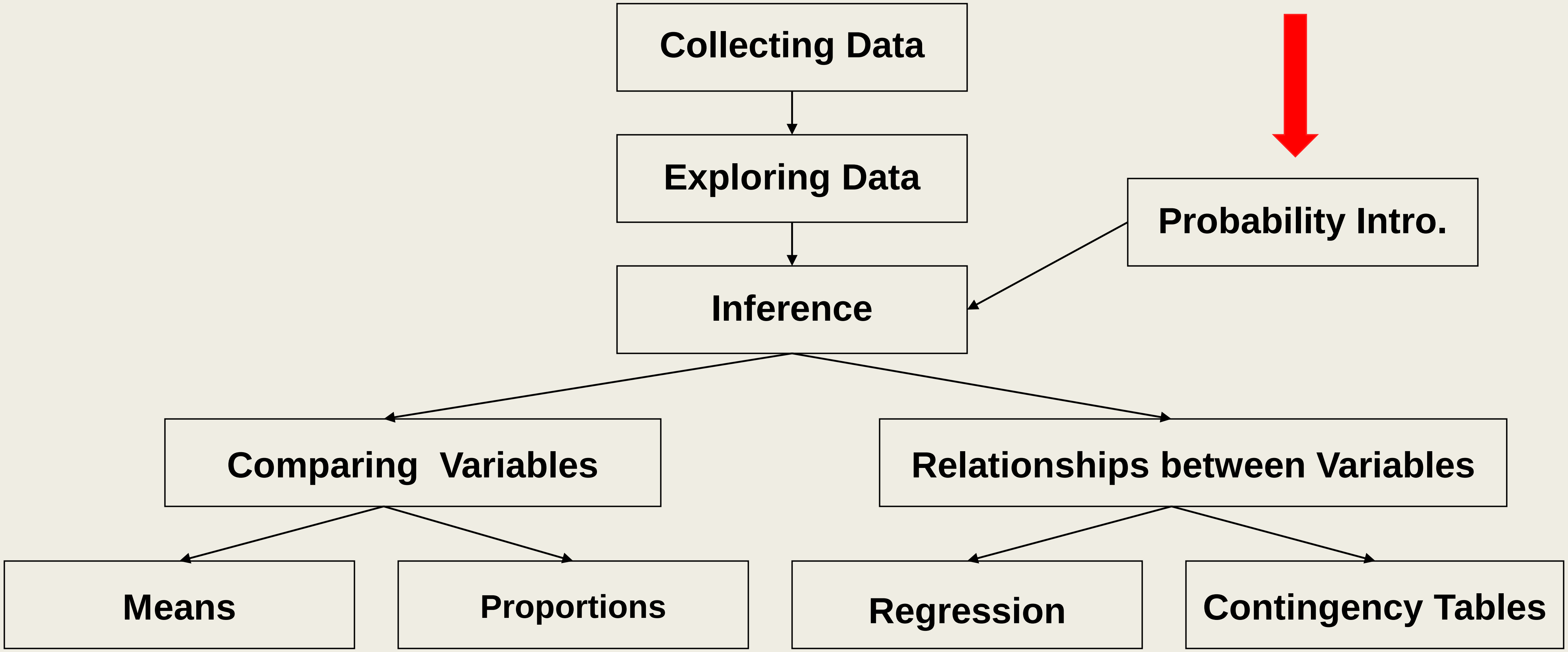
Probability

Assistant Professor Dr. Pornpimol Chaiwuttisak

Course Overview



Basic definitions: Probability

- **Random Experiment** is a process of observation that leads to a single outcome that cannot be predicted with certainty

  Examples:

  1. Pick a integer number X between 1 and 20

  2. Toss a coin

- **Sample Space:** All outcomes of an experiment. Usually denoted by S or Ω.

  Examples:

  S = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20}

  S = {H,T}

  The probability of the whole sample space is 1 : **P(S) = 1**

Basic definitions: Probability

- **Event** denoted by E is any subset of S

  1. E = Chosen odd number = {1,3,5,7,9,11,13,15,17,19}

  2. E = Getting exactly Head = {H}

- **P(E)** denotes the probability of the event E

  1. P(E) = P(Odd Number)

  2. P(E) = P(Head)

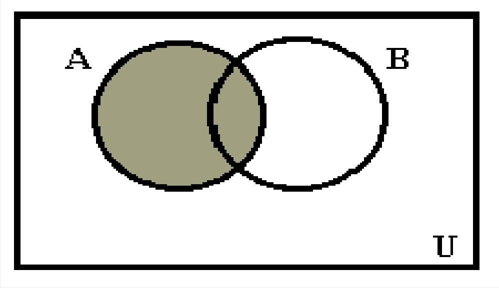
  Notation: P(E) = Probability of event E

  Probability Rule : **$0 \leq P(E) \leq 1$ for any event E**

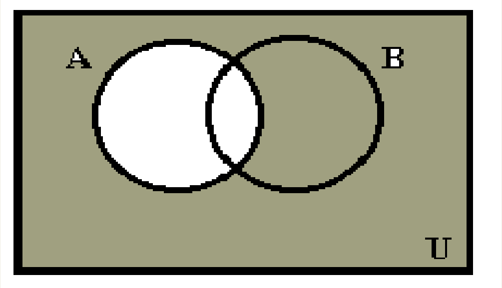
Combinations of Events

- The **complement** A^c of an event A is the event that A does not occur
- **Probability Rule : P(Ac) = 1 - P(A)**
- The **union** of two events A and B is the event that either A or B or both occurs
- The **intersection** of two events A and B is the event that both A and B occur

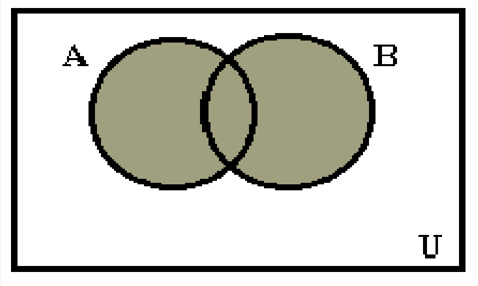
Event A



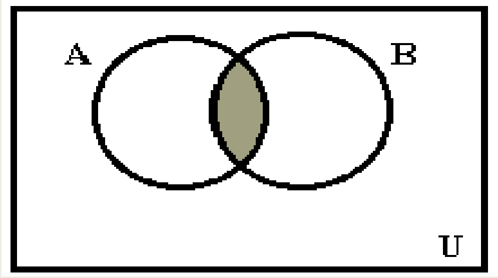
Complement of A



Union of A and B



Intersection of A and B



Disjoint Events

- Two events are called **disjoint** if they can not happen at the same time
  - *Events A and B are disjoint means that the intersection of A and B is zero*
- Example: coin is tossed twice
  - *S = {HH,TH,HT,TT}*
  - *Events A={HH} and B={TT} are disjoint*
  - *Events A={HH,HT} and B = {HH} are not disjoint*
- Probability Rule : **If A and B are disjoint events then**

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

Independent events

- Events A and B are **independent** if knowing that A occurs does not affect the probability that B occurs
- Example: tossing two coins

  Event A = first coin is a head

  Event B = second coin is a head

  Independent

- Disjoint events cannot be independent!
  - *If A and B can not occur together (disjoint), then knowing that A occurs does change probability that B occurs*
- Property: P(A) = P(A|B) and P(B) = P(B) = P(B|A)

Independent events

- Probability Rule : **If A and B are independent**

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

multiplication rule for independent events

Proof

1. $P(A) = P(A|B)$
2. $P(B) = P(B|A)$

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

$$\begin{aligned} P(A \cap B) &= P(A|B) \times P(B) \\ &= P(A) \times P(B) \end{aligned}$$

and

$$\begin{aligned} P(B|A) &= P(A \cap B)/P(A) \\ P(A \cap B) &= P(B|A) \times P(A) \\ &= P(B) \times P(A) \end{aligned}$$

Conditional Probability

- **The notation** **P(F | E)** is read "the probability of event F given event E". It is the probability of an event F given the occurrence of the event E.
- **Definition** Let F and E be events in the sample space S, with P(E)≠ 0. The *conditional probability* of F given E is

$$P(\mathrm{F} \mid \mathrm{E}) = \frac{P(E \cap F)}{P(\mathrm{E})}$$

- The order is very important do not think that P(A|B)=P(B|A)! THEY ARE DIFFERENT.

Exercise

Suppose that A and B are events with probabilities:

$$P(A)=1/3,\ P(B)=1/4,\ P(A \cap B)=1/10$$

Find each of the following:

1. $P(A \mid B) = ?$

Exercise

Suppose that A and B are events with probabilities:

$$P(A)=1/3,\ P(B)=1/4,\ P(A \cap B)=1/10$$

Find each of the following:

1. $P(A \mid B) = P(A \cap B)/P(B)=1/10/1/4=4/10$

Exercise

Suppose that A and B are events with probabilities:

$$P(A)=1/3,\ P(B)=1/4,\ P(A \cap B)=1/10$$

Find each of the following:

1. $P(A \mid B) = P(A \cap B)/P(B)=1/10/1/4=4/10$
2. $P(B \mid A) = ?$

Exercise

Suppose that A and B are events with probabilities:

$$P(A)=1/3,\ P(B)=1/4,\ P(A \cap B)=1/10$$

Find each of the following:

1. $P(A \mid B) = P(A \cap B)/P(B)=1/10/1/4=4/10$
2. $P(B \mid A) = P(A \cap B)/P(A)=1/10/1/3=3/10$

Exercise

Suppose that A and B are events with probabilities:

$$P(A)=1/3,\ P(B)=1/4,\ P(A \cap B)=1/10$$

Find each of the following:

1. $P(A \mid B) = P(A \cap B)/P(B)=1/10/1/4=4/10$
2. $P(B \mid A) = P(A \cap B)/P(A)=1/10/1/3=3/10$
3. $P(A' \mid B') = ?$

Exercise

Suppose that A and B are events with probabilities:

$$P(A)=1/3,\ P(B)=1/4,\ P(A \cap B)=1/10$$

Find each of the following:

1. $P(A \mid B) = P(A \cap B)/P(B)=1/10/1/4=4/10$
2. $P(B \mid A) = P(A \cap B)/P(A)=1/10/1/3=3/10$
3. $P(A'| B')= P(A' \cap B')/P(B')$

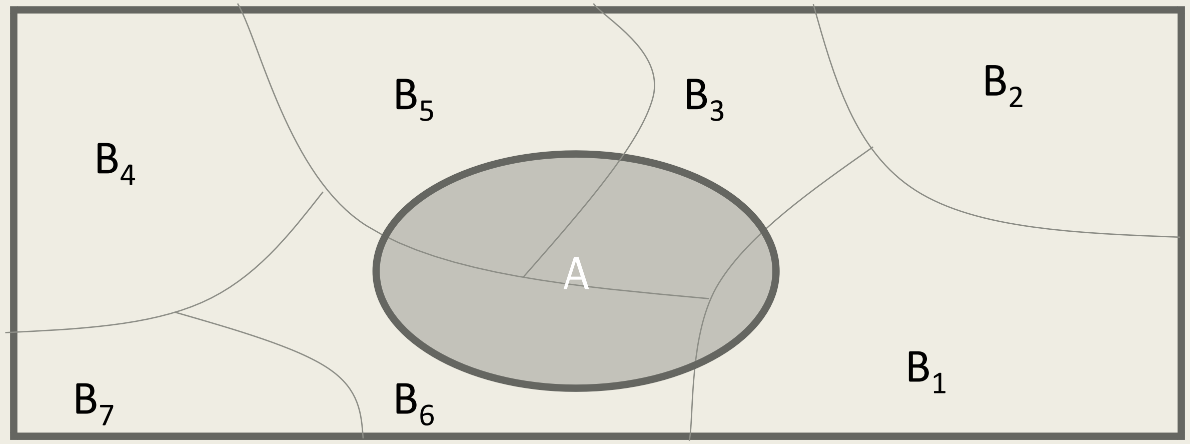
$$= P((A \cup B)')/(1-P(B))$$
$$= (1-P(A \cup B))/(1 - P(B))$$
$$= (1 - (P(A)+P(B)-P(A \cap B)))/(1-P(B))$$
$$= (1 - (1/3+1/4-1/10))/(1-1/10)$$
$$= (1-29/60)/9/10$$
$$= 31/60/9/10=31/54.$$

Problem

In a recent study it was found that the probability that a randomly selected student is a girl is .51 and is a girl and plays sports is .10. If the student is female, what is the probability that she plays sports?

$$P(S \mid F) = \frac{P(S \cap F)}{P(F)} = \frac{.1}{.51} = .1961$$

Rule of total probability



$$p(A) = \sum P(B_i)P(A \mid B_i)$$

Bayes' Rule

Let B_1 , B_2 , B_3 ,...., B_k be mutually exclusive and exhaustive events with prior probabilities $P(B_1)$, $P(B_2)$,…,$P(S_k)$. If an event A occurs, the posterior probability of B_i, given that A occurred is

$$P(B_i \mid A) = \frac{P(B_i)P(A \mid B_i)}{\sum P(B_i)P(A \mid B_i)} \quad \text{for } i = 1, 2, \ldots k$$

Proof

$$P(A \mid B_i) = \frac{P(A \cap B_i)}{P(B_i)} \longrightarrow P(A \cap B_i) = P(B_i)P(A \mid B_i)$$

$$P(B_i \mid A) = \frac{P(A \cap B_i)}{P(A)} = \frac{P(B_i)P(A \mid B_i)}{\sum P(B_i)P(A \mid B_i)}$$

Bayes' Rule

probability of seeing the evidence
if the hypothesis is true

probability a hypothesis is true
(before any evidence is present)

Likelihood

Prior Probability

Posterior

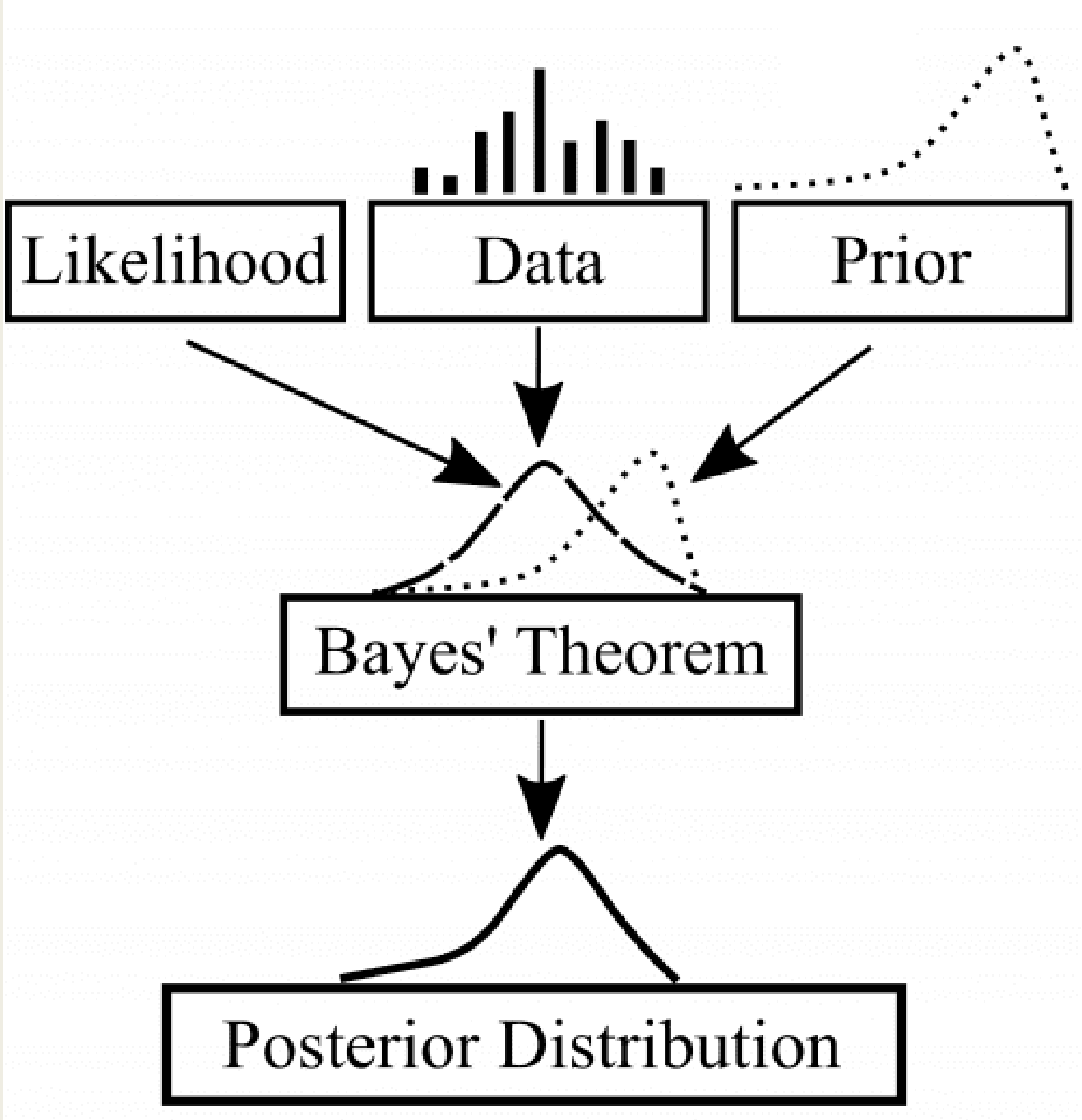
$$P(\theta|D) = \frac{P(D|\theta) \times P(\theta)}{P(D)} \propto P(D|\theta) \times P(\theta)$$

$$P(D) \rightarrow \int P(D|\theta) \times P(\theta)\, d\theta$$

Evidence

probability of observing the evidence

- How good are our parameters given the data
- **Prior** knowledge is incorporated and used to **update** our beliefs about the parameters


Likelihood
Data
Prior
Bayes' Theorem
Posterior Distribution

Principles of Bayesian Inference

- Formulation of a generative model

Model → Likelihood function $P(D|\theta)$

Model → Prior distribution $P(\theta)$

ความเป็นไปได้ที่โมเดลเราจะมีค่าพารามิเตอร์ Weight 1

โอกาสที่ค่าพารามิเตอร์เท่ากับ Weight 1

- Observation of data

Measurement → data *D*

- Model inversion – updating one's belief

$P(\theta|D) \propto P(D|\theta) \times P(\theta)$ → Posterior distribution

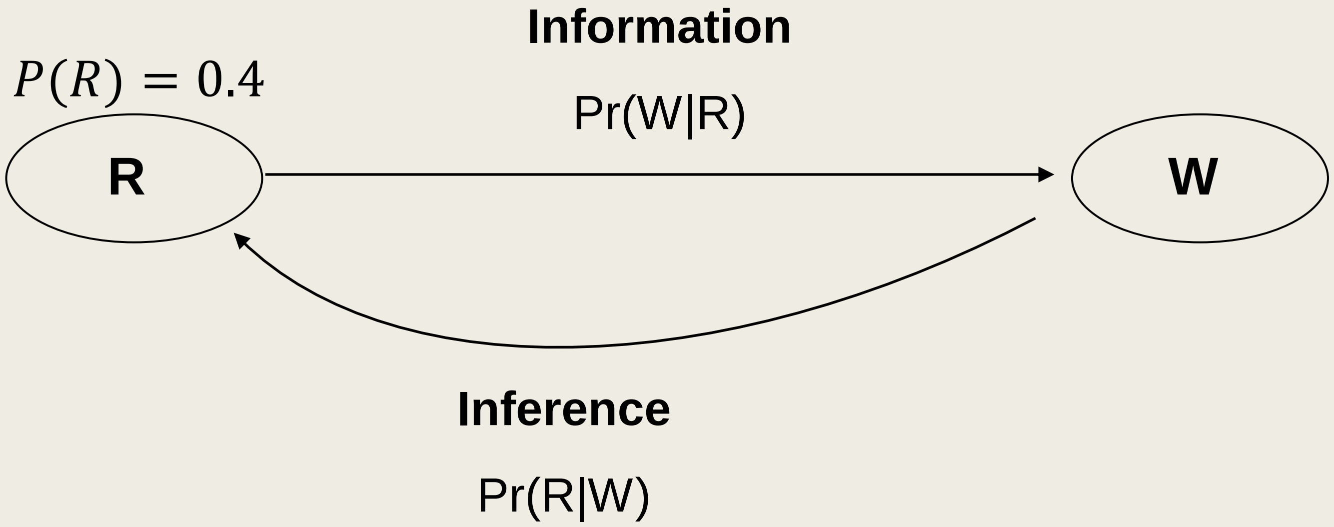
$P(\theta|D) \propto P(D|\theta) \times P(\theta)$ → Model evidence

Bayes' Rule

	R	¬R
W	0.9	0.2
¬W	0.1	0.8

R: It rains

W: The grass is wet



$$P(R|W) = \frac{P(W|R) \times P(R)}{P(W)} = \frac{0.9 \times 0.4}{(0.9\ x\ 0.4)+(0.2\ x\ 0.6)} = \frac{0.36}{0.36+0.12} = 0.75$$

Example: Coin Flipping Model

- Someone flips a coin
- We don't know if the coin is fair or not
- We are told only the outcome of the coin flipping

Example: Coin Flipping Model

- 1st Hypothesis: Coin is fair, 50% Heads or Tails

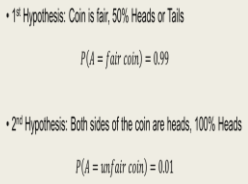
 

- 2nd Hypothesis: Both sides of the coin are heads, 100% Heads

Example: Coin Flipping Model

- 1st Hypothesis: Coin is fair, 50% Heads or Tails



$$P(A = fair\ coin) = 0.99$$

- 2nd Hypothesis: Both sides of the coin are heads, 100% Heads

$$P(A = unfair\ coin) = 0.01$$

Example: Coin Flipping Model

- 1^{st} Flip

$$P(A = fair|B = Heads) = \frac{P(B=Heads|A=fair) \times P(A=fair)}{P(B=Heads)}$$

- $P(A = fair) = 0.99$
- $P(B = Heads|A = fair) = 0.5$
- $P(B = Heads) = P(B = Heads, A = fair) + P(B = Heads, A = unfair)$

$$= P(B|A)P(A) + P\left(B|\overline{A}\right)P\left(\overline{A}\right)$$

$$= 0.5 \times 0.99 \quad + \quad 1 \times 0.01 \quad = \quad 0.5050$$

Example: Coin Flipping Model

- 1^{st} Flip

$$P(A = fair) = 0.99$$

$$P(B = Heads|A = fair) = 0.5$$

$$P(B = Heads) = 0.5050$$

$$P(A = fair|B = Heads) = \frac{P(B = Heads|A = fair) \times P(fair)}{P(B = Heads)} = \frac{0.5 \times 0.99}{0.5050} = 0.9802$$

Example: Coin Flipping Model


Coin is flipped a second time and it is heads again

- Posterior in the previous time step becomes the new prior!!

Example: Coin Flipping Model

- $P(A = fair) = 0.9802$
- $P(B = H|A = fair) = 0.5$
- $P(B = H) = P(B = H, A = fair) + P\left(B = H, \overline{A} = unfair\right)$

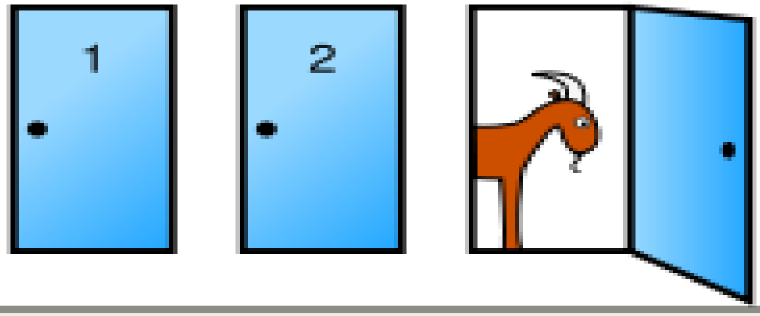
$$= P(B|A)P(A) + P\left(B|\overline{A}\right)P\left(\overline{A}\right)$$

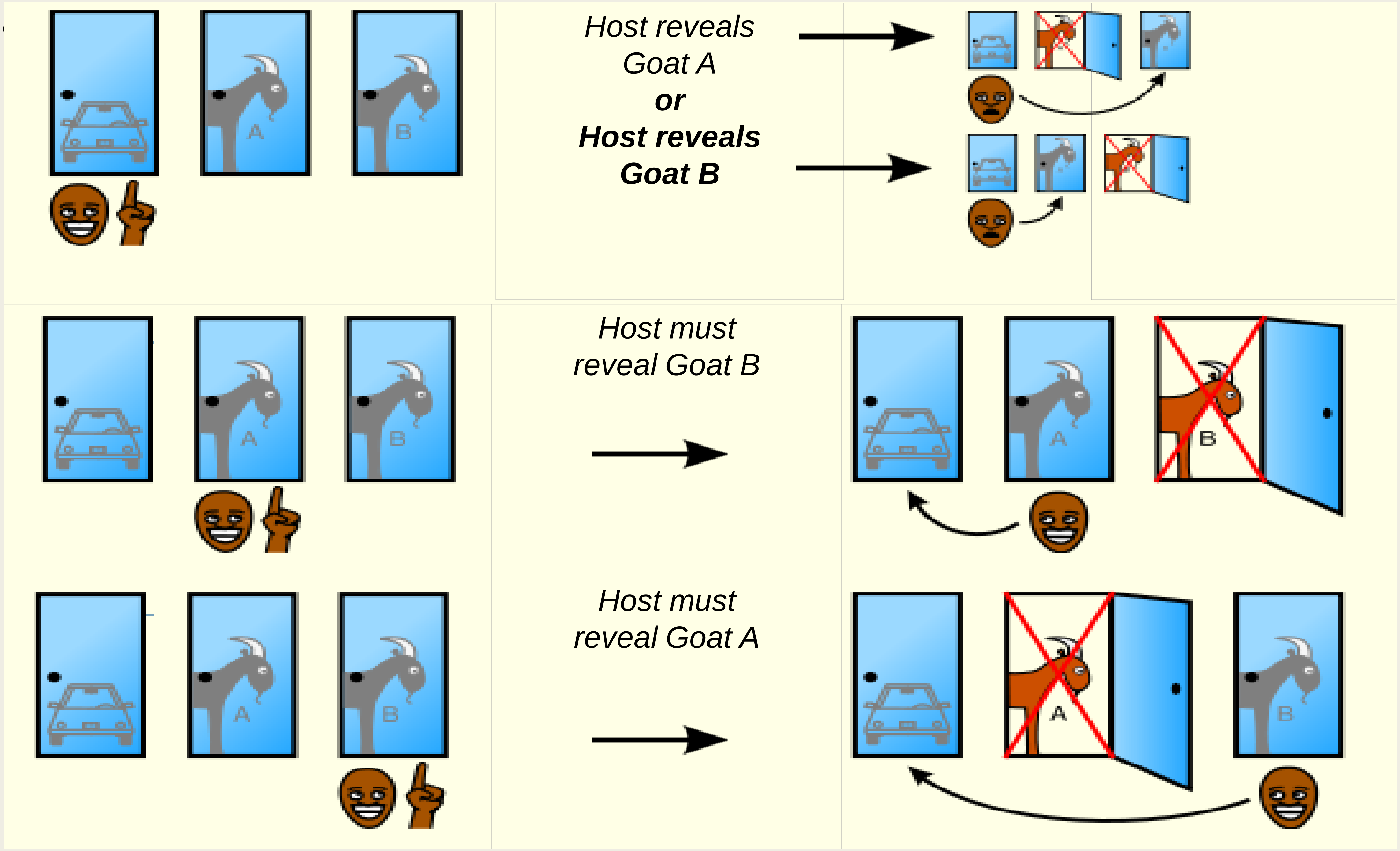
$$= 0.5 \times 0.9802 + 1 \times 0.0198 = 0.5099$$

- $P(A = fair|B = H) = \frac{P(B=H|A=fair) \times P(fair)}{P(B=H)} = \frac{0.5 \times 0.9802}{0.5099} = 0.9612$

Monty Hall Problem

- You're given the choice of three doors: Behind one door is a car; behind the others, goats.
- You pick a door, say No. 1
- The host, who knows what's behind the doors, opens another door, say No. 3, which has a goat.
- Do you want to pick door No. 2 instead?





A
B
Host reveals Goat A
or
Host reveals Goat B
Host must reveal Goat B
Host must reveal Goat A

Monty Hall Problem: Bayes Rule

C_i : the car is behind door *i*, *i* = 1, 2, 3

$$P(C_i) = 1/3$$

H_{ij} : the host opens door *j* after you pick door *i*

$$P(H_{ij} \mid C_k) = \begin{cases} 0 & i = j & \text{Host จะไม่เปิดประตูที่เลือก} \\ 0 & j = k & \text{Host จะไม่เปิดประตูที่มีรถอยู่หลังประตู} \\ 1/2 & i = k & \text{Host จะเปิดประตูอื่นๆ ที่คุณไม่เลือกและไม่มีรถอยู่หลังประตู} \\ 1 & i \neq k, j \neq k & \text{Host จะเปิดประตูถ้าหลังประตูที่เลือกไม่มีรถ} \end{cases}$$

Monty Hall Problem: Bayes Rule cont.

- WLOG, i=1, j=3
- $$P(C_1|H_{13}) = \frac{P(H_{13}|C_1)P(C_1)}{P(H_{13})}$$
- $$P(H_{13}|C_1)P(C_1) = \frac{1}{2}\cdot\frac{1}{3} = \frac{1}{6}$$

Monty Hall Problem: Bayes Rule cont.

- $$\begin{aligned} P(H_{13}) &= P(H_{13}, C_1) + P(H_{13}, C_2) + P(H_{13}, C_3) \\ &= P(H_{13} \mid C_1) P(C_1) + P(H_{13} \mid C_2) P(C_2) \\ &= \frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} \\ &= \frac{1}{2} \end{aligned}$$
- $$P(C_1 \mid H_{13}) = \frac{1/6}{1/2} = \frac{1}{3}$$

Monty Hall Problem: Bayes Rule cont.

$$P\left(C_1 \middle| H_{13}\right) = \frac{1/6}{1/2} = \frac{1}{3}$$

$$P\left(C_2 \middle| H_{13}\right) = 1 - \frac{1}{3} = \frac{2}{3} > P\left(C_1 \middle| H_{13}\right)$$

You should switch!

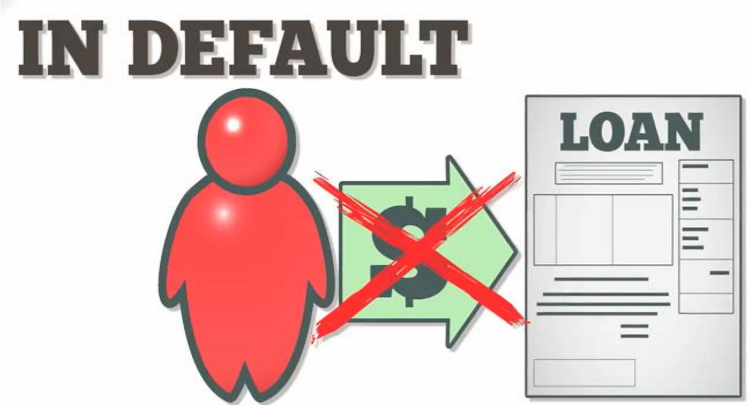
Classification

- Customer Churn

Classification

- Default Problem

A default can occur when a borrower is unable to make timely payments, misses payments, or avoids or stops making payments. Individuals, businesses, and even countries can default if they cannot keep up their debt obligations. Default risks are often calculated well in advance by creditors.



Classification

- Fraud Detection

Naïve Bayes

$$\underbrace{P(c \mid x)}_{\text{Posterior Probability}} = \frac{\overbrace{P(x \mid c)}^{\text{Likelihood}}\,\overbrace{P(c)}^{\text{Class Prior Probability}}}{\underbrace{P(x)}_{\text{Predictor Prior Probability}}}$$

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Example Naïve Bayes

Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	false	N
sunny	hot	high	true	N
overcast	hot	high	false	P
rain	mild	high	false	P
rain	cool	normal	false	P
rain	cool	normal	true	N
overcast	cool	normal	true	P
sunny	mild	high	false	N
sunny	cool	normal	false	P
rain	mild	normal	false	P
sunny	mild	normal	true	P
overcast	mild	high	true	P
overcast	hot	normal	false	P
rain	mild	high	true	N

P (ออกไปเล่น) และ N (ไม่ออกไปเล่น)

Example Naïve Bayes

P(c) Prior probability คือ ความน่าจะเป็นของ Class

- P(P) = 9/14 =(0.64)
- P(N) =5/14 = (0.36)

Example Naïve Bayes

P(x|c) Likelihood คือ ความน่าจะเป็นของแต่ละแอททริบิวต์

Outlook

P(sunny|P) = 2/9	P(sunny|N) = 3/5
P(overcast|P) = 4/9	P(overcast|N) = 0/5
P(rain|P) = 3/9	P(rain|N) = 2/5

Temperature

P(hot|P) = 2/9	P(hot|N) = 2/5
P(mild|P) = 4/9	P(mild|N) = 2/5
P(cool|P) = 3/9	P(cool|N) = 1/5

Humitdity

P(high|P) = 3/9	P(high|N) = 4/5
P(normal|P) = 6/9	P(normal|N) = 1/5

Windy

P(true|P) = 3/9	P(true|N) = 3/5
P(false|P) = 6/9	P(false|N) = 2/5

Example Naïve Bayes

P(C= P| X (Outlook = Sunny, Temp = Cool, Hum = High, Wind = True) คือ ความน่าจะเป็นที่จัดในClass Play (P) เมื่อแอททริบิวต์เป็นดังนี้ **Outlook = Sunny, Temp = Cool, Hum = High, Wind = True**

= 0.0053

P(C= N| X (Outlook = Sunny, Temp = Cool, Hum = High, Wind = True) คือ ความน่าจะเป็นที่จัดในClass No Play (N) เมื่อแอททริบิวต์เป็นดังนี้ **Outlook = Sunny, Temp = Cool, Hum = High, Wind = True**

= 0.0206

Example Naïve Bayes

P(x) คือ ความน่าจะเป็นของแต่ละแอททริบิวต์

- P(Outlook = Sunny) = 5/14
- P(Temp = Cool) = 4/14
- P(Hum = High) = 7/14
- P(Wind = True) = 6/14
- …

Example

From a previous example, we know that 49% of the population are female. Of the female patients, 8% are high risk for heart attack, while 12% of the male patients are high risk. A single person is selected at random and found to be high risk. <u>What is the probability that it is a male?</u>

Define H: high risk F: female M: male

We know:

P(F) =	.49
P(M) =	.51
P(H|F) =	.08
P(H|M) =	.12

Prior Probability

likelihood

$$P(M \mid H) = \frac{P(M)P(H \mid M)}{P(M)P(H \mid M) + P(F)P(H \mid F)}$$

$$= \frac{.51\,(.12)}{.51\,(.12) + .49\,(.08)} = .61$$

Evidence

Example

Suppose a rare disease infects one out of every 1000 people in a population. And suppose that there is a good, but not perfect, test for this disease: if a person has the disease, the test comes back positive 99% of the time. On the other hand, the test also produces some false positives: 2% of uninfected people are also test positive. <u>And someone just tested positive. What are his chances of having this disease?</u>

Example

Define A: has the disease   B: test positive

We know:

$P(A) = .001$ $P(A^c) = .999$

$P(B|A) = .99$ $P(B|A^c) = .02$

We want to know $P(A|B)=?$

$$P(A \mid B) = \frac{P(A)P(B|A)}{P(A)P(B|A)+P(A^C)P(B|A^C)}$$

$$= \frac{.001 \times .99}{.001 \times .99 + .999 \times .02} = .0472$$

Example

A survey of job satisfaction[2] of teachers was taken, giving the following results

		Job Satisfaction		
		Satisfied	Unsatisfied	Total
LEVEL	College	74	43	117
	High School	224	171	395
	Elementary	126	140	266
	Total	424	354	778

[2] "Psychology of the Scientist: Work Related Attitudes of U.S. Scientists" (*Psychological Reports* (1991): 443 – 450).

Example

If all the cells are divided by the total number surveyed, 778, the resulting table is a table of empirically derived probabilities.

		Job Satisfaction		
		Satisfied	Unsatisfied	Total
LEVEL	College	0.095	0.055	0.150
	High School	0.288	0.220	0.508
	Elementary	0.162	0.180	0.342
	Total	0.545	0.455	1.000

		Job Satisfaction		
		Satisfied	Unsatisfied	Total
LEVEL	College	0.095	0.055	0.150
	High School	0.288	0.220	0.508
	Elementary	0.162	0.180	0.342
	Total	0.545	0.455	1.000

For convenience, let C stand for the event that the teacher teaches college, S stand for the teacher being satisfied and so on. Let's look at some probabilities and what they mean.

$P(C) = 0.150$ is the proportion of teachers who are college teachers.

$P(S) = 0.545$ is the proportion of teachers who are satisfied with their job.

$P(C \cap S) = 0.095$ is the proportion of teachers who are college teachers and who are satisfied with their job.

		Job Satisfaction		
		Satisfied	Unsatisfied	Total
LEVEL	College	0.095	0.055	0.150
	High School	0.288	0.220	0.508
	Elementary	0.162	0.180	0.342
	Total	0.545	0.455	1.000

$$P(C \mid S) = \frac{P(C \cap S)}{P(S)}$$

$$= \frac{0.095}{0.545} = 0.175$$

is the proportion of teachers who are college teachers given they are satisfied. Restated: This is the proportion of satisfied that are college teachers.

$$P(S \mid C) = \frac{P(S \cap C)}{P(C)}$$

$$= \frac{P(C \cap S)}{P(C)} = \frac{0.095}{0.150}$$

$$= 0.632$$

is the proportion of teachers who are satisfied given they are college teachers. Restated: This is the proportion of college teachers that are satisfied.

		Job Satisfaction		
		Satisfied	Unsatisfied	Total
LEVEL	College	0.095	0.055	0.150
	High School	0.288	0.220	0.508
	Elementary	0.162	0.180	0.342
	Total	0.545	0.455	1.000

Are C and S independent events?

$$P(C) = 0.150 \text{ and } P(C \mid S) = \frac{P(C \cap S)}{P(S)} = \frac{0.095}{0.545} = 0.175$$

P(C|S) $\neq$ P(C) so C and S are dependent events.

		Job Satisfaction		
		Satisfied	Unsatisfied	Total
L E V E L	College	0.095	0.055	0.150
	High School	0.288	0.220	0.508
	Elementary	0.162	0.180	0.658
	Total	0.545	0.455	1.000

$P(C \cup S)$?

$P(C) = 0.150$, $P(S) = 0.545$ and

$P(C \cap S) = 0.095$, so

$$P(C \cup S) = P(C)+P(S) - P(C \cap S)$$

$$= 0.150 + 0.545 - 0.095$$

$$= 0.600$$

Example

Tom and Dick are going to take a driver's test at the nearest DMV office. Tom estimates that his chances to pass the test are 70% and Dick estimates his as 80%. Tom and Dick take their tests independently.

Define D = {Dick passes the driving test}

T = {Tom passes the driving test}

T and D are independent.

$P(T) = 0.7$, $P(D) = 0.8$

Example

What is the probability that at most one of the two friends will pass the test?

P(At most one person pass)

$= P(D^c \cap T^c) + P(D^c \cap T) + P(D \cap T^c)$

$= (1 - 0.8)\,(1 – 0.7) + (0.7)\,(1 – 0.8) + (0.8)\,(1 – 0.7)$

$= .44$

P(At most one person pass)

$= 1\text{-}P(\text{both pass}) = 1\text{-}\ 0.8 \times 0.7 = .44$

Example

What is the probability that at least one of the two friends will pass the test?

P(At least one person pass)

$= P(D \cup T)$

$= 0.8 + 0.7 - 0.8 \times 0.7$

$= .94$

P(At least one person pass)

$= 1-P(\text{neither passes}) = 1- (1-0.8) \times (1-0.7) = .94$

Example

Suppose we know that only one of the two friends passed the test. What is the probability that it was Dick?

P(D | exactly one person passed)

= P(D $\cap$ exactly one person passed) / P(exactly one person passed)

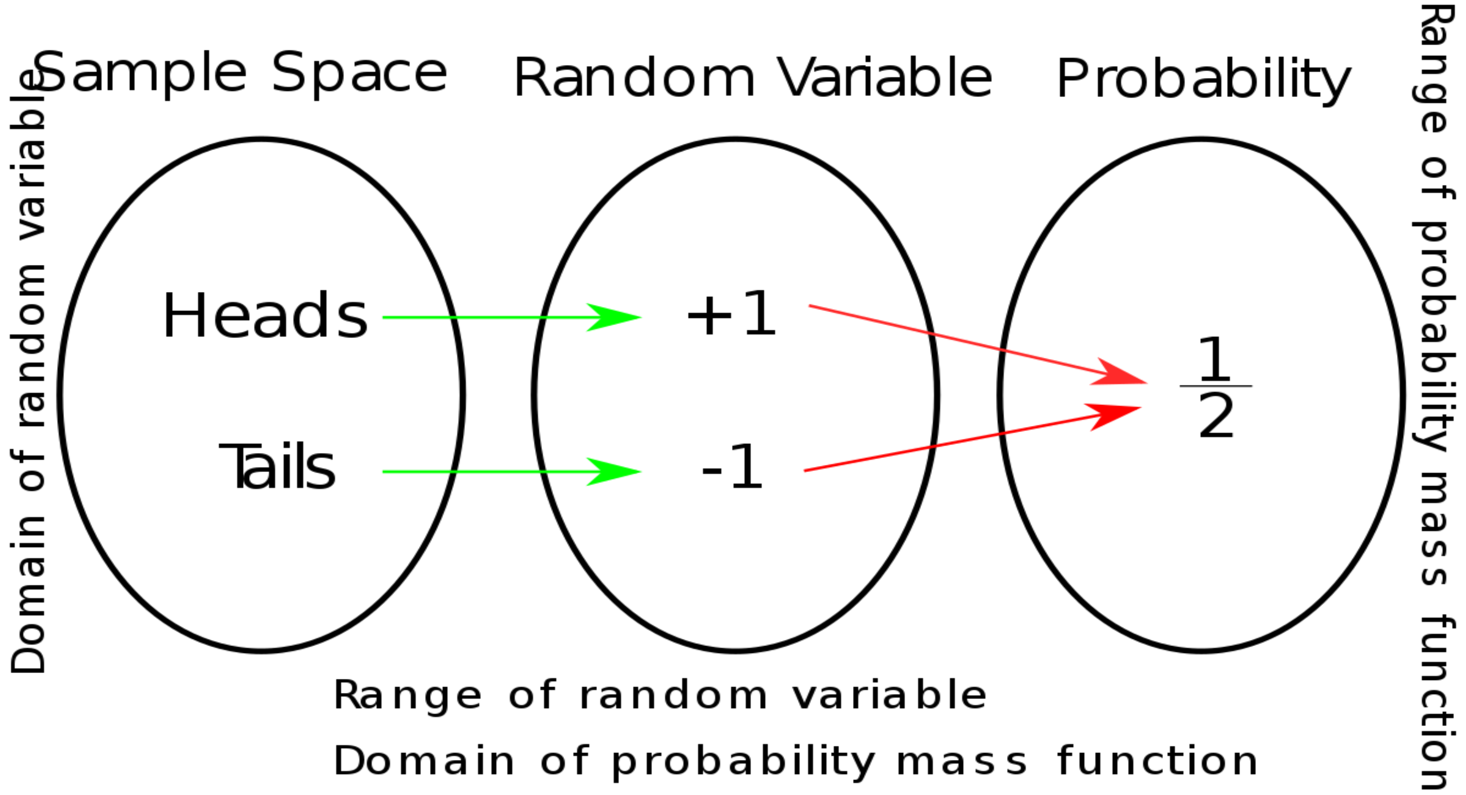
= $P(D \cap T^c) / (P(D \cap T^c) + P(D^c \cap T))$

= 0.8 x (1-0.7)/(0.8 x (1-0.7)+(1-.8) x 0.7)

= .63

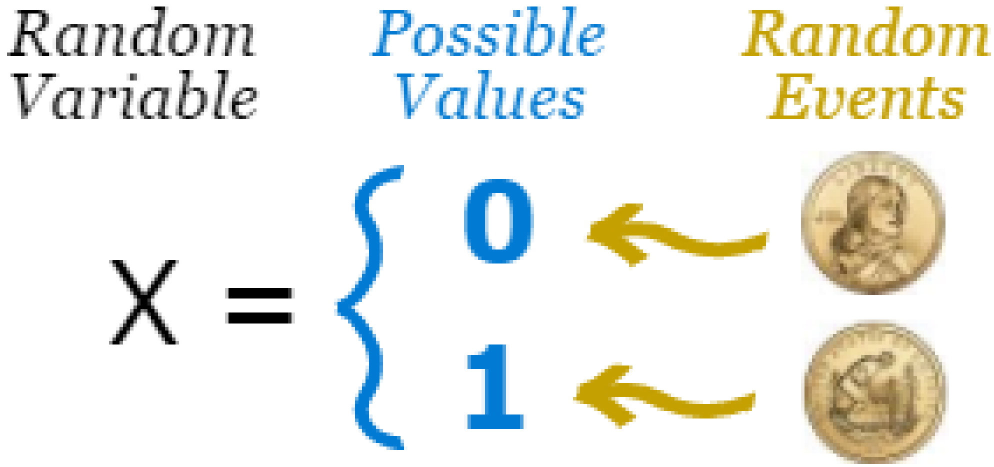
Random Variable

- A random variable is a function that assigns a real number to each outcome in the sample space of a random experiment



Random variables

- Example: three tosses of a coin
  - S = {HHH,THH,HTH,HHT,HTT,THT,TTH,TTT}
  - *Random variable X = number of observed tails*
  - *Possible values for X = {0,1, 2, 3}*
- Why do we need random variables?
  - *We use them as a **model for our observed data***



Types of Random variables

- **Discrete random variables** have a countable number of outcomes
  - *Examples: Dead/alive, treatment/placebo, dice, counts, etc.*
- **Continuous random variables** have an infinite continuum of possible values.
  - *Examples: blood pressure, weight, the speed of a car, the real numbers from 1 to 6.*

Distributions of Discrete Variables

- Frequency table
- Frequency distribution

Color	Frequency
Brown	17
Red	18
Yellow	7
Green	7
Blue	2
Orange	4

Distributions of Continuous Variables

Example: Response times

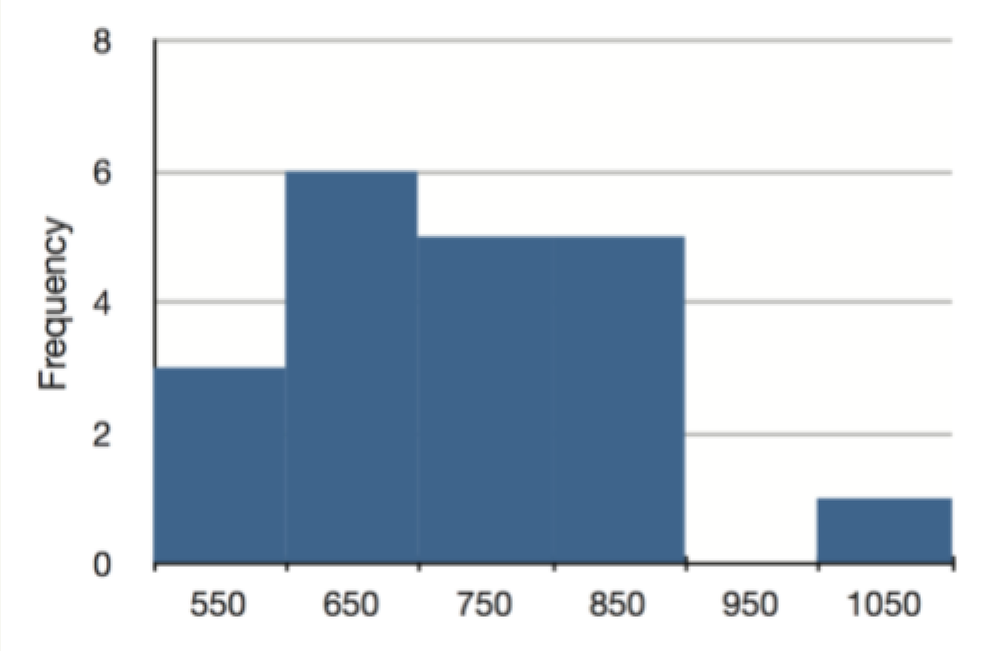
568.00	720.25
577.10	728.00
581.00	729.00
640.00	777.00
641.00	808.00
645.00	824.00
657.13	825.00
673.32	865.00
696.00	875.00
703.00	1007.00

Grouped frequency distribution

Range	Frequency
500-600	3
600-700	6
700-800	5
800-900	5
900-1000	0
1000-1100	1

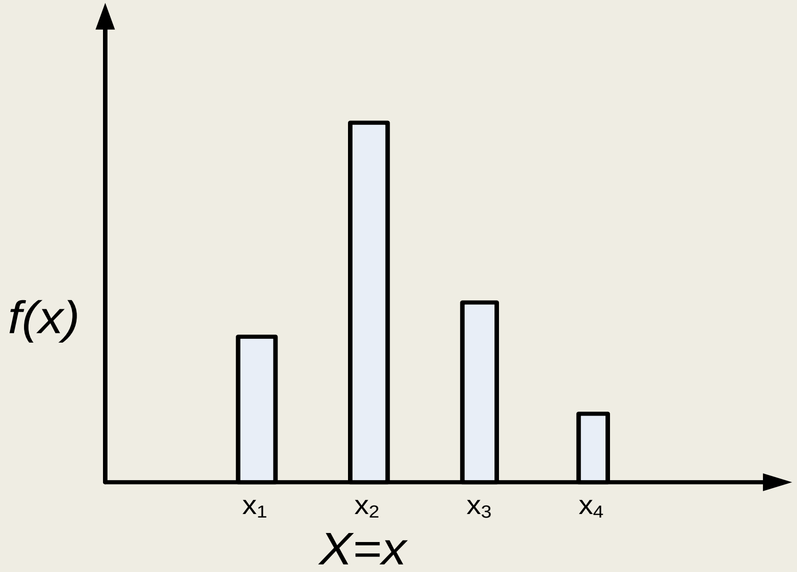
Distributions of Continuous Variables

A histogram of the grouped frequency distribution:

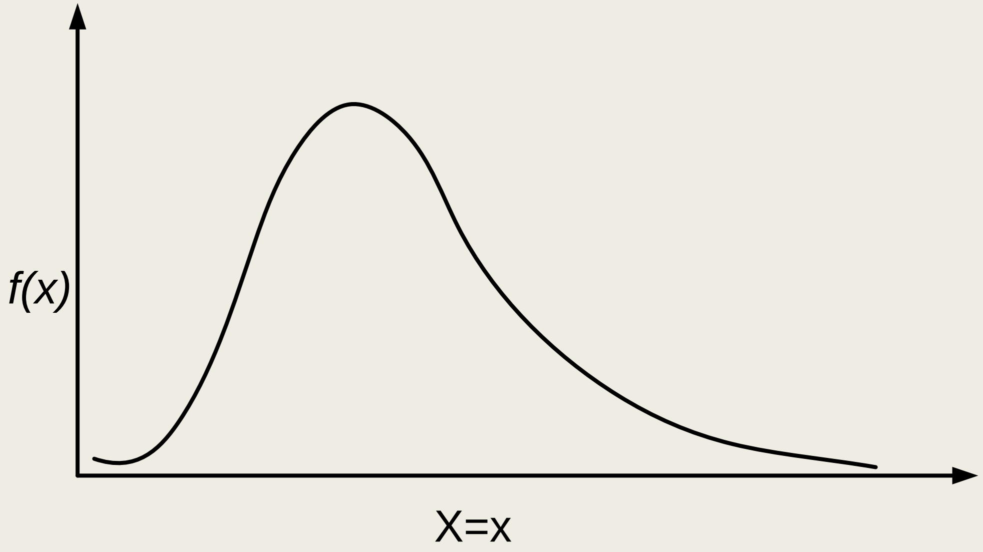


The labels on the X-axis are the middle values of the range they represent

Discrete VS Continuous Probability Distributions



Discrete Probability distribution



Continuous Probability Distribution

When the random variable of interest can take any value in an interval, it is called continuous random variable.

- Every continuous random variable has an infinite, uncountable number of possible values (i.e., any value in an interval)

Consequently, continuous random variable differs from discrete random variable.

Probability Functions

- A probability function maps the possible values of *x* against their respective probabilities of occurrence, *p(x)*
- *p(x)* is a number from 0 to 1.0.
- The area under a probability function is always 1.
- For a **discrete random variable**, a probability function is called "probability mass function (pmf)" which contains the probability of each possible outcomes
- For a **continuous random variable**, a probability function is called "probability density function (pdf)" which the probability of any one outcome is zero (if you specify it to enough decimal places)

Probability Distributions for Discrete Random Variables

The **probability distribution for a discrete random variable *x*** resembles the relative frequency distributions we constructed in Chapter 1. It is a graph, table or formula that gives the possible values of *x* and the probability *p(x)* associated with each value.

We must have

$$0 \le p(x) \le 1 \text{ and } \sum p(x) = 1$$

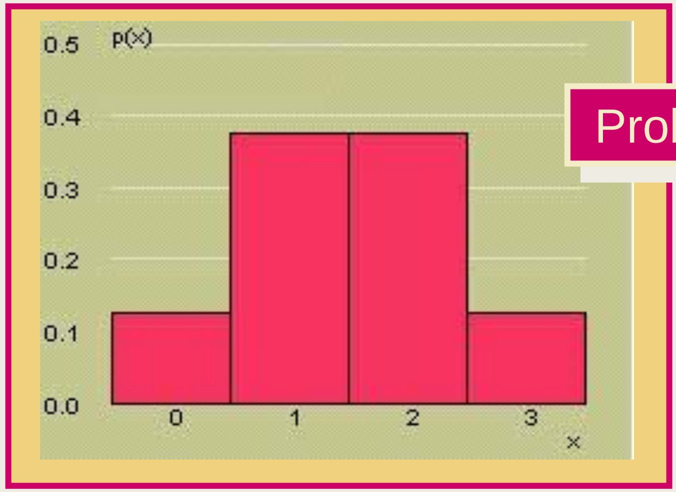
Example

Toss a fair coin three times and define ***x* = number of heads**.

		X
HHH	1/8	3
HHT	1/8	2
HTH	1/8	2
THH	1/8	2
HTT	1/8	1
THT	1/8	1
TTH	1/8	1
TTT	1/8	0

$P(x = 0) = 1/8$
$P(x = 1) = 3/8$
$P(x = 2) = 3/8$
$P(x = 3) = 1/8$

x	$p(x)$
0	1/8
1	3/8
2	3/8
3	1/8

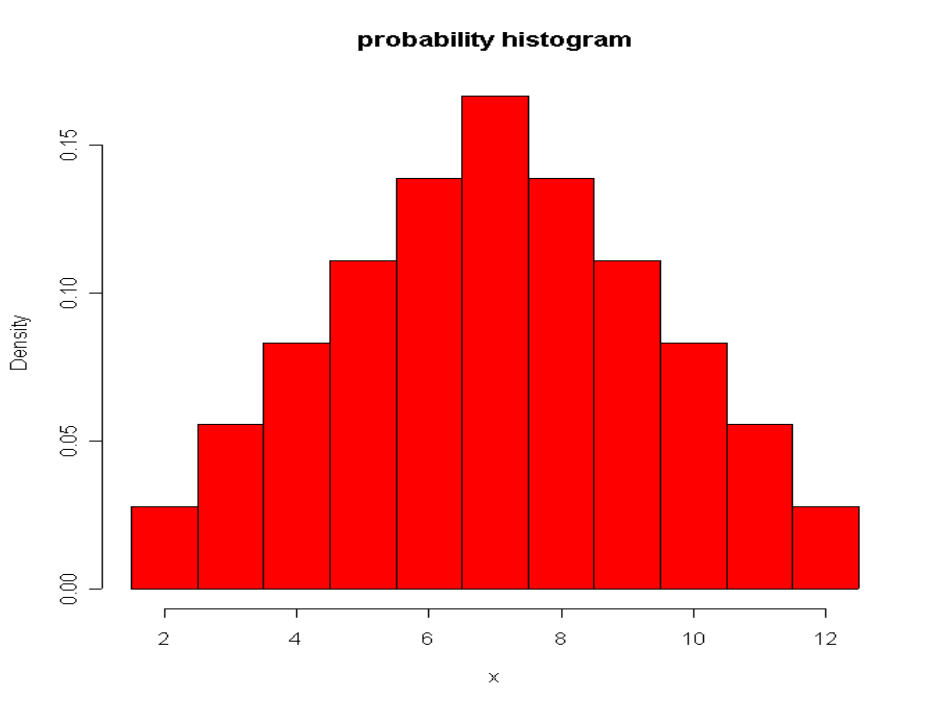


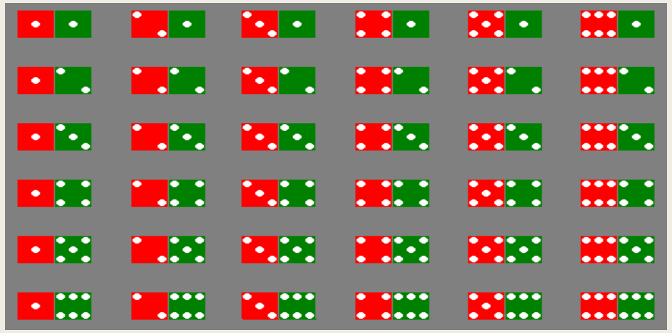
Probability Histogram for *x*

Example

Toss two dice and define

X = **sum of two dice.**





x	$p(x)$
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

Probability Distributions

Probability distributions can be used to describe the population, just as we described samples in Chapter 1.

- ***Shape:*** *Symmetric, skewed, mound-shaped…*
- ***Outliers:*** *unusual or unlikely measurements*
- ***Center and spread:*** *mean and standard deviation. A population mean is called* ***m*** *and a population standard deviation is called* ***s.***

Discrete Random Variables

- A discrete random variable has a **finite** or **countable** number of distinct values

- Discrete random variables can be summarized by listing all values along with the probabilities
  - *Called a **probability distribution***

- Example: number of members in US families

X	2	3	4	5	6	7
P(X)	0.413	0.236	0.211	0.090	0.032	0.018

Another Example

- Random variable X = the sum of two dice
  - *X takes on values from 2 to 12*

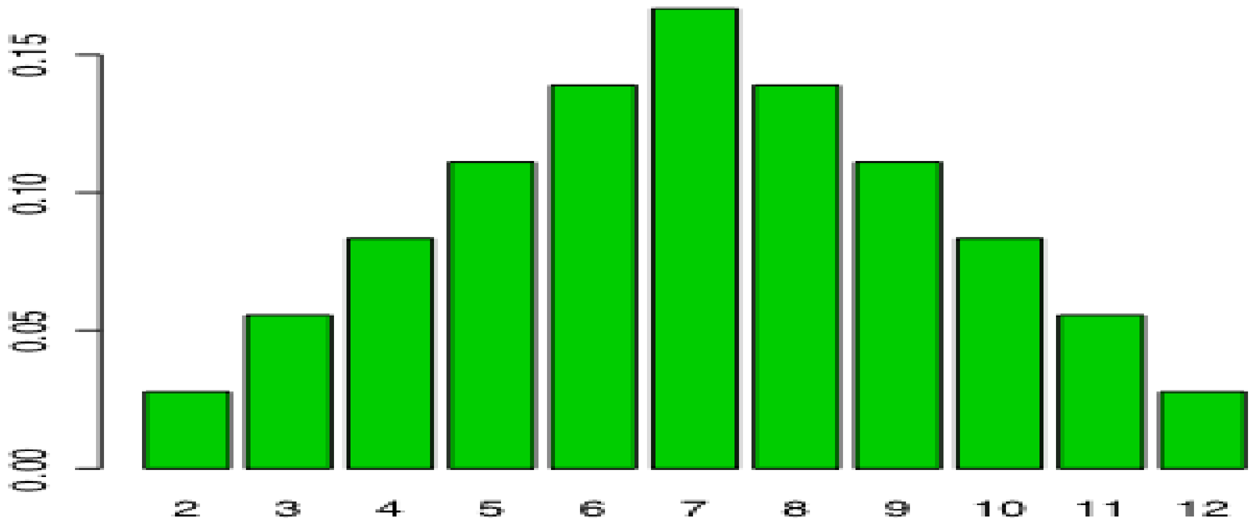
- Use "equally-likely outcomes" rule to calculate the probability distribution:

X	2	3	4	5	6	7	8	9	10	11	12
# of Outcomes	1	2	3	4	5	6	5	4	3	2	1
P(X)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

- If discrete r.v. takes on many values, it is better to use a **probability histogram**

Probability Histograms

- Probability histogram of sum of two dice:



- Using the disjoint addition rule, probabilities for discrete random variables are calculated by adding up the “bars” of this histogram:

$$P(sum > 10) = P(sum = 11) + P(sum = 12) = 3/36$$

Expected Values of Discrete Random Variables

- The **mean**, or **expected value**, of a **discrete random variable** is

$$\mu = E(x) = \sum xp(x).$$

Variance of Discrete Random Variables

- The **variance** of a **discrete random variable** *x* is

$$\sigma^2 = E[(x-\mu)^2] = \sum (x-\mu)^2 p(x) = E(x^2) - [E(\mathrm{x})]^2.$$

- The **standard deviation** of a **discrete random variable** *x* is

$$\sqrt{\sigma^2} = \sqrt{E[(x-\mu)^2]} = \sqrt{\sum (x-\mu)^2 p(x)}.$$

Example

Toss a fair coin 3 times and record *x* the number of heads.

x	$p(x)$	$xp(x)$	$(x-\mu)^2 p(x)$
0	1/8	0	$(-1.5)^2(1/8)$
1	3/8	3/8	$(-0.5)^2(3/8)$
2	3/8	6/8	$(0.5)^2(3/8)$
3	1/8	3/8	$(1.5)^2(1/8)$

$$\mu = \sum xp(x) = \frac{12}{8} = 1.5$$

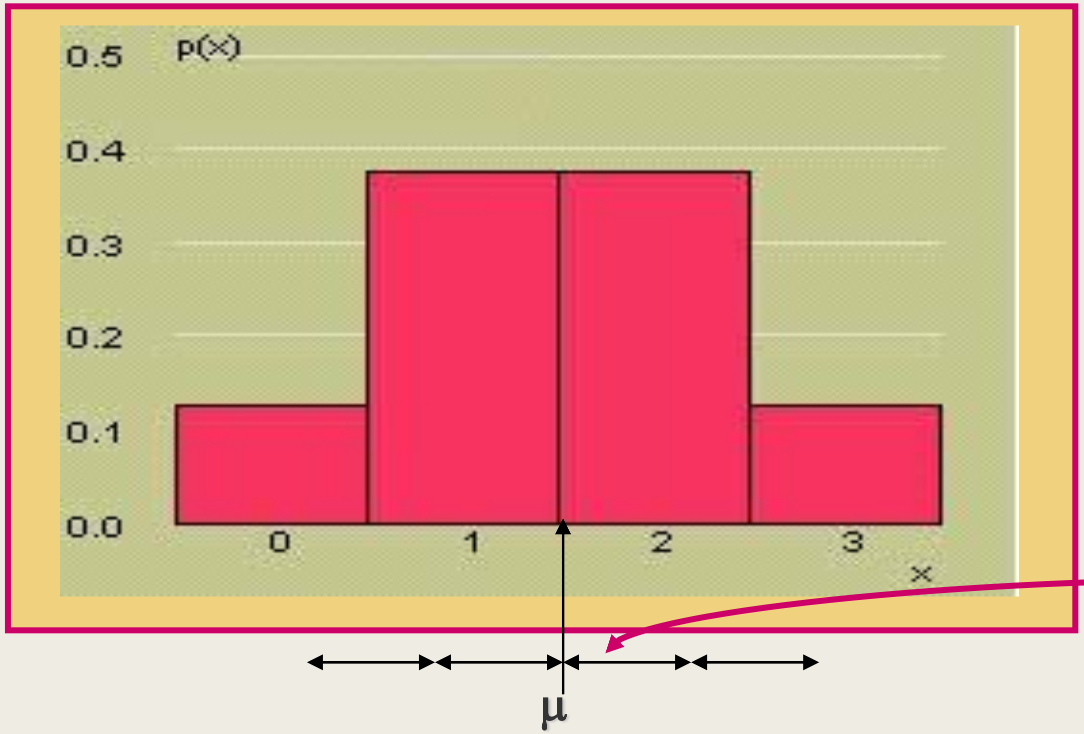
$$\sigma^2 = \sum (x-\mu)^2 p(x)$$

$$\sigma^2 = .28125 + .09375 + .09375 + .28125 = .75$$

$$\sigma = \sqrt{.75} = .688$$

Example

The probability distribution for *x* the number of heads in tossing 3 fair coins.



- Shape? Symmetric; mound-shaped
- Outliers? None
- Center? $\mu = 1.5$
- Spread? $\sigma = .688$

The Binomial Distribution

- A Binomial Random Variable
  - *n identical trials*
  - *Two outcomes: **S**uccess or **F**ailure*
  - *$P(\mathbf{S}) = p$; $P(\mathbf{F}) = q = 1 - p$*
  - *Trials are independent*
  - *x is the number of **S**uccesses in n trials*

The Binomial Distribution

- A Binomial Random Variable
  - *n identical trials* → *Flip a coin 3 times*
  - *Two outcomes: **S**uccess or **F**ailure* → *Outcomes are Heads or Tails*
  - $P(\mathbf{S}) = p;\ P(\mathbf{F}) = q = 1 - p$ → *P(H) = .5; P(T) = 1-.5 = .5*
  - *Trials are independent* → *A head on flip i doesn't change P(H) of flip i + 1*
  - *x is the number of **S**'s in n trials*

The Binomial Distribution

Results of 3 flips	Probability	Combined	Summary
HHH	$(p)(p)(p)$	p^3	$(1)p^3q^0$
HHT	$(p)(p)(q)$	p^2q	$(3)p^2q^1$
HTH	$(p)(q)(p)$	p^2q	
THH	$(q)(p)(p)$	p^2q	
HTT	$(p)(q)(q)$	pq^2	$(3)p^1q^2$
THT	$(q)(p)(q)$	pq^2	
TTH	$(q)(q)(p)$	pq^2	
TTT	$(q)(q)(q)$	q^3	$(1)p^0q^3$

The Binomial Distribution

- The Binomial Probability Distribution
  - *p = P(**S**) on a single trial*
  - *q = 1 – p*
  - *n = number of trials*
  - *x = number of successes*

$$P(x) = \binom{n}{x} p^x q^{n-x}$$

The Binomial Distribution

- The Binomial Probability Distribution

The number of ways of getting the desired results

The probability of getting the required number of successes

The probability of getting the required number of failures

$$P(x) = \binom{n}{x} p^{x} q^{n-x}$$

The Binomial Distribution

- Say 40% of the class is female.
- What is the probability that 6 of the first 10 students walking in will be female?

$$P(x) = \binom{n}{x} p^x q^{n-x}$$

$$= \binom{10}{6} (.4^6)(.6^{10-6})$$

$$= 210(.004096)(.1296)$$

$$= .1115$$

The Binomial Distribution: Descriptive measures

- A Binomial Random Variable has

$$\textit{Mean} \quad \mu = np$$

$$\textit{Variance} \quad \sigma^2 = npq$$

$$\textit{Standard Deviation} \quad \sigma = \sqrt{npq}$$

The Binomial Distribution

- For 1,000 coin flips,

$$\mu = np = 1000 \cdot .5 = 500$$

$$\sigma^2 = npq = 1000 \cdot .5 \cdot .5 = 250$$

$$\sigma = \sqrt{npq} = \sqrt{250} \cong 16$$

The actual probability of getting exactly 500 heads out of 1000 flips is just over 2.5%, but the probability of getting between 484 and 516 heads (that is, within one standard deviation of the mean) is about 68%.

The Poisson Distribution

- There are some experiments, which involve the occurring of the number of outcomes during a given time interval (or in a region of space). Such a process is called Poisson process.
- Example : Number of clients visiting a ticket selling counter in a metro station.

The Poisson Distribution

- Evaluates the probability of a (usually small) number of occurrences out of many opportunities in a …
  - *Period of time*
  - *Area*
  - *Volume*
  - *Weight*
  - *Distance*
  - *Other units of measurement*

The Poisson Distribution

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

- λ = mean number of occurrences in the given unit of time, area, volume, etc.
- e = 2.71828….

The Poisson Distribution: Descriptive measures

- $\mu = \lambda$
- $\sigma^2 = \lambda$

The Poisson Distribution

Suppose there is a disease, whose average incidence is 2 per million people. What is the probability that a city of 1 million people has at least twice the average incidence?

X ~ Poisson (λ=2 million)

$$P(X \geq 4) = 1 - P(X \leq 3)$$

$$= 1 - \frac{e^{-2}2^{0}}{0!} - \frac{e^{-2}2^{1}}{1!} - \frac{e^{-2}2^{2}}{2!} - \frac{e^{-2}2^{3}}{3!}$$

$$= 0.143$$

The Hypergeometric Distribution

- In the binomial situation, each trial was independent.
  - *Drawing cards from a deck and replacing the drawn card each time*
- If the card is *not* replaced, each trial depends on the previous trial(s).
  - *The hypergeometric distribution can be used in this case.*

The Hypergeometric Distribution

- Randomly draw n elements from a set of N elements, without replacement. Assume there are r successes and N-r failures in the N elements.
- The hypergeometric random variable is the number of successes, x, drawn from the r available in the n selections.

The Hypergeometric Distribution

$$P(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

where
N = the total number of elements
r = number of successes in the N elements
n = number of elements drawn
X = the number of successes in the n elements

The Hypergeometric Distribution

$$\mu = \frac{nr}{N}$$

$$\sigma^2 = \frac{r(N-r)n(N-n)}{N^2(N-1)}$$

The Hypergeometric Distribution

- Suppose a customer at a pet store wants to buy two hamsters for his daughter, but he wants two males or two females (i.e., he wants only two hamsters in a few months)
- If there are ten hamsters, five male and five female, what is the probability of drawing two of the same sex? (With hamsters, it's virtually a random selection.)

$$P(M=2)=P(F=2)=\frac{\binom{5}{2}\binom{10-5}{2-2}}{\binom{10}{2}}=\frac{(10)(1)}{45}=.22$$

$$P(M=2\,\text{or}\,F=2)=P(M=2)+P(F=2)=2\times.22=.44$$

Probability Density

Probability Density Function

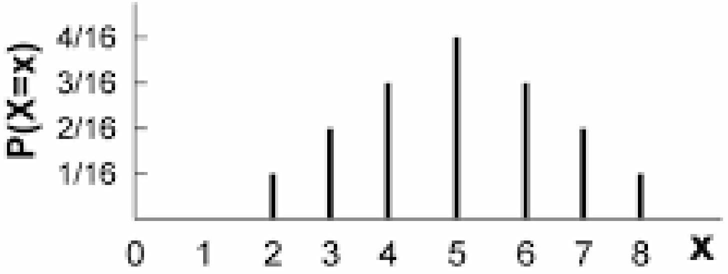
- For a **discrete random variable**, a probability distribution contains the probability of each possible outcomes
- For a **continuous random variable**, the probability of any one outcome is zero (if you specify it to enough decimal places)
- A probability density function is a formula that can be used to compute probabilities of a range of outcomes for a continuous random variable
- The sum of all densities is always 1.0 and the value of function is always greater or equal to zero

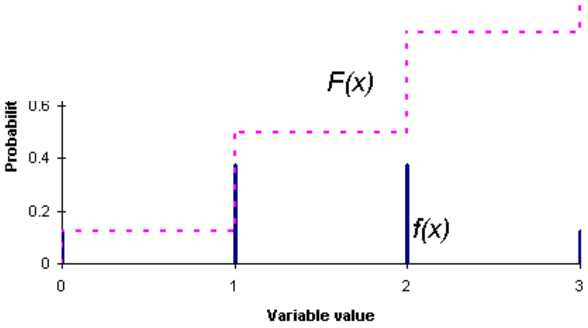
Properties of Probability Mass Function

The function $f(x)$ is a probability mass function for the discrete random variable X, defined over the set of integer numbers R, if

1. $f(x) \geq 0, \text{for all } x \in R$
2. $\sum_{-\infty}^{\infty} f(x) = 1$
3. $f(x_i) = P(X = x_i)$
4. $\mu = \sum_{-\infty}^{\infty} x f(x)$
5. $\sigma^2 = \sum_{-\infty}^{\infty} (x-\mu)^2 f(x)$

x	P(x)
2	1/16
3	2/16
4	3/16
5	4/16
6	3/16
7	2/16
8	1/16

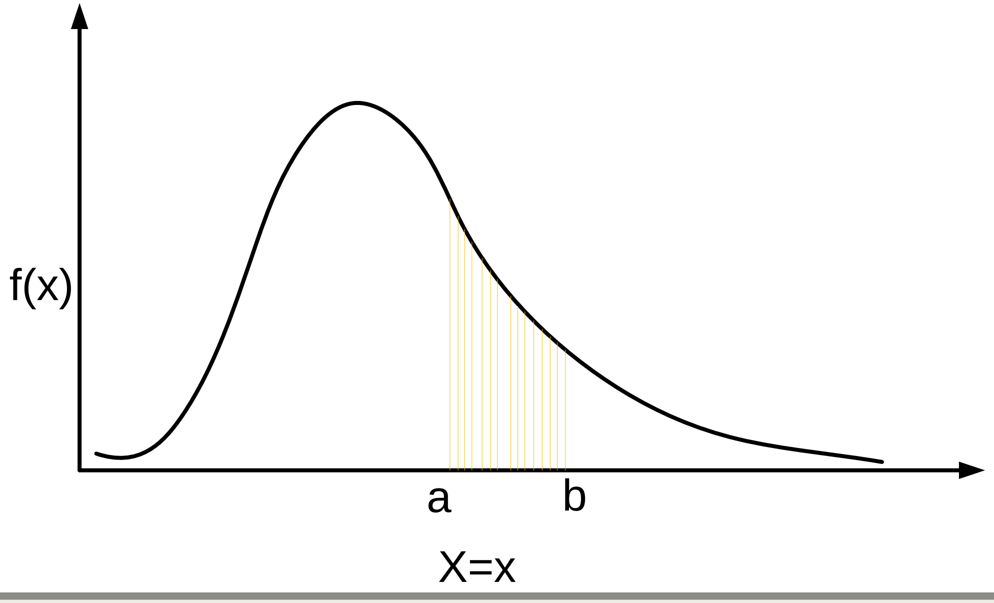




Properties of Probability Density Function

The function $f(x)$ is a probability density function for the continuous random variable X, defined over the set of real numbers R, if

1. $f(x) \geq 0$, for all $x \in R$
2. $\int_{-\propto}^{\propto} f(x)dx = 1$
3. $P(a \leq X \leq b) = \int_{a}^{b} f(x)\,dx$
4. $\mu = \int_{-\propto}^{\propto} xf(x)\,dx$
5. $\sigma^2 = \int_{-\propto}^{\propto} (x-\mu)^2 f(x)dx$



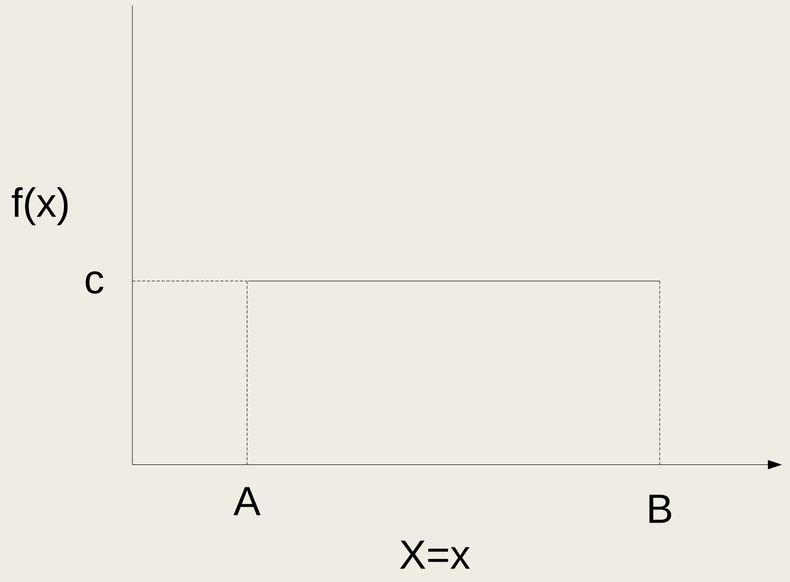
Continuous Uniform Distribution

- One of the simplest continuous distribution in all of statistics is the continuous uniform distribution.

The density function of the continuous uniform random variable X on the interval $[A, B]$ is:

$$f(x{:}\, A, B) = \begin{cases} \dfrac{1}{B - A} & A \leq x \leq B \\ 0 & Otherwise \end{cases}$$

Continuous Uniform Distribution



Note:

a) $\int_{\infty}^{-\infty} f(x)dx = \frac{1}{B-A} \times (B-A) = 1$

$b) P(c < x < d)$= $\frac{d-c}{B-A}$ where both c and d are in the interval (A,B)

$c)\ \mu = \frac{A+B}{2}$

$d) \sigma^2 = \frac{(B-A)^2}{12}$

Normal Distribution

- Also called "bell-shaped distribution"
- Density Function of normal distribution

$$N(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x+\mu)^2}{2\sigma^2}}$$

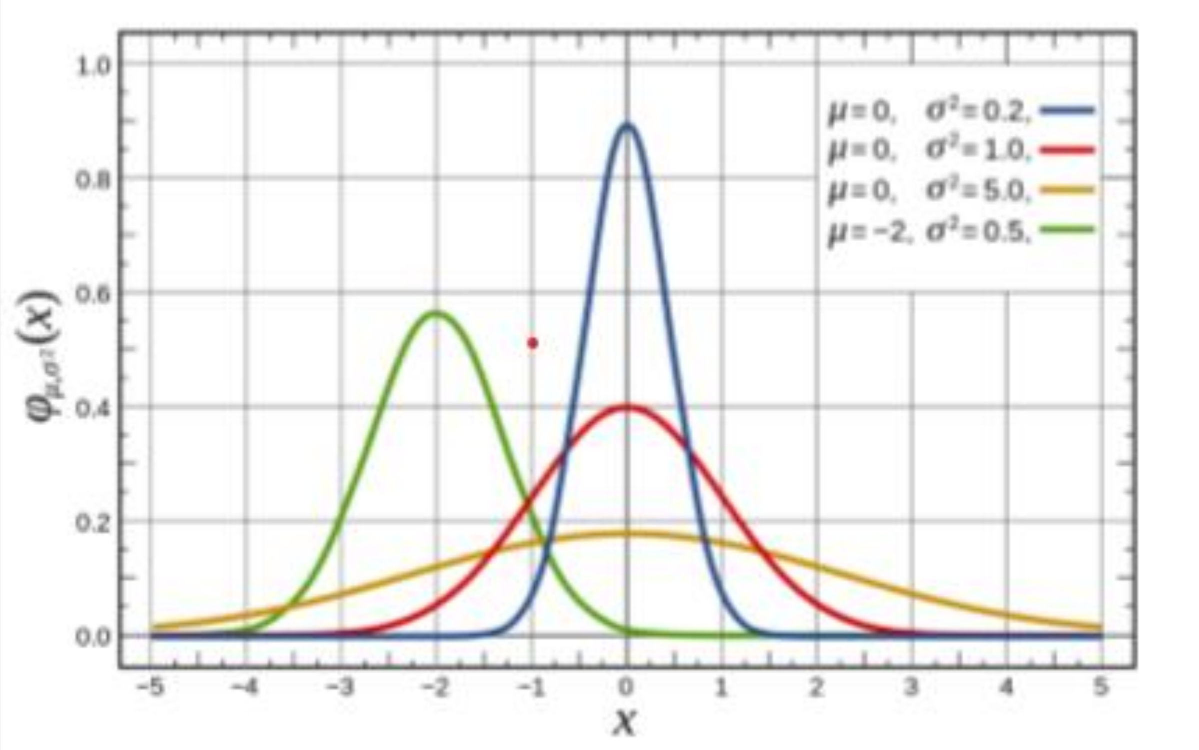
e = 2.71828, π = 3.14159

- A continuous random variable X having the bell-shaped distribution is called a "normal random variable".

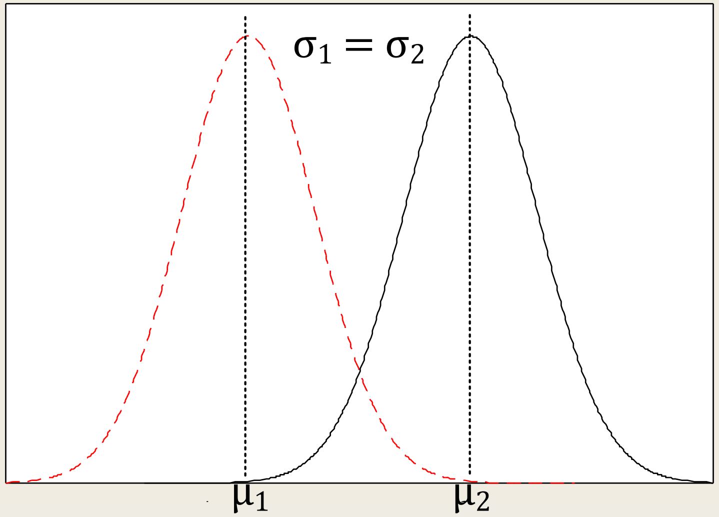
Shapes of Distribution

- Normal distribution vary
- Distribution can be asymmetric
- Distribution can have more than one peak

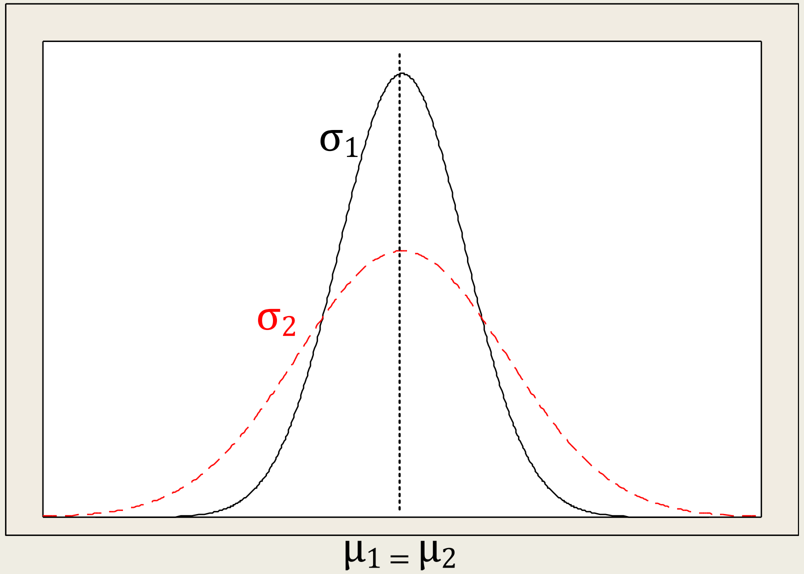
Different normal distributions



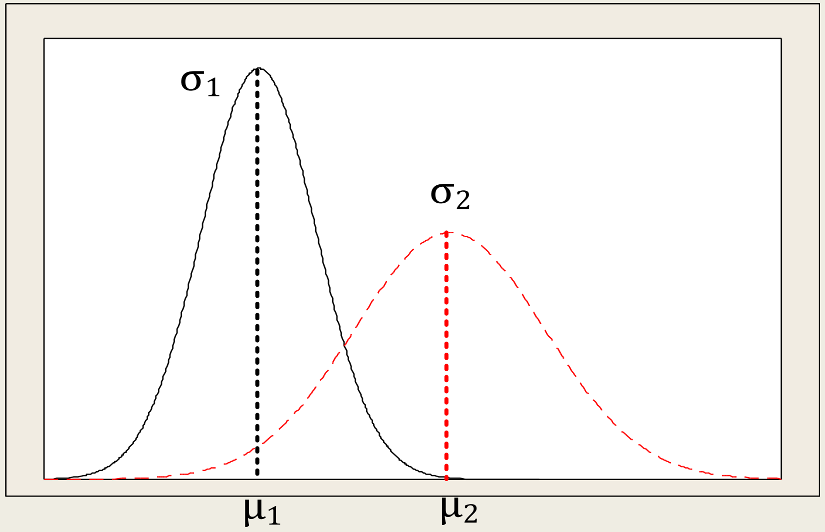
Normal Distribution



Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 = \sigma_2$



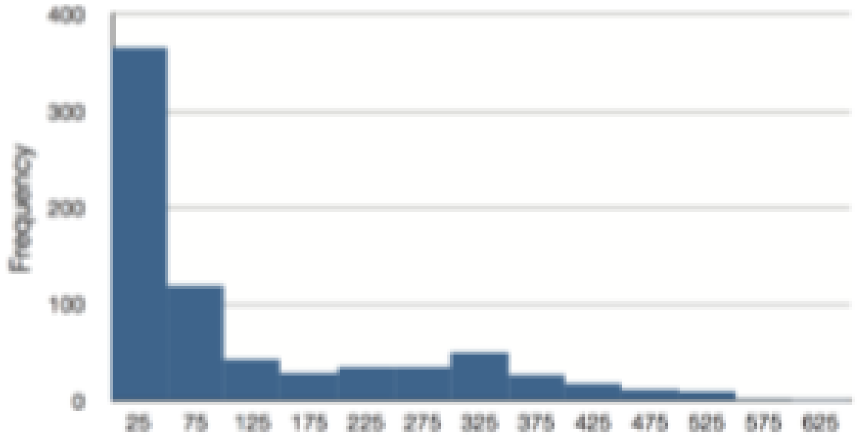
Normal curves with $\mu_1 = \mu_2$ and $\sigma_1 < \sigma_2$



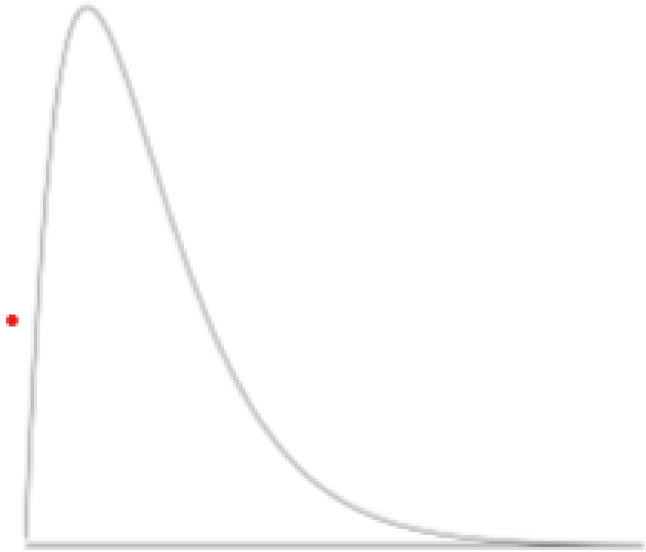
Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 < \sigma_2$

Shapes of Distributions: Skewed Distribution

Positive skew



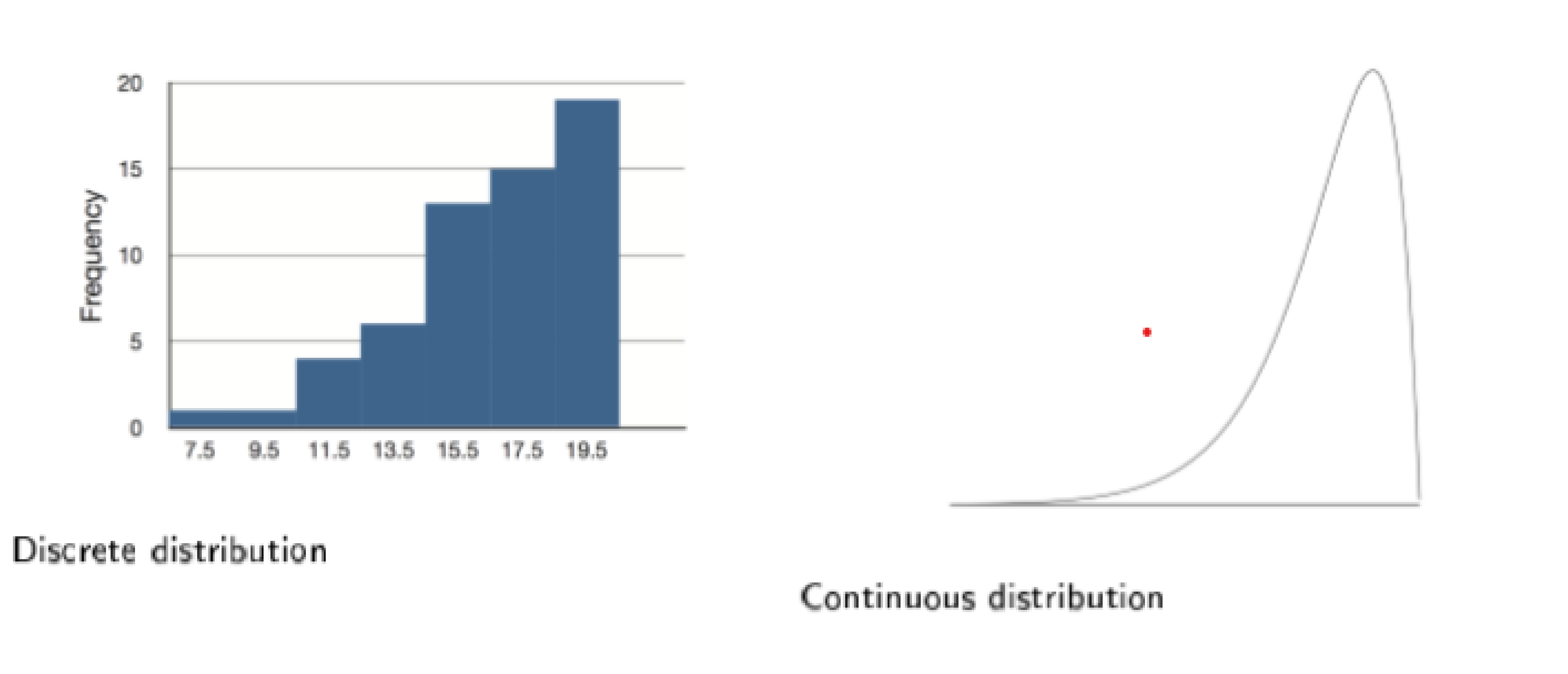
Discrete distribution



Continuous distribution

Shapes of Distributions: Skewed Distribution

Negative skew



Standard Normal Distribution

- The normal distribution has computational complexity to calculate $P(x_1 < x < x_2)$ for any two $(x_1,\ x_2)$ and given μ and σ

- To avoid this difficulty, the concept of z-transformation is followed.

$$\boxed{z = \frac{x-\mu}{\sigma} \qquad \text{[Z-transformation]}}$$

- X: Normal distribution with mean μ and variance σ^2.

- Z: Standard normal distribution with mean $\mu = 0$ and variance σ^2 = 1.

Standard Normal Distribution

- Therefore, if *f(x)* assumes a value, then the corresponding value of $f(z)$ is given by

$$f(x:\mu,\sigma): \quad P(x_1 < x < x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx$$

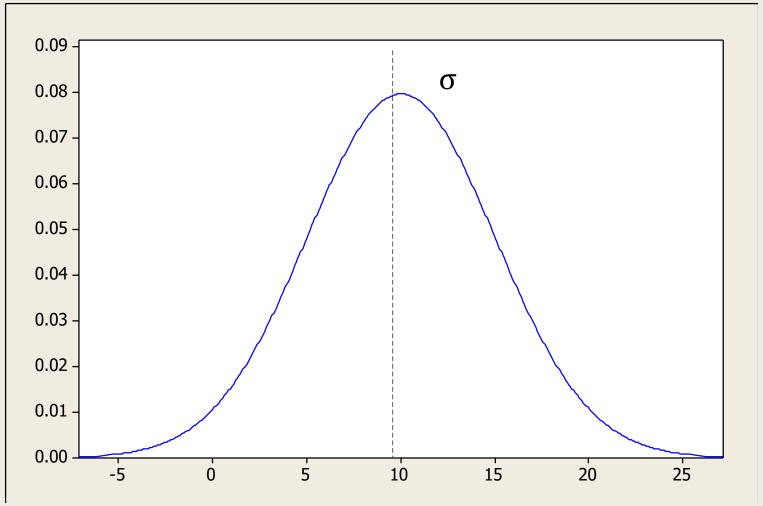
$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{1}{2}z^2} dz$$

$$= f(z:0,\sigma)$$

Standard Normal Distribution

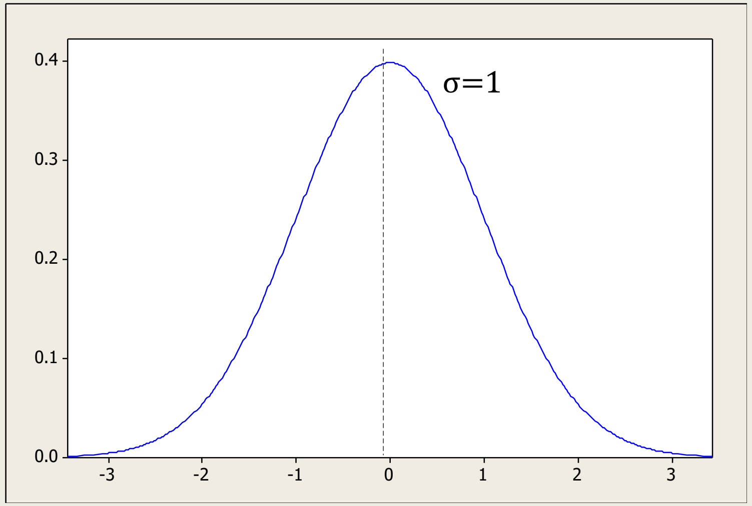
Definition: **Standard Normal Distribution**

The distribution of a normal random variable with mean 0 and variance 1 is called a standard normal distribution.



x=μ

f(x: μ, σ)



μ=0

f(z: 0, 1)