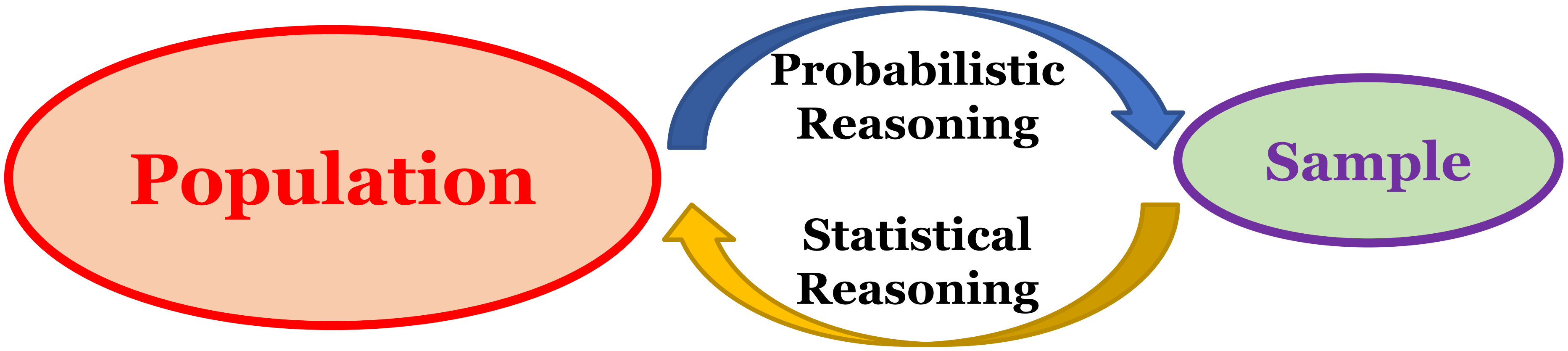
IS4303 Week 9

Probability and Statistics

Overview: Probability and Statistics

- Everything (e.g., decision making) is **uncertain**. Probability provides a **quantitative description of the uncertainty** (i.e., chances or likelihoods) associated with **various outcomes**.
- The link between **Probabilistic Reasoning** and **Statistical Reasoning** tasks:



- Example: Age distribution of people in Singapore
  - I know the overall age distribution, so I can **describe** the proportion of retired people in Clememti
  - I only have a random sample from Clementi, but I can roughly **estimate** the overall proportion of retired people in Singapore

Introduction to Probability

Probability: Some Terms

- **Random Experiment:** A process that generates an observation/measurement with a well-defined outcome
  - Example: (1) Flip coin; (2) Roll dice
- **Basic Outcome:** One possible outcome/result of a random experiment
  - Example: (1) "Head" or "Tail"; (2) "1" or "2" or "3" or "4" or "5" or "6"
- **Sample Space:** A collection of all possible outcomes of a random experiment
  - Example: (1) S={Head, Tail}; (2) S={1, 2, 3, 4, 5, 6}
- **Event:** Any subset of basic outcomes from the sample space
  - Experiment: Roll dice
  - Event A: The number is even → A={2, 4, 6}
  - Event B: The number is at least 5 → B={5, 6}
  - **Simple Event:** The event that has only one outcome in the sample space

Probability: Definitions

A numerical measure of the likelihood that an event will occur:

- **Classical Probability**
  - Assumes all outcomes in the sample space are equally likely to occur:

$$\Pr(\text{event A}) = \frac{N_A}{N_S} = \frac{\text{Number of outcomes that satisfy the event A}}{\text{Total number of possible outcomes in the sample space S}}$$

- **Empirical Probability (Frequentist View)**
  - Relative Frequency: Randomly choose a sample from the population (or repeat experiments/observations n times), and calculate the relative frequency of an event

$$\Pr(\text{event A}) = \frac{f_A}{n} = \frac{\text{Number of observations in the population/sample that satisfy event A}}{\text{Total number of observations in the population/sample}}$$

- Subjective Probability (Bayesian View)
  - Individual measure of belief that an event will occur

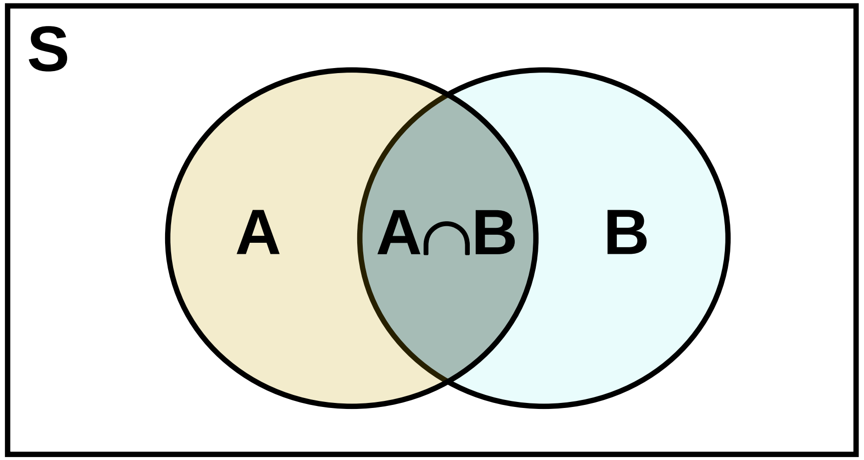
Probability: Some Terms

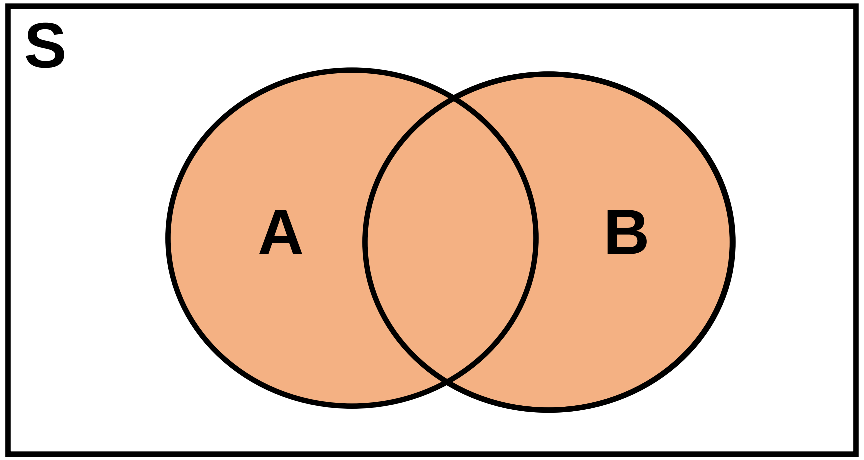
- **Intersection** of events
  - Suppose a sample space S, has two events A and B, then the intersection of A and B (i.e., A∩B) is the **overlapping outcomes** that belong to both A and B
- **Union** of events
  - Suppose a sample space S, has two events A and B, then the union of A and B (i.e., A∪B) is the outcomes that belong to either A or B
  - Additive Law:

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

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

Venn Diagram





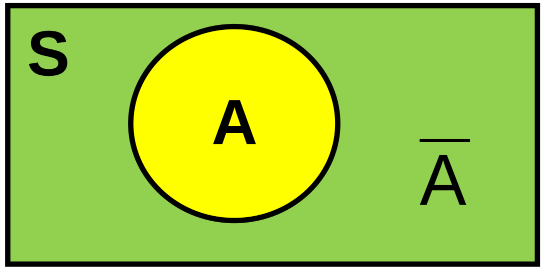
Probability: Some Terms

- **Complement** of an event
  - The complement of event A (i.e., $\bar{A}$), is the outcomes that do not belong to A

$$A \cup \bar{A} = S \Leftrightarrow \Pr(A \cup \bar{A}) = \Pr(A) + \Pr(\bar{A}) = 1$$

$$A \cap \bar{A} = \varnothing \Leftrightarrow \Pr(A \cap \bar{A}) = 0$$

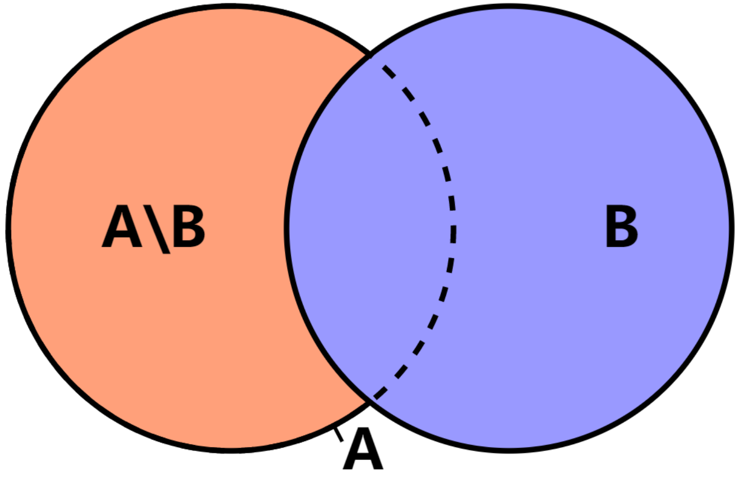
Venn Diagram



- **Difference** of events
  - The difference of A and B (i.e., A\B or A-B) is the outcomes that belong to A but not B

$$A \setminus B = A - B$$

$$\Pr(A \setminus B) = \Pr(A) - \Pr(A \cap B)$$



Probability: Some Terms

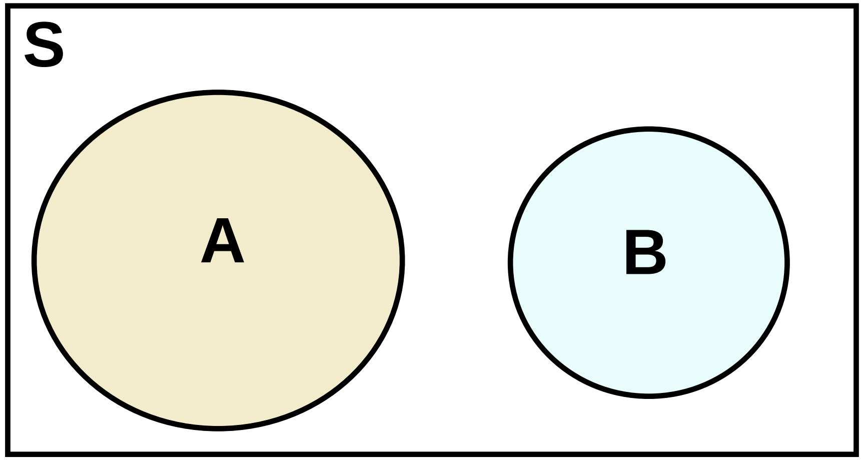
- **Mutually Exclusive** events
  - A and B are mutually exclusive events if they have **no basic outcomes** in common (i.e., A∩B=∅, which is an empty set)

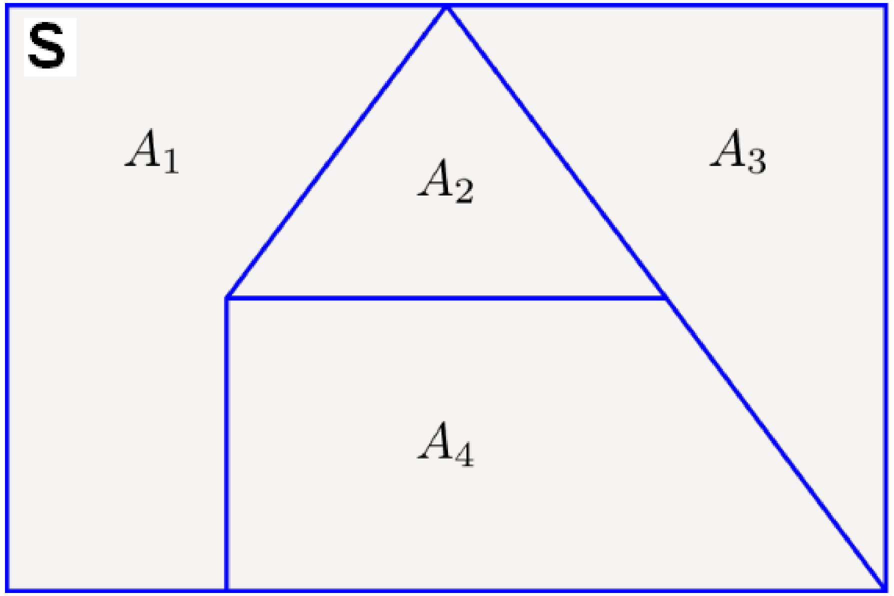
$$A \cap B = \varnothing \Leftrightarrow \Pr(A \cap B) = 0$$

- **Exhaustive** events
  - A_i (i=1,…,4) are collectively exhaustive events if the union of them equals to the whole sample space S
  - A_i (i=1,…,4) are mutually exclusive and exhaustive events if A and B are mutually exclusive and collectively exhaustive events

$$\Pr(A_i \cap A_j) = 0 \text{ for } i \neq j, \text{ and } \Pr(\bigcup_{i=1}^{4} A_i) = 1$$

Venn Diagram





Probability: Example of Roll Dice

Event A: Observe an Odd number

Event B: Observe a number greater than 2

Event C: Observe a 2

Event D: Observe either 3 or 4

Q1. A and B are mutually exclusive? C and D are mutually exclusive?

Q2. $Pr(A)=?$ $Pr(B)=?$ $Pr(C)=?$ $Pr(D)=?$

Q3. $Pr(A \cap B)=?$ $Pr(A \cap C)=?$ $Pr(A \cap D)=?$

Q4. $Pr(A \cup B)=?$ $Pr(A \cup C)=?$ $Pr(A \cup D)=?$

Q5. $Pr(A \backslash B)=?$ $Pr(A \backslash C)=?$ $Pr(A \backslash D)=?$

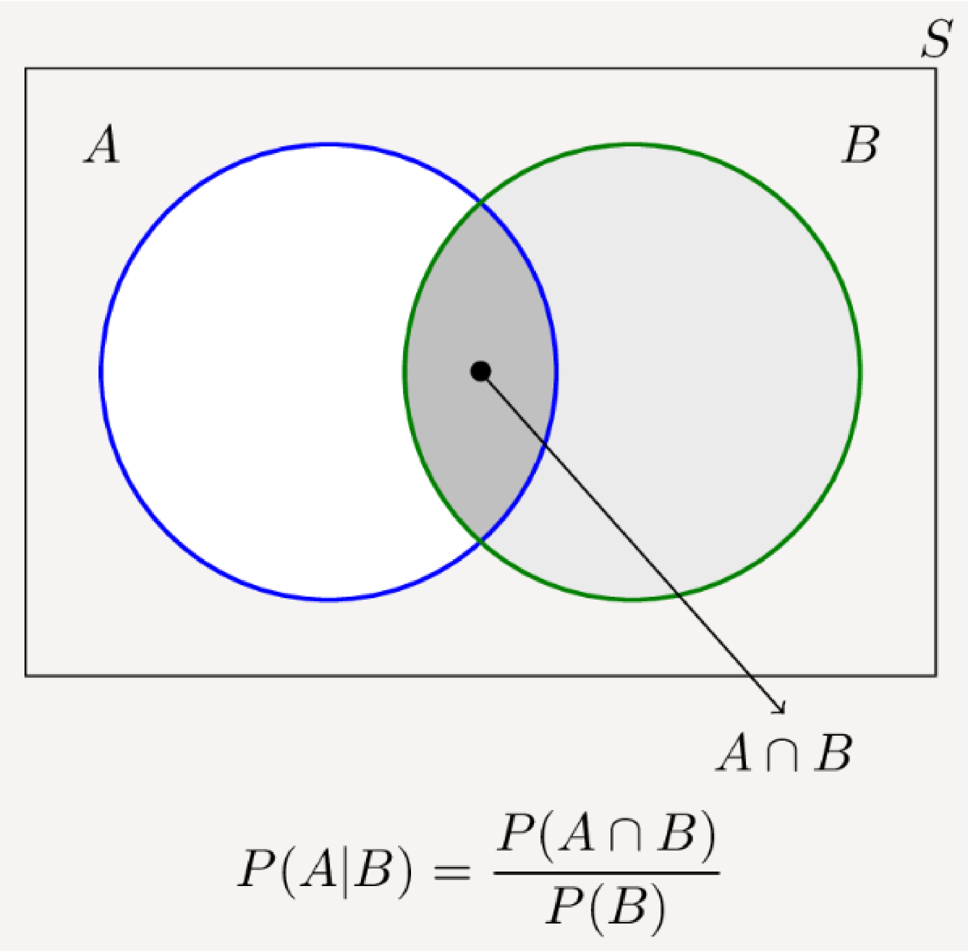
Conditional Probability

- If two events, A and B, are in a sample space S, then the conditional probability of A given B is defined as:

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}, \text{ when } \Pr(B) > 0$$

- It means the probability that A occurs, given that B has already occurred
- Your original sample space S is now shrunk to the area of B

Venn Diagram



Conditional Probability: Roll Two Dice

- Roll two fair dice:
- Event A: 1st red die shows 1
- Event B: 2nd green die shows 2
- Event C: 1st red die shows an even number

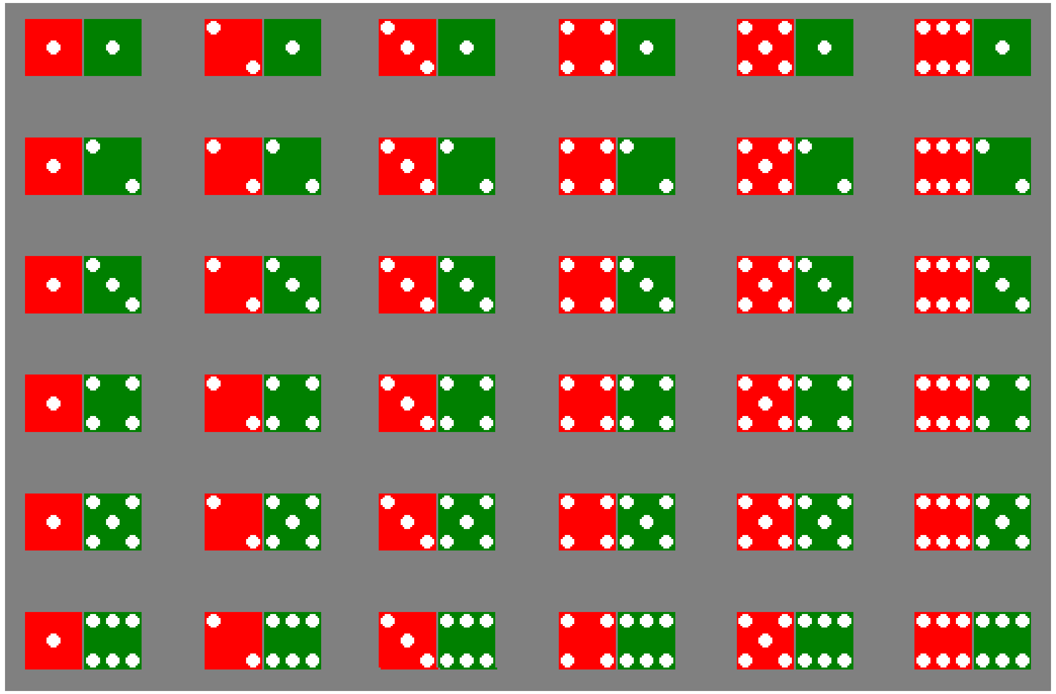
Q1. $\Pr(A|B) = \Pr(A \cap B)/\Pr(B)$

$= (1/36)/(1/6) = 1/6$

Q2. $\Pr(B|C) = \Pr(B \cap C)/\Pr(C)$

$= (3/36)/(3/6) = 1/6$

Sample Space of Two Dice



$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}, \text{ when } \Pr(B) > 0$$

Probability: Dependence and Independence

- Two events A and B are **independent** if: $\Pr(A \cap B) = \Pr(A) \times \Pr(B)$
- It means the occurrence of event A has **nothing to do with** the occurrence of event B, and vice versa
- We can further derive the equation:

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\Pr(A) \times \Pr(B)}{\Pr(B)} = \Pr(A) \text{ and } \Pr(B \mid A) = \Pr(B)$$

- It implies that the event B **does not convey information** to event A, or the occurrence of B **does not influence** the occurrence of A, and vice versa
  - Example:
  - A: It will rain tomorrow; B: I flip a coin and get Head
  - A and B are independent

Probability: Dependence and Independence

- Two events A and B are dependent if: $\Pr(A \cap B) \neq \Pr(A) \times \Pr(B)$
- We can further derive the equation:

$$\Pr(A \mid B) \neq \Pr(A) \text{ or } \Pr(B \mid A) \neq \Pr(B)$$

- It means the occurrence of event A **is influenced by** the occurrence of event B, or the occurrence of event B is influenced by the occurrence of event A
  - Example: I have 10 credit cards in my pocket, 5 DBS and 5 OCBC cards
  - Event A: First time, I randomly pick one and it is DBS card; **Suppose I do not put the first card back to my pocket;**
  - Event B: Second time, I randomly pick one and it is DBS card
  - $\Pr(A \cap B) = (5/10) \times (4/9)$
  - $\Pr(A) = 5/10$ and $\Pr(B) = 5/10$
  - $\Pr(A \cap B) \neq \Pr(A) \times \Pr(B)$ or $\Pr(B \mid A) \neq \Pr(B)$

Probability: Law of Total Probability

- If $B_1, B_2, \ldots B_n$, is a partition (i.e., mutually exclusive and exhaustive) of the Sample Space S, then for any event A we have:

$$\Pr(A) = \sum_{i=1}^{n} \Pr(A \cap B_i) = \sum_{i=1}^{n} \Pr(A \mid B_i) \Pr(B_i)$$

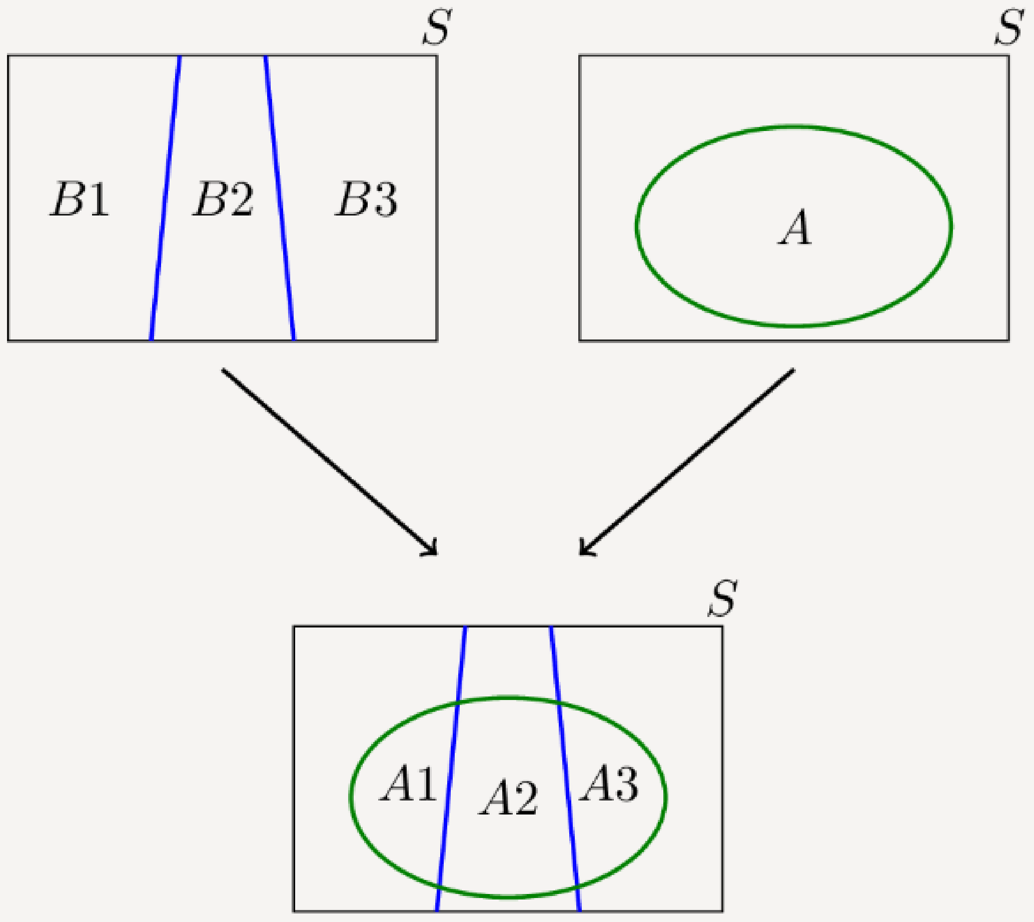
- In the right-side example:

$$\Pr(A_i) = \Pr(A \cap B_i) \text{ for } i = 1, 2, 3$$

$$\Pr(A) = \sum_{i=1}^{3} \Pr(A_i) = \sum_{i=1}^{3} \Pr(A \cap B_i)$$

$$= \sum_{i=1}^{3} \Pr(A \mid B_i) \Pr(B_i)$$

Venn Diagram



Probability: Bayes Rule/Formula

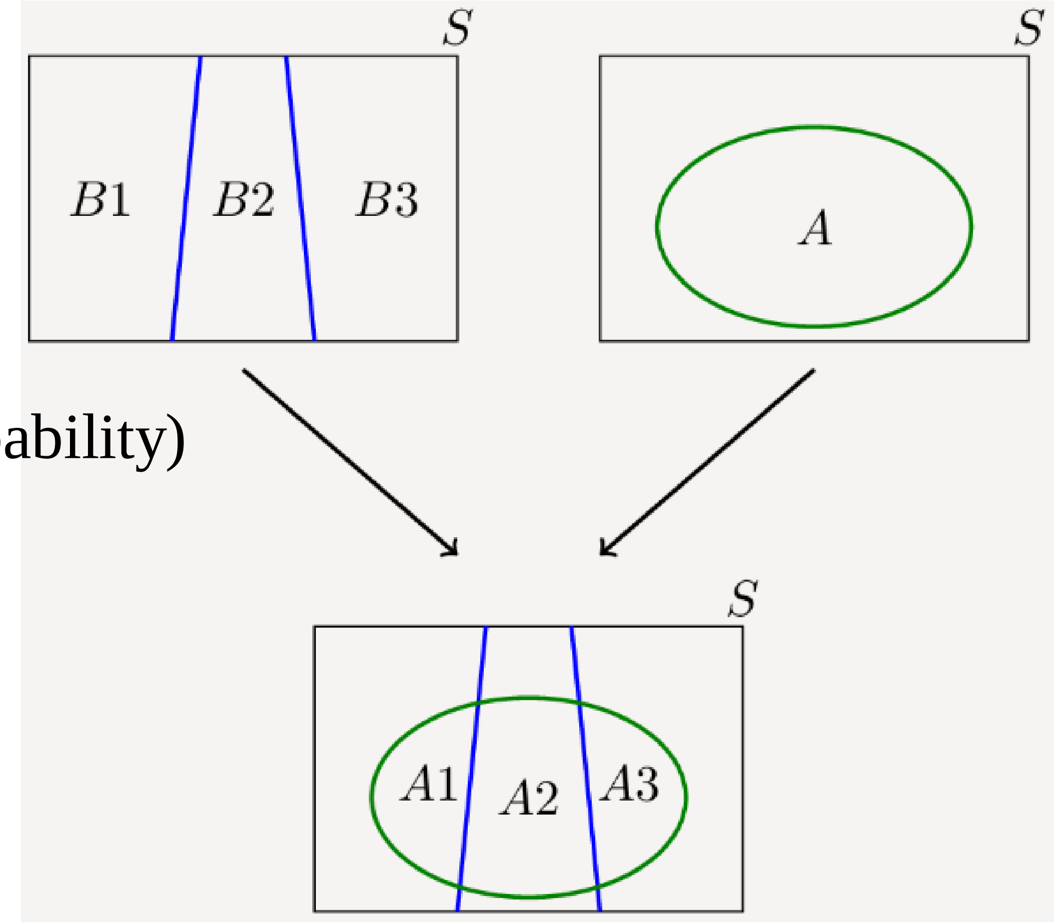
- If $B_1, B_2, \ldots B_n$, is a partition (i.e., mutually exclusive and exhaustive) of the Sample Space S, and A is any event with $\Pr(A) \neq 0$, then we have:

$$\Pr(B_j \mid A) = \frac{\Pr(B_j \cap A)}{\Pr(A)} \qquad \text{(Conditional Probability)}$$

$$= \frac{\Pr(A \mid B_j)\Pr(B_j)}{\sum_{i=1}^{n} \Pr(A \mid B_i)\Pr(B_i)} \qquad \text{(Law of Total Probability)}$$

- Implications:
  - When Prior Belief on an event B_j: $\Pr(B_j)$ meets with a new event A, then the initial belief on B_j will be updated into Posterior Belief $\Pr(B_j \mid A)$
  - This is the foundation of **Bayesian Statistics**

Venn Diagram



Example 1: Medical Test

- Suppose one disease infects 1 out of every 1,000 people in a population. And there is a rapid screening test:
- If a person is infected by the disease, the test comes back positive 99.98% of the time.
- However, the test also produces some false positives: 0.005% of uninfected people are also test positive.
- If one person just tested positive, what is the chance of having this disease?

Event A: This person has HIV

Event B: Test outcome is positive

Question: Pr(A|B)=?

Condition 1: Pr(A)=0.001

Condition 2: Pr(B|A)=0.9998

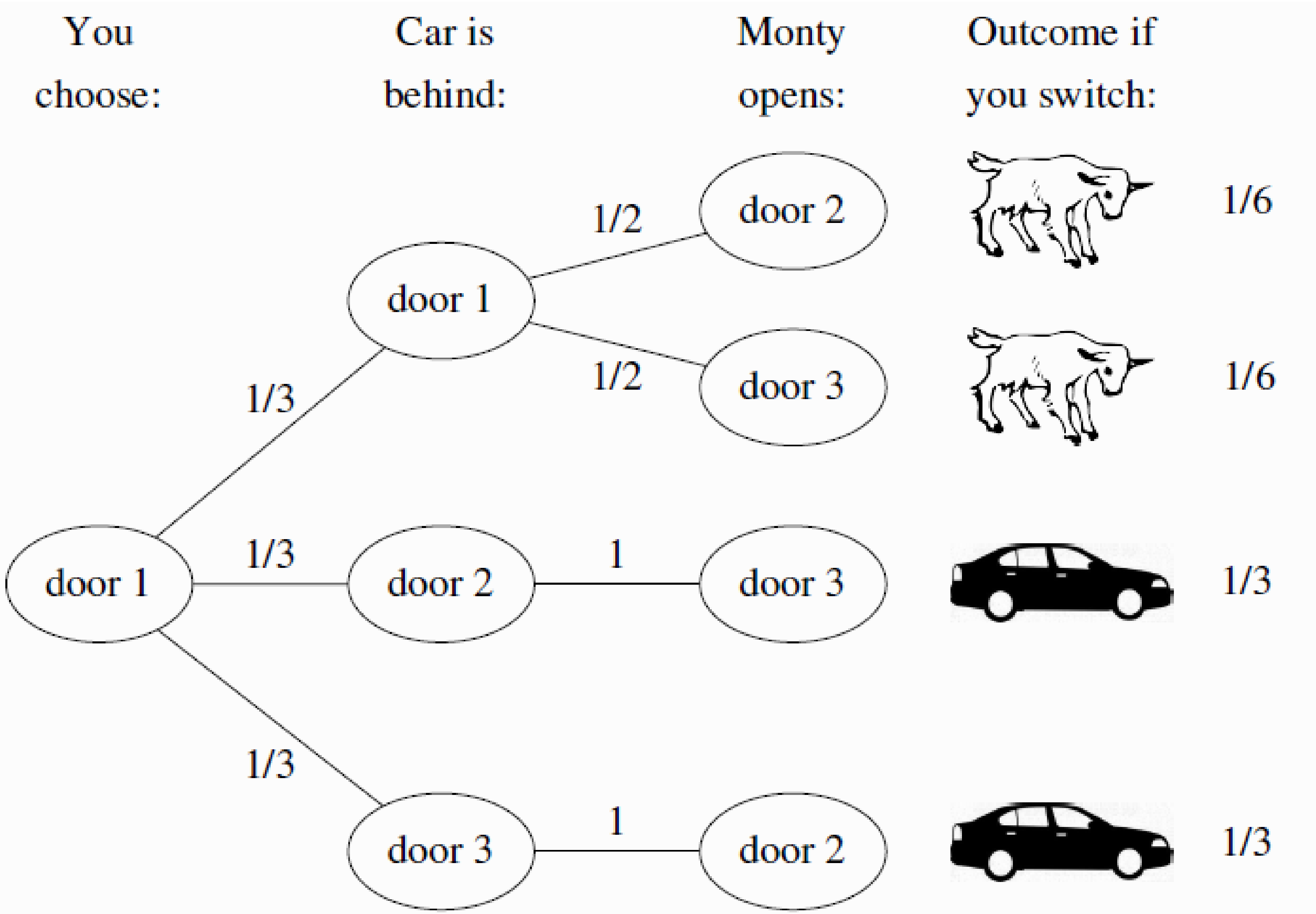
Condition 3: $\Pr(B|\bar{A})=5\times10^{-5}$

$$\Pr(A\,|\,B)$$
$$=\frac{\Pr(A)\times\Pr(B\,|\,A)}{\Pr(A)\times\Pr(B\,|\,A)+\Pr(\bar{A})\times\Pr(B\,|\,\bar{A})}$$
$$=\frac{0.001\times0.9998}{0.001\times0.9998+0.999\times5\times10^{-5}}\approx 95.24\%$$

Example 2: Monty Hall Problem

- On the game show **<Let's Make a Deal>** that is hosted by Monty Hall, a player is asked to choose one of three closed doors, two of which have a goat behind them and one of which has a car.
- The host, Monty Hall, knows where the car is, and then opens one of the two remaining doors. The door he opens always has a goat behind it (i.e., he never reveals the car).
- Monty Hall offers the player the option of switching to the other unopened door. **If the player's goal is to get the car, should s/he switch doors?**

Example 2: Monty Hall Problem



Example 2: Monty Hall Problem

- Use **Bayes Rule** to solve this problem:
  - Event A: Car is behind door 1
  - Event B: Car is behind door 2
  - Event C: Car is behind door 3
  - Event M_i: Monty Hall opens door i (i=1,2,3)
  - **Without loss of generality, suppose the contestant picks door 1**

$$\Pr(A)=\Pr(B)=\Pr(C)=\frac{1}{3}$$

$$\Pr(M_2\,|\,A)=\frac{1}{2}\ \text{and}$$

$$\Pr(M_2\,|\,B)=0\ \text{and}$$

$$\Pr(M_2\,|\,C)=1$$

$$\Pr(M_2\cap A)=\Pr(M_2\,|\,A)\times\Pr(A)=\frac{1}{2}\times\frac{1}{3}=\frac{1}{6}$$

$$\Pr(M_2\cap B)=\Pr(M_2\,|\,B)\times\Pr(B)=0\times\frac{1}{3}=0$$

$$\Pr(M_2\cap C)=\Pr(M_2\,|\,C)\times\Pr(C)=1\times\frac{1}{3}=\frac{1}{3}$$

$$\Rightarrow$$

$$\Pr(M_2)=\Pr(M_2\cap A)+\Pr(M_2\cap B)+\Pr(M_2\cap B)$$

$$=\frac{1}{6}+0+\frac{1}{3}=\frac{1}{2}$$

Example 2: Monty Hall Problem

- Use **Bayes Rule** to solve this problem:
  - Event A: Car is behind door 1
  - Event B: Car is behind door 2
  - Event C: Car is behind door 3
  - Event M_i: Monty Hall opens door i (i=1,2,3)
  - **Without loss of generality, suppose the contestant picks door 1**

$$\Pr(A) = \Pr(B) = \Pr(C) = \frac{1}{3}$$

$$\Pr(M_2 \mid A) = \frac{1}{2} \text{ and}$$

$$\Pr(M_2 \mid B) = 0 \text{ and}$$

$$\Pr(M_2 \mid C) = 1$$

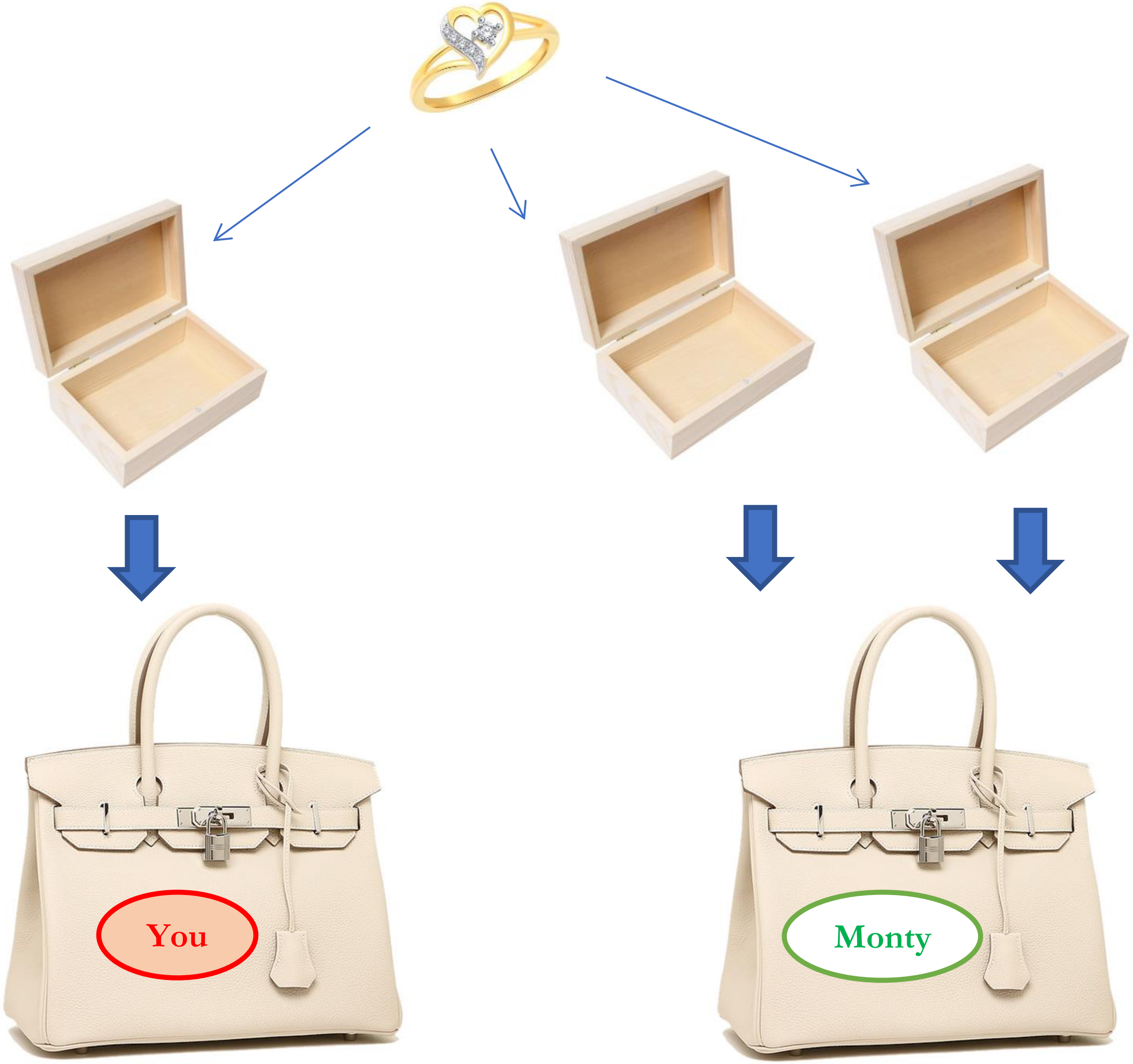
$$\Pr(A \mid M_2) = \frac{\Pr(A) \times \Pr(M_2 \mid A)}{\Pr(M_2)} = \frac{\frac{1}{3} \times \frac{1}{2}}{\frac{1}{2}} = \frac{1}{3}$$

$$\Pr(C \mid M_2) = \frac{\Pr(C) \times \Pr(M_2 \mid C)}{\Pr(M_2)} = \frac{\frac{1}{3} \times 1}{\frac{1}{2}} = \frac{2}{3}$$

$\Rightarrow$ **You should switch the door !**

Example 2: Monty Hall Problem

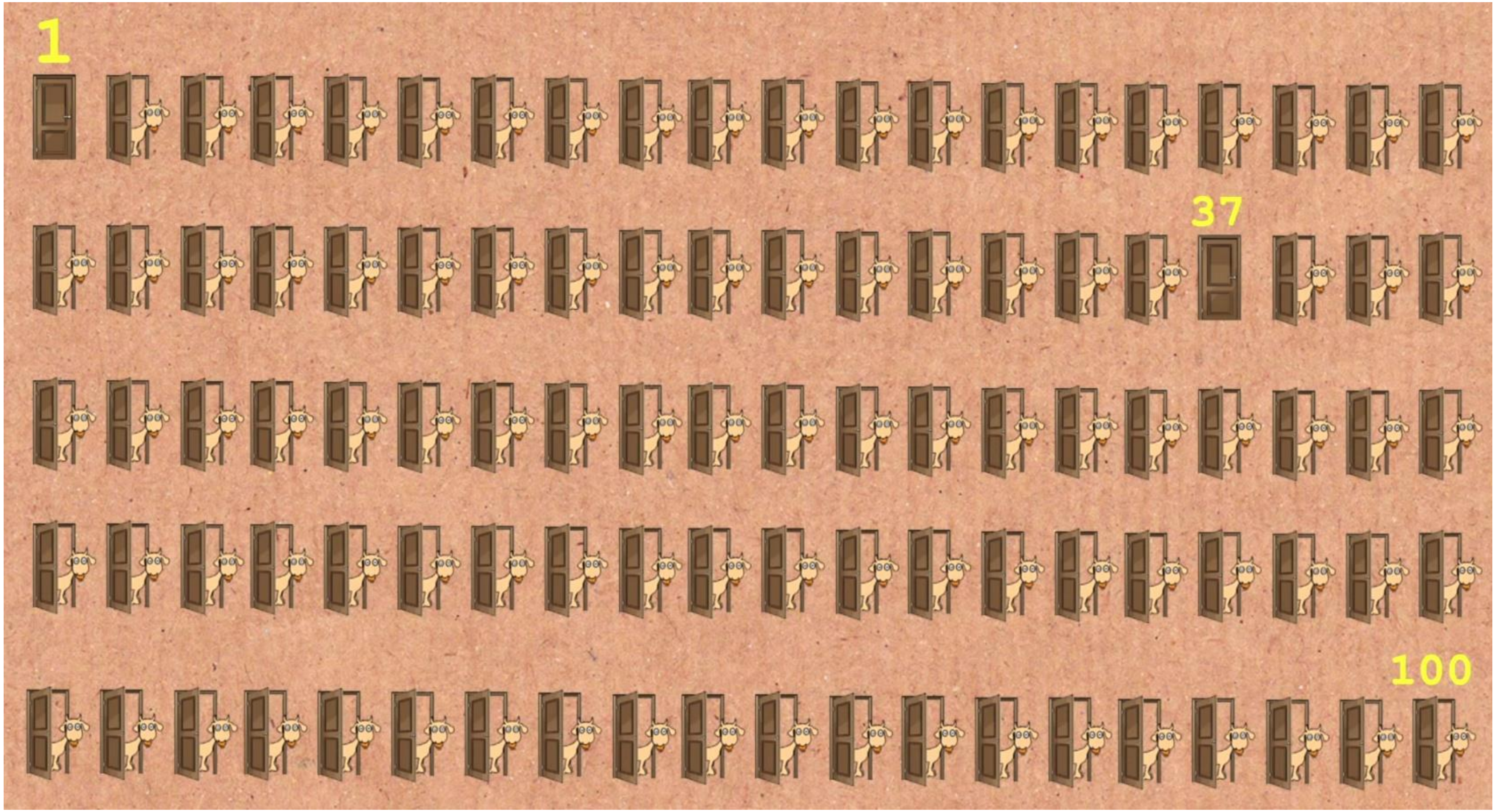
- Suppose there are three boxes, and one of them has a ring, the other two are empty
- You pick one box and put in your bag
- You see that Monty Hall gets another 2 boxes, throws one empty box away, and puts another one into his bag



Will you exchange your bag for his?

Example 2: Monty Hall Problem

If you first pick door 1, and Monty Hall helps you preclude 98 doors with goat, and only leave door 37 unopened

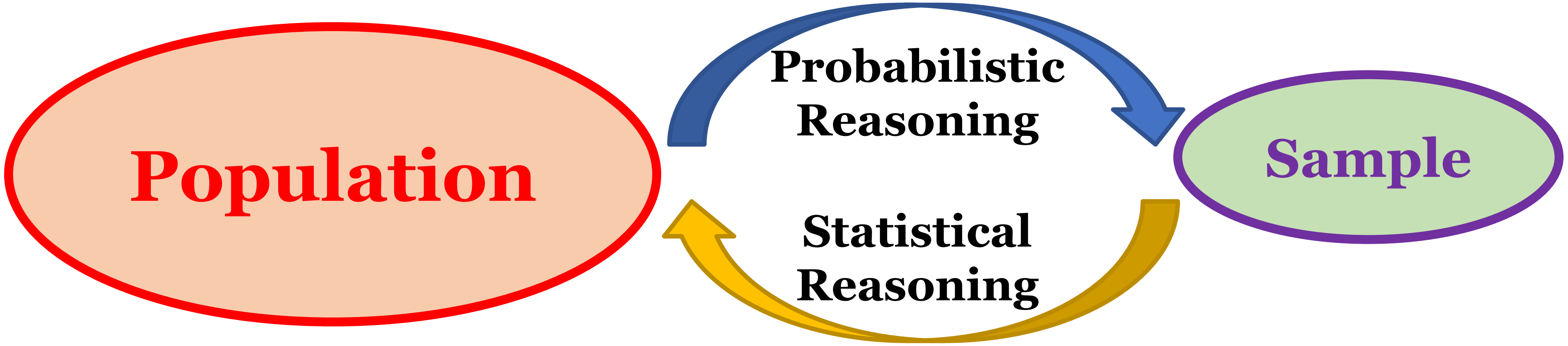


Will you choose door 1 or door 37?

Introduction to Statistics

Overview: Probability and Statistics

- The link between descriptive and inferential tasks:



- Example: Age distribution of people in Singapore
  - I know the overall age distribution, so I can **describe** the proportion of retired people in Clememti
  - I only have a random sample from Clementi, but I can **estimate** the overall proportion of retired people in Singapore

Statistics: Some Terms

- Random Variable: A discrete or continuous quantitative variable X is a **random variable** if its possible values are outcomes of a random experiment/process.
  - Example: (1) X=Number of "Heads" if flipping a coin N times;
  - (2) Y=Body weight of a person who is randomly picked from the whole population
- Probability Distribution For Discrete Random Variable:
  - **Probability Mass Function (PMF)** for a **discrete random variable** X is a function that gives the probability that a discrete random variable is exactly equal to some value x: $\Pr_X(x)=\Pr(X=x)$.
  - PMF must satisfy: $0 \le \Pr_X(x) \le 1$ and $\sum_x \Pr_X(x) = 1$
  - Expectation/Mean: $\mu = E(X) = \sum_x x \times \Pr_X(x)$
  - Variance: $\sigma^2 = Var(X) = E[(X - E(X))^2] = E(X^2) - E^2(X) = \sum_x [x - E(X)]^2 \times \Pr_X(x)$

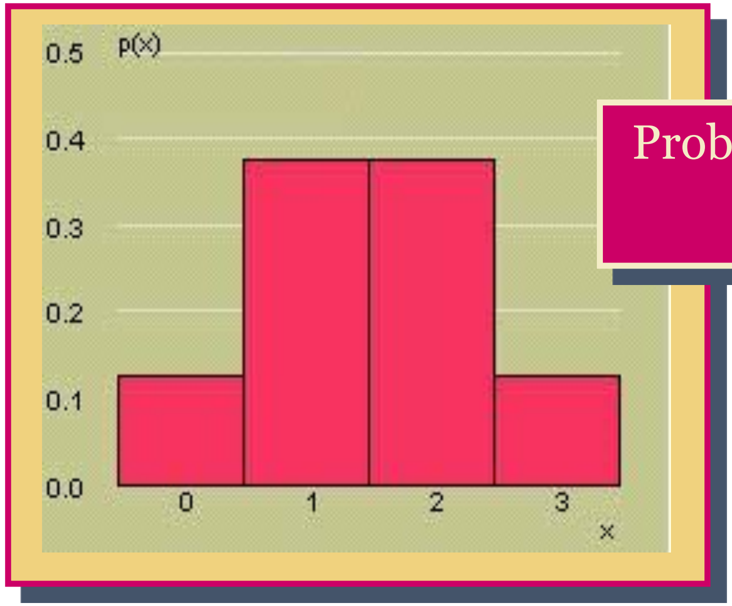
Statistics: PMF For Discrete Variable

- Example: Flip a fair coin 3 times, and define: **X=Number of heads**

HHH	1/8
HHT	1/8
HTH	1/8
THH	1/8
HTT	1/8
THT	1/8
TTH	1/8
TTT	1/8

$$E(X) = \sum_x x \times \Pr_X(x) = 0 + 1\times\frac{3}{8} + 2\times\frac{3}{8} + 3\times\frac{1}{8} = \frac{3}{2}$$

x	Pr(X=x)
0	1/8
1	3/8
2	3/8
3	1/8



Probability Histogram for x

Statistics: Some Terms

- Probability Distribution For Continuous Random Variable:
  - **Cumulative Distribution Function (CDF):** The CDF of a random variable X (or just distribution function of X evaluated at x), is the probability that X will take a value less than or equal to x: $F_X(x)=\Pr(X\le x)$
  - **Continuous random variable** is a random variable whose CDF is continuous everywhere. There are no "gaps" or "jumps". Instead, continuous random variables almost never take an exact prescribed value: $\Pr_X(x)=\Pr(X=x)=0$
  - **Probability Density Function (PDF):** The density (similar to relative frequency/likelihood) of a random variable X (evaluated at x): $f_X(x)$ is defined as:

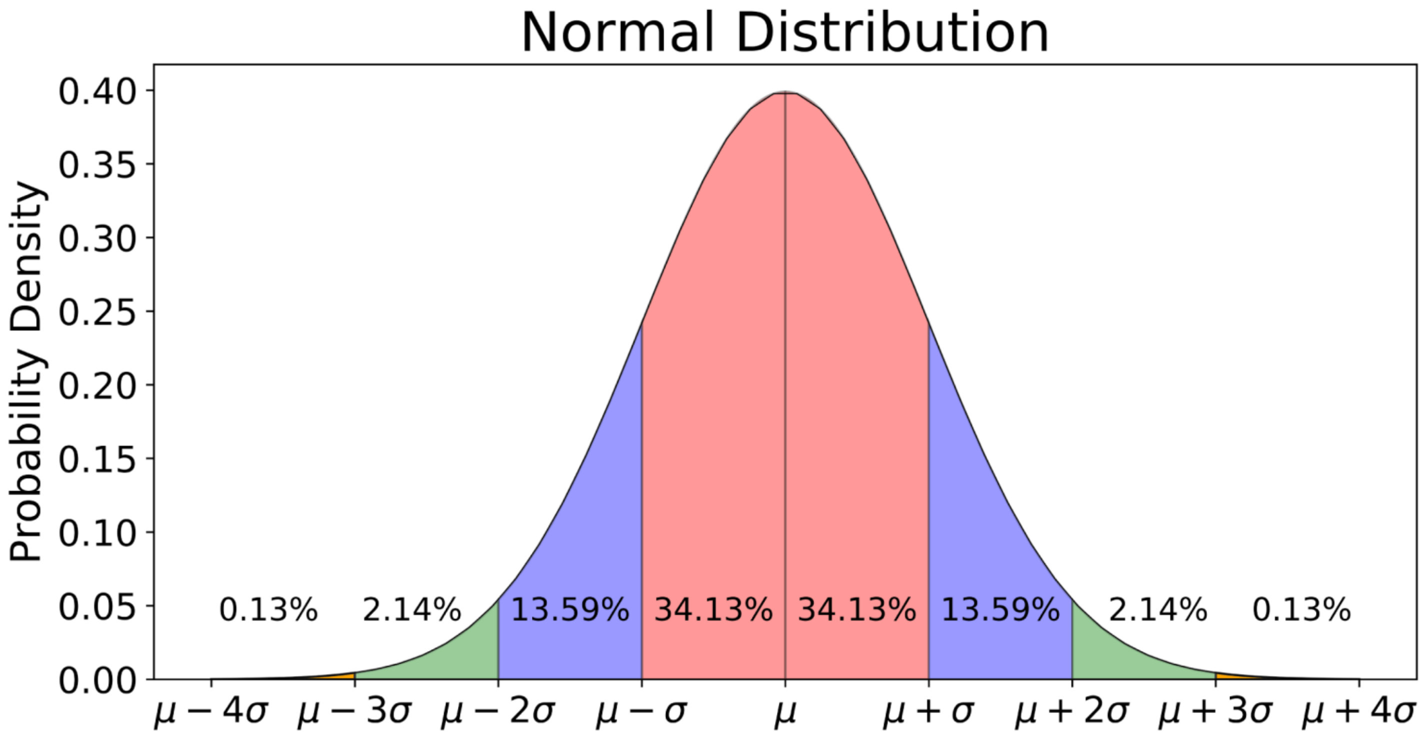
$$f_X(x) = \lim_{\Delta\to 0^+} \frac{\Pr(x < X \le x+\Delta)}{\Delta} = \lim_{\Delta\to 0^+} \frac{F_X(x+\Delta) - F_X(x)}{\Delta}$$

$$= \frac{dF_X(x)}{dx} = F_X'(x) \text{ if } F_X(x) \text{ is differentiable at } x$$

Mean: $\mu=\mathrm{E}(X)=\int_{-\infty}^{\infty} x f_X(x)dx$ and Variance: $\sigma^2=\mathrm{Var}(X)=\int_{-\infty}^{\infty} [x-\mu]^2 f_X(x)dx$

Statistics: CDF And PDF

- Example: Company's Net Profit in the next month (e.g., Normal distribution or Gaussian distribution)



PDF: $f_X(x \mid \mu, \sigma^2) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

CDF (Area under density curve):

$$F_X(x) = \Pr(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

Statistics: Law Of Large Numbers

- Law of Large Numbers (LLN):
- Draw a random sample of n observations from any population (i.e., a sequence of independent and identically distributed i.i.d. random variables) with finite mean μ.
- As the number of observations n increases, the **sample mean** (of the observed values) $\bar{x}_n$ gets **closer and closer** to the **population mean** μ (i.e., sample mean **converges in probability** to the population mean).

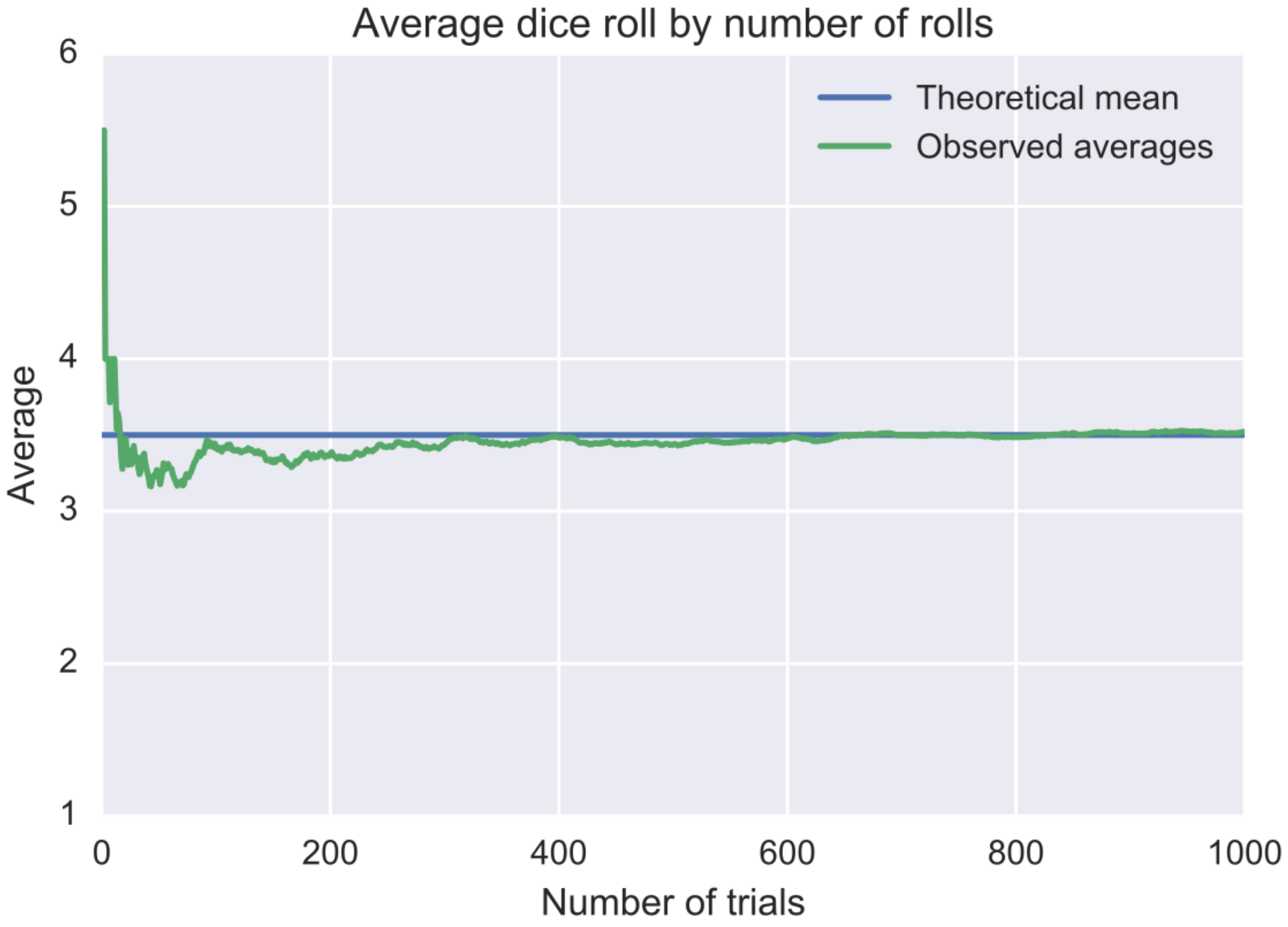
$$\bar{x}_n = \frac{1}{n}(x_1 + \cdots + x_n) \xrightarrow{P} \mu \text{ when } n \to \infty \text{ and } \mu < \infty$$

- This means:

$$\forall \varepsilon > 0: \lim_{n \to \infty} \Pr(|\bar{x}_n - \mu| > \varepsilon) = 0$$

Statistics: Law Of Large Numbers

- Example: Rolling A Single Dice and Get Average Value



As the number of rolls increases, the average of the values of all the results will approach 3.5, which is equal to the theoretical mean value:

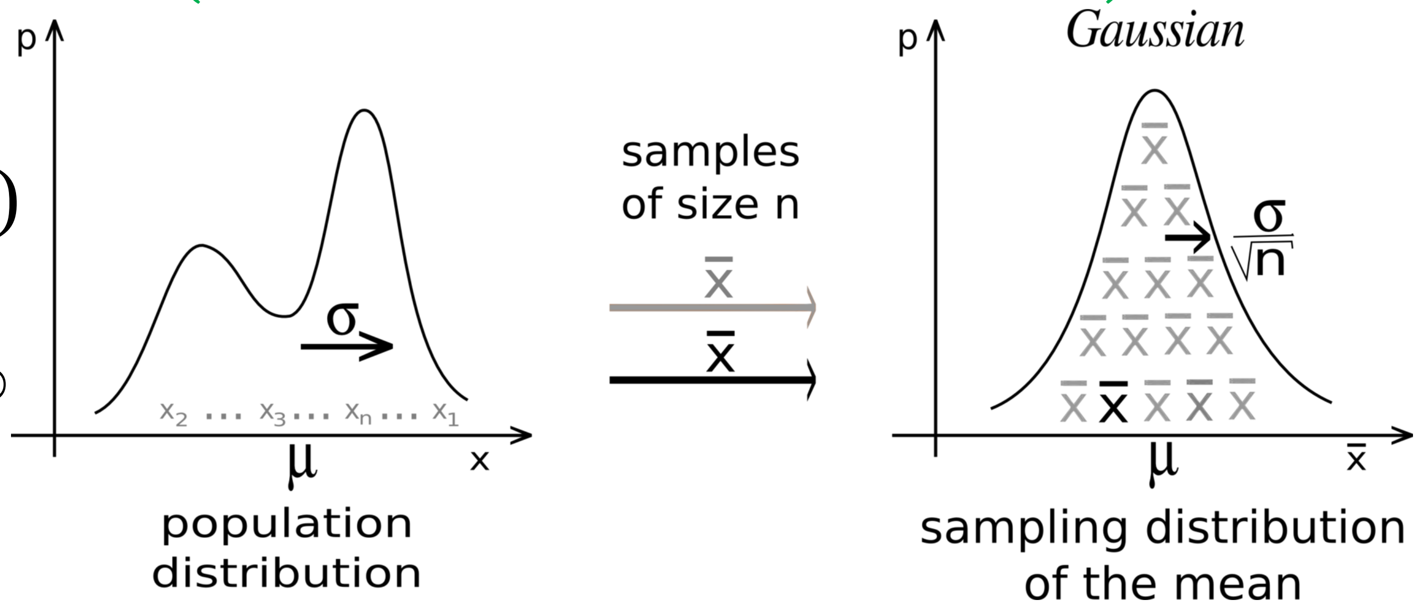
(1+2+3+4+5+6)/6=3.5

Statistics: Central Limit Theorem

- Central Limit Theorem (CLT):
- Classical CLT: Draw a random sample of n observations from any population (i.e., a sequence of independent and identically distributed i.i.d. random variables) with finite mean μ and finite standard deviation σ.
- When n is large, the distribution of the sample mean $\bar{x}_n$ **converges in distribution to Normal Distribution (a.k.a., Gaussian Distribution)**

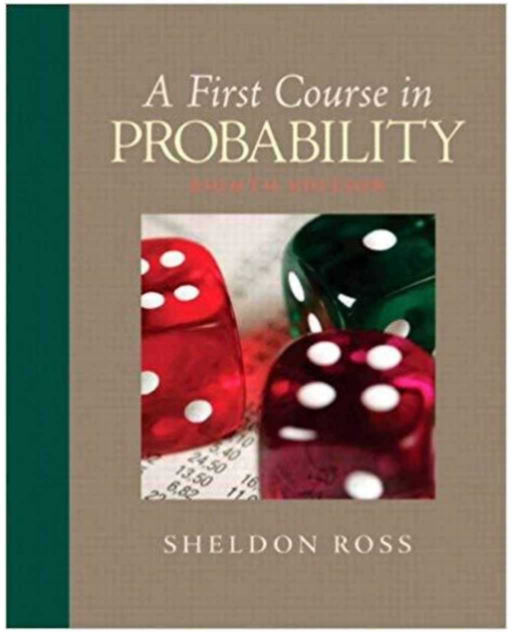
$$\bar{x}_n = \frac{1}{n}(x_1 + \cdots + x_n) \xrightarrow{d} N(\mu, \frac{\sigma}{\sqrt{n}})$$

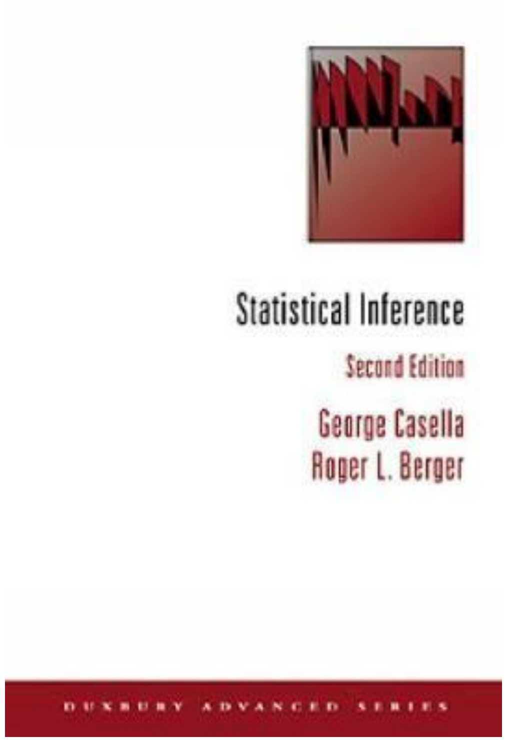
when $n \to \infty$ and $\mu < \infty$ and $\sigma^2 < \infty$



References

- Concepts of Probability
  - http://julio.staff.ipb.ac.id/files/2015/02/Ross_8th_ed_English.pdf
  - http://ocw.uci.edu/courses/math_131a_introduction_to_probability_and_statistics.html
- Concepts of Statistics
  - https://www.math.uh.edu/~zhenwang/Teaching/Math%202433/G.%20Casella,%20R.%20L.%20Berger-Statistical%20Inference.pdf
  - https://youtu.be/VPZD_aij8H0





Any Questions ?

Thank You !

Appendix: Conditional Independence

- Two events A and B are conditionally independent given an event C with Pr(C)>0 if:

$$\Pr(A \cap B \mid C) = \Pr(A \mid C) \times \Pr(B \mid C)$$

- If Pr(B)>0, by conditioning on event C, the conditional probability Pr(A|B) is:

$$\Pr(A \mid B, C) = \frac{\Pr(A \cap B \mid C)}{\Pr(B \mid C)}$$

- If Pr(B|C)>0 and Pr(C)>0, and If A and B are conditionally independent given C, we have:

$$\Pr(A \mid B, C) = \frac{\Pr(A \cap B \mid C)}{\Pr(B \mid C)} = \frac{\Pr(A \mid C)\Pr(B \mid C)}{\Pr(B \mid C)} = \Pr(A \mid C)$$

Appendix: List Of Probability Distributions

- http://en.statmania.info/2015/08/list-of-probability-distribution.html
- https://en.wikipedia.org/wiki/List_of_probability_distributions

Distribution	p.d.f	Mean	Variance	Limit
Uniform	$\frac{1}{k}$	$\frac{k+1}{2}$	$\frac{k^2-1}{12}$	k = 1, 2, ..., n
Bernoulli	$p^x(1-p)^{1-x}$	p	pq = p(1-p)	0, 1
Binomial	$n_{C_x} p^x q^{n-x}$	np	Npq	(0,n)
Hypergeometric	$\frac{M_{C_x} N - M_{C_{n-x}}}{N_{C_n}}$	$n.\frac{M}{N}$	$\frac{N-n}{N-1}.n.\frac{M}{N}(1-\frac{M}{N})$	
Poisson	$\frac{e^{-\lambda}\lambda^x}{x!}$	λ	λ	(0,∞)
Geometric	$P(1-p)^{x-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	(0,∞)
Negative Binomial	$\binom{x-1}{r-1} p^r 1 - p^{x-r}$	$\frac{r}{p}$	$\frac{rq}{p^2}$	(r,∞)
Normal	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)}$	μ	σ^2	(-∞,∞)
Standard Normal	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$	0	1	(-∞,∞)
Exponential	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	X, λ > 0
Gamma	$\frac{\lambda^\alpha}{\Gamma n} x^{\alpha-1} e^{-\lambda x}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$	x>0
Beta	$\frac{1}{\beta(a,b)} e^{a-1}(1-x)^{b-1}$	$\frac{a}{a+b}$	$\frac{ab}{(a+b+1)(a+b)}$	0, 1
Chi-squared	$\frac{1}{\Gamma(\frac{k}{2})}(\frac{1}{2})^{\frac{k}{2}} x^{\frac{k}{2}-1} e^{-\frac{1}{2}x}$	k	2k	(0,∞)