P(C|S')} = 4$

(*Relative Risk*)

66

General Multiplication Rule

For any two events A and B,

$P(A \cap B) = P(A|B)\,P(B) = P(B|A)\,P(A)$

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

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

67

General Multiplication Rule

If in the population, **50% of the people smoked**, and **40% of the smokers have lung cancer**, what percentage of the population that are smoker and have lung cancer?

P(S) = 50% of the subjects smoked
P(C|S) = 40% of the smokers have cancer

$P(C \cap S) = P(C|S)\,P(S) = .4 \times .5 = .2$

68

General Multiplication Rule

Box 1 contains 2 red balls and 1 blue ball

Box 2 contains 1 red ball and 3 blue balls

A coin is tossed. If "Head" turns up a ball is drawn from Box 1, and if "Tail" turns up then a ball is drawn from Box 2. Find the probability of selecting a red ball.

$P(H \cap Red) + P(T \cap Red)$
$= P(H)P(Red|H) + P(T)P(Red|T)$
$= (1/2) \times (2/3) + (1/2) \times (1/4)$
$= 1/3 + 1/8 = \mathbf{11/24}$

69

Chapter 1 Probability

1.5 Independent Events

70

Independent Events

Events A and B are independent if

$P(A|B) = P(A)$

or

$P(B|A) = P(B)$

or

$P(A \cap B) = P(A) \cdot P(B)$

71

Special Multiplication Rule

If event A and B are independent,

$P(A \cap B) = P(A) \cdot P(B)$

72

Example

If a balanced die is rolled twice, what is the probability of having two 6's?

6_1 = the event of getting a 6 on the 1st trial

6_2 = the event of getting a 6 on the 2nd trial

$P(6_1) = 1/6, \quad P(6_2) = P(6_2|6_1) = 1/6,$

6_1 and 6_2 are independent events

$P(6_1 \cap 6_2) = P(6_1)\,P(6_2) = (1/6)(1/6) = 1/36$

73

Independent Events

"10%" of the people in a large population has disease H. If a random sample of two subjects was selected from this population, what is the probability that both subjects have disease H?

H_i : Event that the i-th randomly selected subject has disease H.

$P(H_2|H_1) = P(H_2)$ [Events are almost independent]

$P(H_1 \cap H_2) = P(H_1)\,P(H_2) = .1 \times .1 = .01$

74

Independent Events

If events $A_1, A_2, \ldots, A_k$ are independent, then

$P(A_1 \cap A_2 \cap \ldots \cap A_k)$
$= P(A_1) \cdot P(A_2) \cdot \ldots \cdot P(A_k)$

What is the probability of getting all heads in tossing a balanced coin four times experiment?

$P(H_1) \cdot P(H_2) \cdot P(H_3) \cdot P(H_4) = \mathbf{(.5)^4 = .0625}$

75

Independent Events

"10%" of the people in a large population has disease H. If a random sample of **3** subjects was selected from this population, what is the probability that all subjects have disease H?

H_i : Event that the i-th randomly selected subject has disease H.

[Events are almost independent]

$P(H_1 \cap H_2 \cap H_3) = P(H_1)\,P(H_2)\,P(H_3)$
$= .1 \times .1 \times .1 = \mathbf{.001}$

76

Binomial Probability

What is the probability of getting two 6's in casting a balanced die 5 times experiment?

$P(S \cap S \cap S' \cap S' \cap S') = (1/6)^2 \times (5/6)^3 = \mathbf{0.016}$

$P(S \cap S' \cap S \cap S' \cap S') = (1/6)^2 \times (5/6)^3$

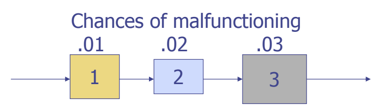
$P(S \cap S' \cap S' \cap S \cap S') = (1/6)^2 \times (5/6)^3$

…

How many of them? $\binom{5}{2} = \frac{5!}{2!\cdot 3!} = 10$

Probability (two 6's) = 0.016 x 10 = **0.16**

77



A medical devise consists of 3 components. The devise will not be functioning if any of these 3 components failed. The performance of these 3 components are independent to each other. When the devise is turned on, **the chance of component 1 would be malfunctioned is .01 , and the chance of the component 2 would be malfunctioned is .02 , and the chance of the component 3 would be malfunctioned is .03.** What is the probability that this medical devise will be functioning next time the system is turned on?

$(1-.01)(1-.02)(1-.03) = .99 \times .98 \times .97 = .941$

78

Chapter 1 Probability

1.6 Bayes's Theorem

(Diagnostic Tests)

(Screening Tests)

Bayes's Theorem

- Box 1 contains 2 red balls and 1 blue ball
- Box 2 contains 1 red ball and 3 blue balls

A coin is tossed. If "Head" turns up a ball is drawn from Box 1, and if "Tail" turns up then a ball is drawn from Box 2. If a red ball is drawn, what is probability that it is from Box 1?

P(H | Red) = ?

Bayes's Theorem

- Box 1 contains 2 red balls and 1 blue ball
- Box 2 contains 1 red ball and 3 blue balls

1/2 P(H) — 2/3 P(Red | H) $\xrightarrow{1/3}$ P(H ∩ Red)

P(H ∩ Red) = P(H)xP(Red|H)

1/2 P(H) — 1/3 P(Blue | H) $\xrightarrow{1/6}$ P(H ∩ Blue)

1/2 P(T) — 1/4 P(Red | T) $\xrightarrow{1/8}$ P(T ∩ Red)

1/2 P(T) — 3/4 P(Blue | T) $\xrightarrow{3/8}$ P(T ∩ Blue)

Bayes's Theorem

1/2 P(H) — 2/3 P(Red | H) $\xrightarrow{1/3}$ P(H ∩ Red)

P(H ∩ Red) = P(H)xP(Red|H)

1/2 P(H) — 1/3 P(Blue | H) $\xrightarrow{1/6}$ P(H ∩ Blue)

1/2 P(T) — 1/4 P(Red | T) $\xrightarrow{1/8}$ P(T ∩ Red)

1/2 P(T) — 3/4 P(Blue | T) $\xrightarrow{3/8}$ P(T ∩ Blue)

$$P(H|R) = \frac{P(H \cap R)}{P(R)} = \frac{P(H \cap R)}{P(H \cap R) + P(T \cap R)}$$

$$= [1/3] / ([1/3] + [1/8]) = \mathbf{8/11}$$

Prior and Posterior Probabilities

P(H | Red) is called the Posterior Probability of H.

P(H) is called the Prior probability of H.

Bayes's Theorem $P(A \mid D) = \dfrac{P(A \cap D)}{P(D)}$

- Plant A(80%) - has 2% of Defective products.
- Plant B(20%) - has 1% of Defective products.

If a defective product is found what is the probability that it was from Plant A?

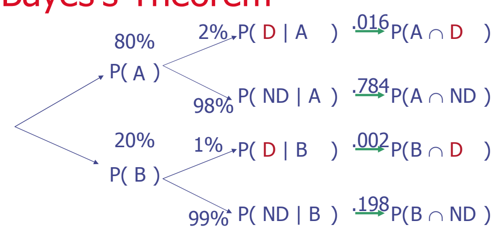
80% P(A) — 2% P(D | A) $\xrightarrow{.016}$ P(A ∩ D)

80% P(A) — 98% P(ND | A) $\xrightarrow{.784}$ P(A ∩ ND)

20% P(B) — 1% P(D | B) $\xrightarrow{.002}$ P(B ∩ D)

20% P(B) — 99% P(ND | B) $\xrightarrow{.198}$ P(B ∩ ND)

Bayes's Theorem

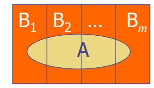


$$P(A \mid D) = \frac{P(A \cap D)}{P(D)} = \frac{P(A \cap D)}{P(A \cap D) + P(B \cap D)}$$

$= .016 / (.016 + .002) = \mathbf{.889}$

85

Bayes's Theorem



Let $B_1, B_2, \ldots, B_m$ be mutually exclusive and exhaustive events that constitute a partition of sample space S. The prior probabilities of events B_i are positive. If A is the union of m mutually exclusive events that, $A = \bigcup_{i=1}^{m}(B_i \cap A)$ and that

$$P(A) = \sum_{i=1}^{m} P(B_i \cap A) = \sum_{i=1}^{m} P(B_i)P(A \mid B_i)$$

then

$$P(B_k \mid A) = \frac{P(B_k)P(A \mid B_k)}{\sum_{i=1}^{m} P(B_i)P(A \mid B_i)}, \qquad k = 1,2,\ldots m$$

86

Some Questions

- If you test positive for HIV, what is the probability that you have HIV?
- If you used an inspection system and detected a defective computer chip what is the probability that this chip is really defective?

87

Example: Coronary artery disease

	Present D^+	Absent D^-	Total
Positive T^+	815	115	930
Negative T^-	208	327	535
Total	1023	442	1465

[From the book, *Medical Statistics* page 30, and the 2x2 table from data of Weiner et al (1979)]

88

	Present D^+	Absent D^-	Total
Positive T^+	815	115	930
Negative T^-	208	327	535
Total	1023	442	1465

The disease **Prevalence** in these patients

$P(D^+) = 1023/1465 = .70$

Predictive Value of a Positive Test: The probability of **having the disease** given that a person has a **positive test** is given by:

$$P(D^+ \mid T^+) = \frac{P(D^+ \cap T^+)}{P(T^+)} = \frac{815/1465}{930/1465} = \frac{815}{930} \approx .88$$

89

	Present D^+	Absent D^-	Total
Positive T^+	815	115	930
Negative T^-	208	327	535
Total	1023	442	1465

Predictive Value of a Negative Test:______

$$P(D^- \mid T^-) = \frac{P(D^- \cap T^-)}{P(T^-)} =$$

90

	Present D^+	Absent D^-	Total
Positive T^+	815	115	930
Negative T^-	208	327	535
Total	1023	442	1465

Things to notice: (Prevalence)

$$P(D^+) = P(D^+ \cap T^+) + P(D^+ \cap T^-)$$
$$= 815/1465 + 208/1465 = 1023/1465 \approx .7$$

Likewise, (Overall percentage that tested positive)

$$P(T^+) = P(T^+ \cap D^+) + P(T^+ \cap D^-)$$
$$= 815/1465 + 115/1465 = 930/1465 \approx .63$$

91

	Present D^+	Absent D^-	Total
Positive T^+	815	115	930
Negative T^-	208	327	535
Total	1023	442	1465

Calculate the following:

$P(D^-)$ = _______

$P(T^-)$ = _______

$P(D^+ \cap T^+) - P(D^+)P(T^+)$ = _________

92

Sensitivity and Specificity

93

	Present D^+	Absent D^-	Total
Positive T^+	815	115	930
Negative T^-	208	327	535
Total	1023	442	1465

Sensitivity: The probability of a person testing positive, given that (s)he has the disease.

$$P(T^+ | D^+) = \frac{P(T^+ \cap D^+)}{P(D^+)} = \frac{815/1465}{1023/1465} = 815/1023 \approx .80$$

94

	Present D^+	Absent D^-	Total
Positive T^+	815	115	930
Negative T^-	208	327	535
Total	1023	442	1465

Specificity: The probability of a person testing negative given that (s)he does not have the disease.

$$P(T^- | D^-) = \frac{P(T^- \cap D^-)}{P(D^-)} = \frac{327/1465}{442/1465} = 327/442 \approx .74$$

95

	Present D^+	Absent D^-	Total
Positive T^+	815	115	930
Negative T^-	208	327	535
Total	1023	442	1465

False Negative rate = 1 – sensitivity

$$P(T^- | D^+) = \frac{P(T^- \cap D^+)}{P(D^+)} = \frac{208/1465}{1023/1465} = 208/1023 \approx .2$$

False Positive rate = 1 – specificity

$$P(T^+ | D^-) = \frac{P(T^+ \cap D^-)}{P(D^-)} = \frac{115/1465}{442/1465} = 115/442 \approx .26$$

96

Sensitivity & False Negatives

For our problem, since the sensitivity is .8, the false negative rate is 1 – .8 = .2.

Interpretation: 20% of the time a person will actually have the disease when the test says that he/she does not.

Likewise, since specificity is .74, the false positive rate is 0.26.

97

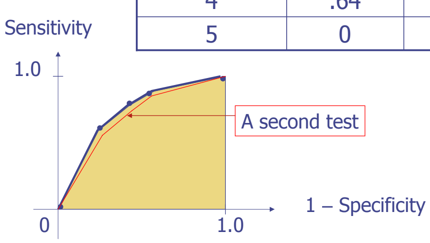
Sensitivity vs Specificity

- In a perfect world, we want both to be high.
- The two components have a see-saw relationship.
- Comparisons of tests' accuracy are done with *Receiver Operator Characteristic* (ROC) curves.

98

ROC

Test Positive Criteria	Sensitivity	Specificity
1	1.0	0
2	.93	.55
3	.81	.59
4	.64	.80
5	0	1.0



99

About ROC Curve

- Area under the ROC represents the probability of correctly distinguishing a normal from an abnormal subject on the relative ordering of the test ratings.
- Two screening tests for the same disease, the test with the higher area under its ROC curve is considered the better test, unless there is particular level of sensitivity or specificity is especially important in comparing the two tests.

100

Bayes's Theorem $P(B|A) = \dfrac{P(A|B)P(B)}{P(A)}$

Applying Bayes's formula, we can find the predictive value of a positive test using sensitivity of a test

$$P(D^+|T^+) = \frac{\overbrace{P(T^+|D^+)}^{\text{sensitivity}}\overbrace{P(D^+)}^{\text{prevalence}}}{\underbrace{P(T^+)}_{\text{Tested positive}}} = \frac{(.8)(.7)}{.63} = .89$$

In words, the predictive value of a positive test is equal to the sensitivity (=.8) times prevalence (=.7) divided by percentage who test positive (=.63).

101

Example

Suppose that 84% of hypertensives and 23% of normotensives are classified as hypertensive by an automated blood-pressure machine. If it is estimated that the prevalence of hypertensives is 20%, what is the predictive value of a positive test? What is the predictive value of a negative test?

Sensitivity = .84 Specificity = 1–.23 = .77

Pretty Good Test!??

$$PV^+ = \frac{(.84)(.2)}{P(T^+)} = \frac{(.84)(.2)}{P(T^+ \cap D^+) + P(T^+ \cap D^-)} = \frac{(.84)(.2)}{(.2)(.84) + (.8)(.23)} = .48$$

$PV^- = .95$ **What if only 2% prevalence?**

102