Descriptive Statistics

Mean
Median
Mode

Standard deviation
Variance
Range

Inferential Statistics

Probability.

Likelihood of an event

- Laws of Probability → Conditional Prob- Bayes' Theorem
- Random Variables
- Probability Mass function
- Counting the number of events.

Laws of Probability:

Probability theory studies assignments of prob. to "Events".

Events are statements like propositions for which it is not immediately clear whether they are true or false. eg:

↳ It will rain in two weeks from now.

↳ India's population will take over china's by 2022.

↳ '3' shows up when you roll a fair die.

Algebra of Events:

$A \cup B$ → is the event that at least one of A and B will occur

≤2 {1,2} even {2,4,6} → {1,2,4,6}

$A \cap B$ → is the event that both A and B will occur

{2}

A^c → is the event that A will not occur.

↳ {3,4,5,6}

We also define an event Ω which is the union of all possible events; $\emptyset$ is the intersection of all possible event.

$A \cap A^c = \phi$ ↳ Empty set

$\Omega^c = \phi$

$A \cup A^c = \Omega$

Rolling of a fair die.

$A = \{1, 2\}$ Numbers less than or equal to 2 show up.

$B = \{2, 4, 6\}$ Even Numbers show up.

$P(A) = 1/3 = 1/6 + 1/6$

$P(B) = 1/2$

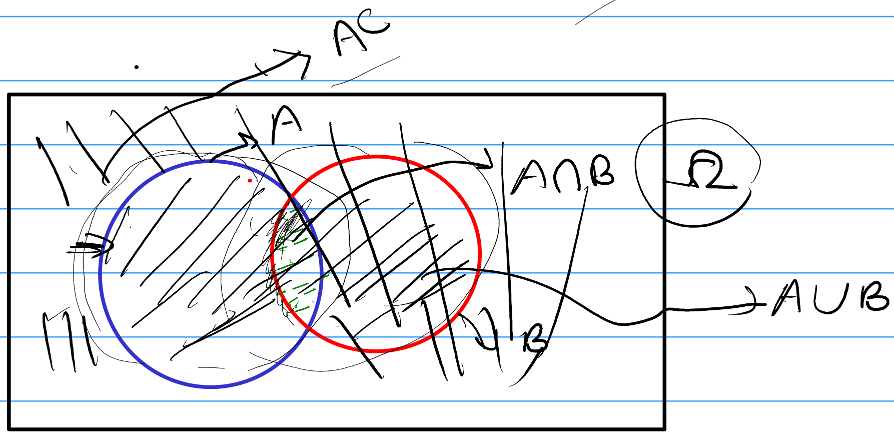
$P(A \cap B) = P(\{2\}) = 1/6$

$P(A \cup B) = P(\{1, 2, 4, 6\}) = 2/3$

$P(A) + P(B) = \frac{5}{6}$

$- P(A \cap B) = 2/3 = P(A \cup B)$

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



Laws of Probability:

$P(A)$ is a real number b/w 0 and 1; the prob. of (occurrence of) the event A.

- Prob. of the universal event is 1, i.e., $P(\Omega) = 1$
- If $A \subset B$, then $P(A) \leq P(B)$

$\{1, 2\} \subset \{1, 2, 3\}$

$1/3$ $\quad$ $1/2$

- The "excluded middle" law, $P(A) + P(A^c) = 1$

- The addition law:

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

Ex: Traffic lights at an intersection cycle through every 75 seconds.

↳ Green → 30 → $\frac{30}{75}$

Yellow → 5 → $\frac{5}{75}$

Red → 40 → $\frac{40}{75}$

+ Counting the sizes of various sets.

Ex: Find the total number of possible 8-bit bytes. 00010100

2^8

$2 \; 2 \; 2 \; \ldots$ $\quad 2^8 = 256$

$0 \to 255$

Ex: Find the # of four letter sequences that can be obtained by ordering the letters A, B, C, D without repetitions. ADCB

$\underbrace{4 \cdot 3 \cdot 2 \cdot 1}_{4!} = 24$

Ex: Find the number of unique five-letter sequences that can be formed from the word QUEUE?

$Q U_1 E_1 U_2 E_2$

$5 \; 4 \; \ldots$ $\quad \frac{5! = 120}{2!\,2!} = 30$

Ex: $0! = 1$? $\quad 5! = 5 \times 4 \times 3 \times 2 \times 1$

$$n! = n(n-1)!$$

$$= n\cdot(n-1)\cdot(n-2)\ldots 1 = n(n-1)!$$

$$(n-1)! = \frac{n!}{n}$$

$n=1$

$$\boxed{0! = \frac{1!}{1} = 1}$$

$n=0$ $\quad (-1)! = \frac{0!}{0} \rightarrow$ undefined

We do not define factorials for negative numbers.

Ex: Suppose nine basketball players are labeled by the letters A, B, C, D, E, F, G, H, I

A lineup consists of five players. The order of the players in the lineup is not important

→ # of distinct lineups?

$\overbrace{9 \quad 8 \quad 7 \quad 6 \quad 5}^{5!}$

A B

B A

$\rightarrow 9\cdot8\cdot7\cdot6\cdot5 = \frac{9!}{4!} \rightarrow {}^9P_5 = \frac{9!}{(9-5)!}$

Had the order been important

A B E G I

5 4

$$\frac{9!}{4!\,5!} = 126 \qquad {}^9C_5 = {}^9C_4$$

$$\binom{9}{5} = {}^9C_5$$

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

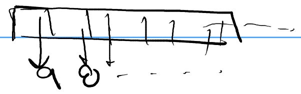
When order is important

$$\boxed{{}^nC_k = \frac{n!}{k!\,(n-k)!} = {}^nC_{n-k}}$$

order is not important

Ex: How many binary seqⁿ of length 9 have exactly 5 ones?

$\hookrightarrow \underline{\left({}^{9}C_5\right)} = \frac{9!}{5!\,4!} = 126$



5 ones, 4 zeros

I_1, I_2, I_3, I_4, I_5 $\quad 0_1, 0_2, 0_3, 0_4$

$$\frac{9!}{5!\,4!} = {}^{9}C_5$$

Ex: Suppose there are 9 socks loose in a drawer in a dark room

6 orange $\quad$ 3 Blue

Someone selects two socks at random. What is the prob. that the two socks are of the same color?

$\hookrightarrow |\Omega| = \binom{9}{2} = {}^{9}C_2 = \frac{9\cdot 8}{2} = 36$

when socks are orange $\binom{6}{2} = {}^{6}C_2 = \frac{6\cdot 5}{2} = 15$

when socks are blue $\binom{3}{2} = 3$

$P(\text{socks are of same color}) = \frac{15+3}{36} = 1/2.$

Alternative Method:

$\hookrightarrow$ Orange $\frac{6}{9} = 2/3$

$\hookrightarrow$ Blue $\frac{3}{9} = 1/3$

$\vec{A}$ is that orange colored sock is picked at first

Sock 1 is orange and Sock 2 is also orange

$\vec{B}$ is the event that we pick orange colored sock in the 2nd draw

$\frac{6}{9} \cdot \frac{5}{8} = \frac{30}{72}$

Sock 1 is Blue and Sock 2 is also blue

$\frac{3}{9} \cdot \frac{2}{8} = \frac{6}{72} \longrightarrow \frac{36}{72} = 1/2.$

* Conditional Prob. and Independence

Two events A and B are said to be independent if

$$\boxed{P(A \cap B) = P(A).P(B)}$$

H , T
$\frac{1}{2}$ $\frac{1}{2}$

$P(\{HH\}) = \frac{1}{2} \cdot \frac{1}{2}$

$A_1, A_2, \dots, A_n$

$$\boxed{P(A_1 \cap A_2 \dots \cap A_n) = P(A_1)\, P(A_2) \dots P(A_n)}$$

"Mutual Independence" A_1, A_2, A_3

$$\begin{cases} P(A_1 \cap A_2 \cap A_3) = P(A_1)\, P(A_2)\, P(A_3) \\ P(A_1 \cap A_2) = P(A_1).P(A_2) \\ P(A_1 \cap A_3) = P(A_1).P(A_3) \\ P(A_2 \cap A_3) = P(A_2)\, P(A_3) \end{cases}$$

A_1, A_2, A_3, A_4

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) P(A_2) P(A_3)$$

- Mutually Exclusive Events:

A and B are mutually exclusive if $\boxed{A \cap B = \phi}$

$A,\ B = A^c$, A and A^c are M.E.

Ω and ϕ are M.E.

$$\boxed{P(A \cup B) = P(A) + P(B)}$$

- Events can also be dependent sometimes, meaning the occurrence of one event is likely to have a bearing on the other event.

↳ Conditional Prob. → A, B with $P(B) \neq 0$

$$\underbrace{P(A|B)}_{\text{conditional prob. of A given B}} \neq P(A)$$

"Knowing that B is true, what is the prob. of occurrence of A"

If $P(B) \neq 0$,

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

A: odd numbers show up

B: even numbers show up

$\frac{1}{2}$ $P(A)$ $\quad$ $1/2$ $P(B)$

$P(A\cap B) = 0$

$P(A|B) = 0$

A: numbers ≤ 2 show up $\to \{1,2\}$ $\qquad \boxed{P(A) = 1/3}$ $\to P(A|\Omega)$

B: even numbers ≤ 4 show up $\{2,4\}$ 50% 50%

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

$A \cap B = \{2\}$

$n(B) = 2$

$n(A\cap B) = 1$

$\boxed{P(A|B) = 1/2}$

- If two events are mutually exclusive, then $P(A|B) = 0$
- If two events are independent, $P(A\cap B) = P(A)\cdot P(B)$

$\boxed{P(A|B) = P(A)}$

* Bayes' theorem: $P(A) \neq 0$, $P(B) \neq 0$

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

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

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

y_T

$\hat{x}_1, \dots, \hat{x}_{T-1}$

$\{x_T\} \longrightarrow \{y_1, \dots, y_T\}$ $\quad$ P(rel. | pos. measure)

Ex: What is the prob. that one of the dice shows 2 when the sum is 7. (we roll two fair dice)

$n(B) = \underbrace{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)}_{6}$

$n(A\cap B) = 2$ $\qquad P(A|B) = 2/6 = 1/3$

$A = \{(2,1), (2,2), (2,3), \ldots, (2,6)$
$(1,2), (3,2), (4,2), (5,2), (6,2)\}$

$n(A) = 11$

$p(A) = 11/36$

$P(A|B) = 1/3$

* Non-intuitive example:

$\xrightarrow{} 1/1000 = .001$

For every 1000 persons being tested, only one traveler is expected to be infected. The RT-PCR test has a false positive in 5% cases, in which the traveler is healthy & uninfected.
But the test is highly sensitive, and never fails to detect a truly infected person. X returned with a positive infection in the test. What is the prob. that X was indeed infected?

1/50 $\quad$ 0.95

$$P(\text{infected}) = \frac{1}{1000} = 0.001$$

$$P(\text{infected} \mid \text{pos}) = \frac{P(\text{infected and pos})}{P(\text{pos})}$$

$$= \frac{\overbrace{P(\text{pos} \mid \text{infected})}^{=1}\, P(\text{infected})}{P(\text{pos})} \quad \text{Bayes' Theorem}$$

$$= \frac{0.001}{P(\text{pos})} = \frac{0.001}{0.05095} = 0.019627 \approx 0.02$$

infected $\quad$ uninfected

$$\to P(\text{pos}) = \underbrace{P(\text{pos} \mid \text{infected})}_{=1}\, \underbrace{P(\text{infected})}_{=0.001} + \underbrace{P(\text{pos} \mid \text{uninfected})}_{0.05}\, \underbrace{P(\text{uninfected})}_{=0.999}$$

$$= P(\text{pos} \cap \text{infected}) + P(\text{pos} \cap \text{uninfected})$$

$$= 0.001 + 0.05 \times 0.999$$

$$= 0.05095$$

We can also arrive at the same sol^n using a different approach.

999 healthy   1 infected

$\#$ of positive tests $= 999 \times 0.05 + 1$

$= 50.95$

$\therefore$ Prob. of choosing an infected person $= \frac{1}{50.95} \approx 0.02$

Random Variables

A real valued function on Ω.

$\vec{X_1} \quad \vec{X_2}$

$X \to 1, 2, 3, 4, 5, 6$

Roll two fair dice

$S \to$ sum of numbers showing up

$\hookrightarrow S = 2, \ldots, 12$ $S = X_1 + X_2$

$(1,1) \quad (1,2), (1,3) \ldots$

$2 \quad 3$

H, T 0.5

$X = \begin{cases} 0 & \text{when } H \\ 1 & \text{when } T \end{cases}$

$2, 1, \ldots$

$\bar{X} = \frac{1+2+3+4+5+6}{6} = 3.5$ rand

$X \to$ Discrete-type random variable

$X \cancel{\to} 0.34$

$X \in \{1,2,3,4,5,6\}$ $\cancel{X} \in \{1, 2, 3, \ldots\}$

$1.1, \sqrt{2}, \pi, \ldots$

Discrete type random var.

Prob. Mass. function
Population Mass " (pmf)

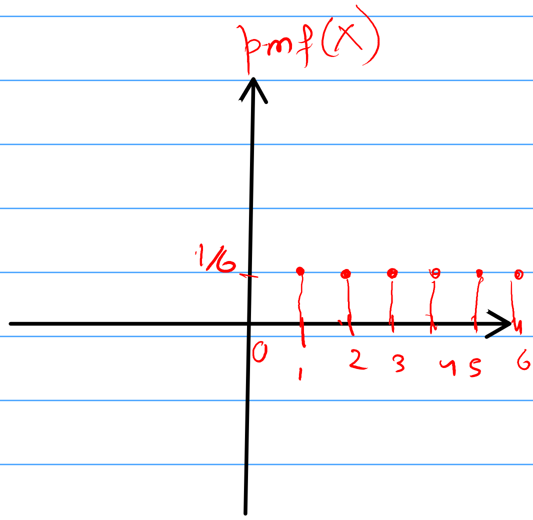
$u \in \{u_1, \ldots, u_n\}$

$P(X = u)$

$\{1, 2, 3, 4, 5, 6\}$

$$\begin{cases} P(X=1) = 1/6 \\ P(X=2) = 1/6 \\ \vdots \\ P(X=6) = 1/6 \end{cases}$$

$P(X=7) = 0$



Ex. Toss a fair coin twice. → H $X=0$, → T $X=1$

Sum of two independent coin tosses. H H

$P(S=0) = \frac{1}{2} \cdot \frac{1}{2} = 0.25$
H H

$P(S=1) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = 0.5$
H T T H $=1$

$P(S=2) = \frac{1}{2} \cdot 1/2 = 0.25$
T T