06-08-2020

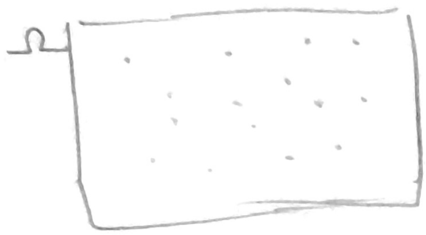
MIT RES. 6 012 Introduction to Probability

* Sample Space:

Any Experiment

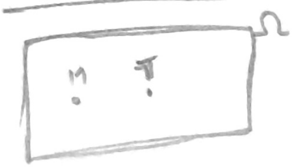
- Two steps
  - Describe possible outcomes
  - Describe beliefs about likelihood of outcomes

- Set of possible outcomes Ω



- Set must be:
  - Mutually exclusive
  - Collectively exhaustive
  - At the right granularity.

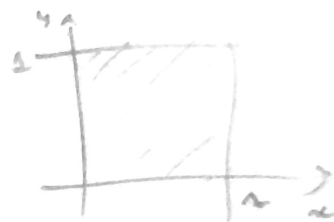
Example: Coin Toss



* Sample space: discrete / finite example
- Two rolls of a tetrahedral die

(1,1) (1,2) (1,3) (1,4)
(2,1) (2,2) (2,3) (2,4)
(3,1) (3,2) (3,3) (3,4)
(4,1) (4,2) (4,3) (4,4)

* Sample space: continous example
- (x,y) such that $0 \le x, y \le 1$



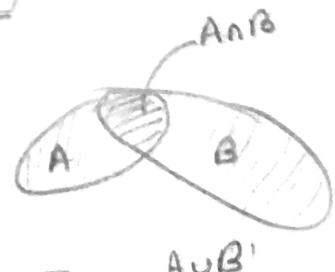
* Probability Axioms:
- Event: a subset of the sample space
  - Probability is assigned to events

P(A)

- Axiom:
  - Non-negativity: $P(A) \ge 0$
  - Normalization: $P(\Omega) = 1$
  - (finite) additivity: (to be strengthened later)

$$\boxed{\text{If } A \cap B = \phi, \text{ then } P(A \cup B) = P(A) + P(B)}$$

(disjoint events)

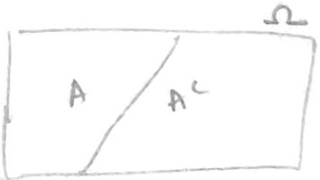


Simple consequences of the axiom:

Axioms	Consequences
(a) $P(A) \geq 0$	$P(A) \leq 1$
(b) $P(\Omega) = 1$	$P(\phi) = 0$
For disjoint events ($A \cap B = \phi$)	$P(A) + P(A^c) = 1$
(c) $P(A \cup B) = P(A) + P(B)$	$P(A \cup B \cup C) = P(A) + P(B) + P(C)$

& similarly for k disjoint events

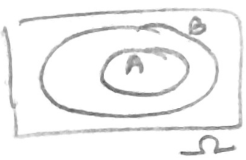
$P(\{s_1, s_2 \dots s_k\}) = P(\{s_1\}) + P(\{s_2\}) \dots P(\{s_k\})$



$A \cup A^c = \Omega$

$A \cap A^c = \phi$

* If $A \subset B$, then $P(A) \leq P(B)$



Proof: $B = A \cup (B \cap A^c)$

$P(B) = P(A) + \underbrace{P(B \cap A^c)}_{\geq 0}$

$P(B) \geq P(A)$

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

$\boxed{P(A \cup B) \leq P(A) + P(B)}$ union bound

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

Examples: Probability calculations: discrete / finite

- Two rolls of tetrahedral die.
- Let every possible outcome have probability of $1/16$
- $P(X = 1) = 4 \cdot \frac{1}{16} = \frac{1}{4}$

Let $Z = \min(X, Y)$

- $P(Z = 4) = 1/16$. $(4,4)$
- $P(Z = 2) = 5/16$ $[(2,2)(2,3)(2,4),(3,2),(4,2)]$

→ Discrete Uniform Law:

- Assume Ω consists of n equally likely elements
- Assume k consists of k elements.

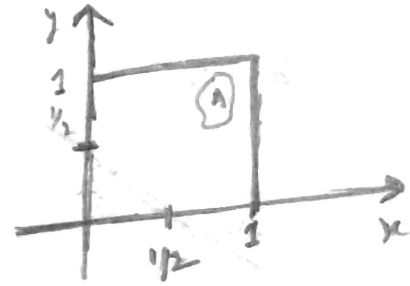
$$\boxed{P(A) = \frac{k}{n}}$$

• Probability calculation: continuous example

• Uniform probability law: Probability = Area.

example • (x,y) such that $0 \le x, y, \le 1$

$$P(\{(x,y) \mid x+y \le 1/2\}) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$$



• Probability calculation: discrete but infinite sample space

• Sample space = $\{1, 2, \ldots\}$

- We are given $P(n) = \frac{1}{2^n}$, $n = 1, 2 \ldots$

• $P(\text{outcome is even}) = P(\{2, 4, 6, \ldots\})$

$= P(\{2\}) + P(\{4\}), + \ldots$

$= P(2) + P(4) + P(6) + \ldots$

$\frac{1}{2^2} + \frac{1}{2^4} + \frac{1}{2^6} + \ldots$

$$\frac{\frac{1}{2^2}}{1 - \frac{1}{2^2}} = \frac{1/4}{3/4} = \frac{1}{3}$$

* **Sets :**
  - A collection of distinct elements.
    $\{a, b, c, d\}$ finite
    $\mathbb{R}$: real number infinite

$x \in S$ or $x \notin S$

$\{x \in \mathbb{R} : \cos(x) > 1/2\}$

S^c

$x \in S^c$ if $x \in \Omega$ & $x \notin S$.

$(S^c)^c = S$

Ω : Universal set

ϕ : empty set $\boxed{\Omega^c = \phi}$

$S \subset T : x \in S \Rightarrow x \in T$

- Union & Intersection :

$S \cup T : x \in S \cup T \Leftrightarrow x \in S$ or $x \in T$

$S \cap T : x \in S \cap T \Leftrightarrow x \in S$ & $x \in T$

$S_n \quad n = 1, 2 \ldots$

$x \in \bigcup_n S_n$ iff $x \in S_n$ for some n.

$x \in \bigcap_n S_n$ iff $x \in S_n$ for all n.

$S \cup (T \cup U) = (S \cup T) \cup U = S \cup T \cup U$

→ $S \cup T = T \cup S$

→ $S \cap (T \cup U) = (S \cap T) \cup (S \cap U)$, $S \cup (T \cap U) = (S \cup T) \cap (S \cup U)$

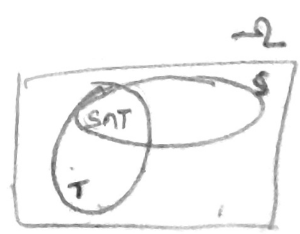
→ $(S^c)^c = S$ $\quad S \cap S^c = \phi$

$S \cap \Omega = S$

→ $S \cup \Omega = \Omega$

→ $S \subset T$ and $T \subset S \Leftrightarrow S = T$

* **De Morgan's Law :** $\boxed{(S \cap T)^c = S^c \cup T^c}$



$x \in (S \cap T)^c \Leftrightarrow x \notin S \cap T$

$\Leftrightarrow x \notin S$ or $x \notin T$

$\Leftrightarrow x \in S^c$ or $x \in T^c$

$\Leftrightarrow x \in S^c \cup T^c$

- Another form of De Morgan's Law:

$$\boxed{(S \cup T)^c = S^c \cap T^c}$$

General form of De Morgan's Law:

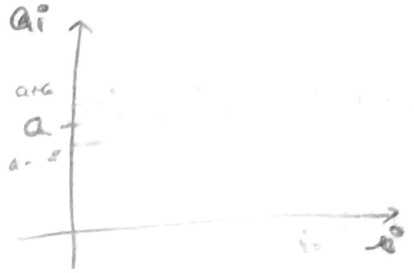
$$\left(\bigcap_n S_n\right)^c = \bigcup_n S_n^c$$

$$\left(\bigcup_n S_n\right)^c = \bigcap_n S^c$$

- Sequence and Limits

$a_1, a_2, a_3, \ldots$ $\quad i \in \mathbb{N} \in \{1, 2, 3 \ldots\}$

$a_i \in S \quad S = \mathbb{R} \quad \mathbb{R}^n$



sequence a_i, $\{a_i\}$

function $f: \mathbb{N} \to S$

$f(i) = a_i$

$\lim_{i \to \infty} a_i = a$

For any $\epsilon > 0$, there exists i_0, such that if $i \geq i_0$, then $|a_i - a| < \epsilon$

$a_i \xrightarrow[i \to \infty]{} a$

as i converges to ∞, a_i converges to a

- When does a sequence converge?

→ if $a_i < a_{i+1}$ for all i, then either
  - the sequence "converges to ∞"
  - the sequence converges to some real number a

→ if $|a_i - a| \leq b_i$ for all i, & $b_i \to 0$, then $a_i \to a$

Infinite Series:

$$\sum_{i=1}^{\infty} a_i = \lim_{n \to \infty} \sum_{i=1}^{n} a_i$$ provided limit exists.

- If $a_i \geq 0$; limit exists
- If term a_i do not have the same sign:
  - limit need not exist
  - limit may exist but be different if we sum in a different order
  - Fact: limit exists & independent of order of summation if

$$\sum_{i=1}^{\infty} |a_i| < \infty$$

Geometric series:

$$S = \sum_{i=0}^{\infty} a^i = 1+\alpha+\alpha^2 + \dots = \frac{1}{1-\alpha}, \ |\alpha| < 1$$

$$(1-\alpha)(1+\alpha+ \quad +\alpha^n) = 1-\alpha^{n+1}$$

$n \to \infty$ $$(1-\alpha)S = 1$$

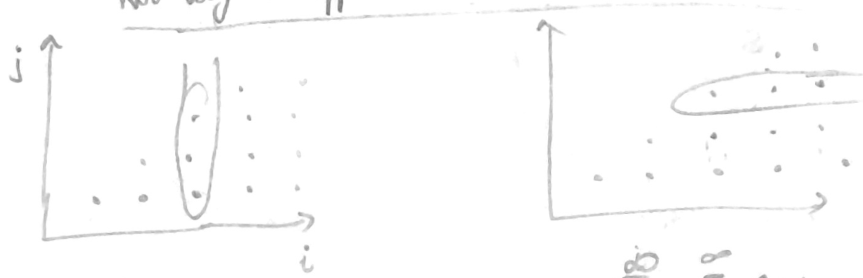
$$\boxed{S = \frac{1}{1-\alpha}}$$

- Series with multiple indices:

$$\sum_{i\geq 1, j\geq 1} a_{ij} \qquad \text{if } \sum |a_{ij}| < \infty \Rightarrow \sum_{i=1}^{\infty}\left(\sum_{j=1}^{\infty} a_{ij}\right) = \sum_{j=1}^{\infty}\left(\sum_{i=1}^{\infty} a_{ij}\right)$$

Two way to approach sum with multiple indices.

$$\sum_{(i,j):\, j\leq i} a_{ij}$$



$$\sum_{i=1}^{\infty}\sum_{j=1}^{i} a_{ij} \qquad\qquad \sum_{j=1}^{\infty}\sum_{i=j}^{\infty} a_{ij}$$

$$\text{if } \sum |a_{ij}| < \infty \Rightarrow \sum_{i=1}^{\infty}\sum_{j=1}^{i} a_{ij} = \sum_{j=1}^{\infty}\sum_{i=j}^{\infty} a_{ij}$$

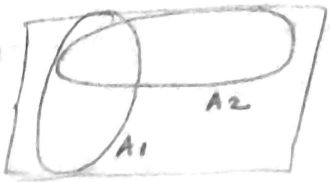
- Countable & Uncountable Infinite Set:

- Countable: can be put in 1-1 correspondence with positive integers Ω a_1, a_2, a_3
(discrete)
  - positive integers $\{a_1, a_2, a_3, \dots\} = \Omega$
  - integers.
  - pairs of positive integers

- Uncountable: not countable
  - the interval eg: $[0,1]$
  - the reals, plane, space

→ **Interpreting the union bound & the Bonferroni inequality:**

$\boxed{P(A_1 \cup P_2) \le P(A_1) + P(A_2)}$ union bound



$\boxed{P(A_1 \cap P_2) \geq P(A_1) + P(A_2) - 1}$

the Bonferroni inequality

Proof: $P((A_1 \cap A_2)^c) = P(A_1^c \cup A_2^c) \le P(A_1^c) + P(A_2^c)$

$1 - P(A_1 \cap A_2) \le 1 - P(A_1) + 1 - P(A_2)$

$\boxed{P(A_1 \cap A_2) \geq P(A_1) + P(A_2) - 1}$

The Bonferroni inequality

$P(A_1 \cap \dots \cap A_n) \geq P(A_1) + P(A_2) + \dots + P(A_n) - (n-1)$

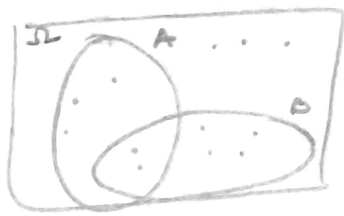
→ **conditioning & Bayes' rule:**

- Conditional probability
- three important tools:
  - Multiplication rule
  - Total probability rule
  - Baye's rule

Example:

$P(A) = \frac{5}{12}$

$P(B) = 6/12$



· If told that B occurred

$P(A|B) = \frac{2}{6}$ $P(B|B) = 6/6$

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

· $P(A|B)$ = "probability of A, given that B occurred"

defined only when $P(B) > 0$

· $P(A|B) \geq 0$ · $P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$.

· $P(B|B) = 1$ · If $A \cap C = \varphi$, then $P(A \cup C|B) = P(A|B) + P(C|B)$

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

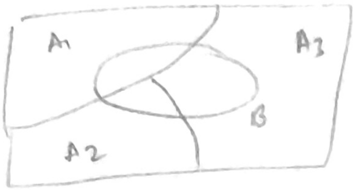
- The multiplication rule:

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

example: $P(A^c \cap B \cap C^c) = P(A^c \cap B).P(C^c | A^c \cap B)$

$= P(A^c).\ P(B|A^c).\ P(C^c | A^c \cap B)$

- Total Probability theorem:



- Partition of sample space into A_1, A_2, A_3
- Have $P(A_i)$ for every i

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

$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \dots$

$$\boxed{P(B) = \sum_i P(A_i)P(B|A_i)}$$

Total probability with partition of sample space

- Baye's Rule:

- Partition of sample space into A_1, A_2, A_3.
- Have $P(A_i)$ for every i initial "belief"
- Have $P(B|A_i)$ for every i

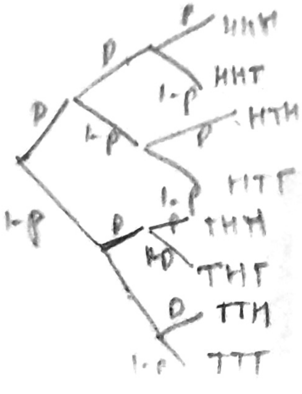
revised "beliefs" given that B occurred

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)\,P(B|A_i)}{\sum_j P(A_j)P(B|A_j)}$$

- Event Independence:

Example: A model based on conditional probability:

- 3 tosses of a baised coin: $P(H)=p$, $P(T)=1-p$



- $P(THT) = (1-p)(p)(1-p)$ 　　 $P(H_2|H_1) = p = P(H_2|T_1)$.
- Total probability

$P(1\ \text{head}) = 3p(1-p)^2$

- Baye's rule:

$$P(\text{first toss } H \mid 1\ \text{head}) = \frac{P(H_1 \cap 1\ \text{head})}{P(1\ \text{head})} = \frac{p(1-p)^2}{3p(1-p)^2} = \frac{1}{3}$$

- **Independence of 2 events:**

→ Intuitive "definition": $P(B|A) = P(B)$
- occurance of A provides no new information about B.

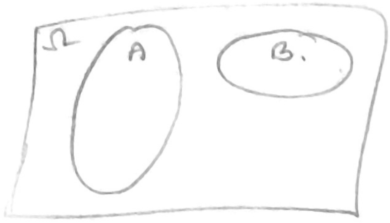
$\therefore P(A \cap B) = P(A)\,P(B|A) = P(A).P(B)$ {since $P(B|A) = P(B)$} for independent events

Definition of Independence:

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

- Symmetric with respect to A and B.
- implies $P(A|B) = P(A)$

example



$P(A)>0$ $P(B)>0$

$P(A \cap B) = 0$

$P(A).P(B) > 0$

$P(A \cap B) \neq P(A).P(B)$

A & B are not independent

- 2 events are independent if the occurence of one does not affect the probability of occurance of the other.
- If A and B are independent, then A & B^c are independent.
  - Intuitive argument.

$A = (A \cap B) \cup (A \cap B^c)$

$P(A) = P(A \cap B) + P(A \cap B^c)$

$= P(A).P(B) + P(A \cap B^c)$

$P(A \cap B^c) = P(A)(1 - P(B))$

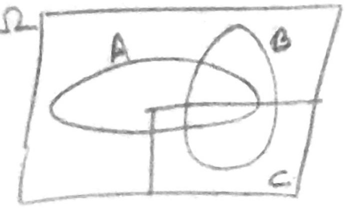
$P(A \cap B^c) = P(A).P(B^c)$

Hence A & B^c are independent

- **Condition Independence:**

Condition independence, given C, is defined as independence under the probability law $P(.|C)$

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



- Independence of collection of events:

- Intuitive "definition": Information on some of the events does not change probabilities related to the remaining events.

$A_1, A_2 \dots$ independent $\Rightarrow P(A_3 \cap A_4^c) = P(A_3 \cap A_4^c \mid A_1 \cap (A_2 \cap A_5^c))$

Definition: Events $A_1, A_2, \dots A_n$ are called independent if.

$$P(A_i \cap A_j \cap A_k \dots A_m) = P(A_i) \cdot P(A_j) \dots P(A_m)$$

for distinct indices $i, j \dots m$

- Independence vs pairwise independence

HH 1/4	HT 1/4
TH 1/4	TT 1/4

- Two independent fair coin tosses
  - H_1: first toss is H
  - H_2: second toss is H.

$$P(H_1) = P(H_2) = 1/2$$

- C: the two tosses had the same result $:= \{HH, TT\}$

$P(H_1 \cap C) = P(H_1 \cap H_2) = 1/4$

$P(H_1)P(C) = \frac{1}{2} \times \frac{1}{2} = 1/4$

$P(H_1 \cap C) = P(H_1) \cdot P(C)$

Hence, H_1 & C are independent events

Similarly, H_2 & C: independent

$P(H_1 \cap H_2 \cap C) = \frac{1}{4}$

$P(H_1) \cdot P(H_2) P(C)$

$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$

H_1, H_2 & C are pairwise independent but not independent.

- Reliability: p_i: probability that unit i is 'up'

U_i: ith unit is up

$U_1, U_2 \dots U_n$ independent

F_i: ith unit down

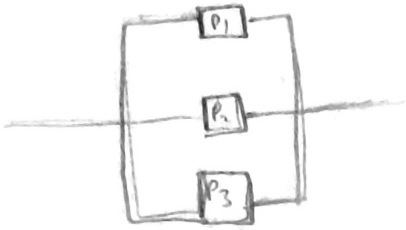
F_i: independent.

probability that system is up!

$P(\text{system up}) = P(U_1 \cap U_2 \cap U_3)$

$= P(U_1) P(U_2) P(U_3)$

$= p_1 p_2 p_3$



$P(\text{system is up}) = P(U_1 \cup U_2 \cup U_3)$

$= 1 - P(F_1 \cap F_2 \cap F_3)$

$= 1 - P(F_1) \cdot P(F_2) \cdot P(F_3)$

$= 1 - (1-p_1)(1-p_2)(1-p_3)$

Example: The king's sibling

- The king comes from the family of 2 children. What is the probability that his sibling is female?

→ $P(\text{boy}) = P(\text{girl}) = 1/2$] → $P(\;) = 1/2$? ✗
independent

BB 1/4	BG 1/4
GB 1/4	GG 1/4

since king is boy

$\underline{\underline{2/3}}$ 3 outcome ⇒ equal probability = 1/3

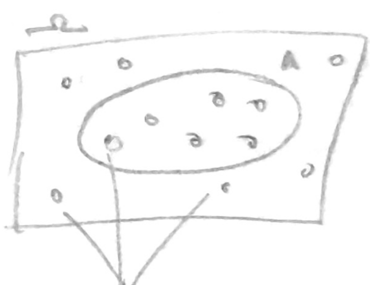
$P(\text{female sibling} \mid \text{boy}) = 2/3$

* **Counting:**

Discrete Uniform Law:

- Assume Ω consists of n equally likely elements
- Assume A consists of k elements

then: $P(A) = \dfrac{\text{no of elements of } A}{\text{no of elements of } \Omega} = \dfrac{k}{n}$



prob = $\frac{1}{n}$

→ Basic counting principle:

- r stages
- n_i choices at stage i

Number of choices is: $n_1 \cdot n_2 \cdots n_r$

- Permutations: Number of ways of ordering n elements. $n!$

n	$(n-1)$	$(n-2)$	...				1

$n \times (n-1) \times (n-2) \cdots 1 = n!$

- Number of subsets: $\{1, \ldots n\}$: 2^n

- <u>Combinations</u>: – How many ways to select k element from given set of n.

$\binom{n}{k}$: number of k-elements subsets of a given n-element set

$$= \frac{n!}{k!\,(n-k)!}$$

- Two ways of constructing an ordered sequence of k distinct items:
  - choose the k items one at a time: $\rightarrow \frac{n!}{(n-k)!}$
  - Chose k items, then order them.
    $\hookrightarrow \binom{n}{k} k!$

$$\boxed{\binom{n}{k} = \frac{n!}{(n-k)!\,k!}} \qquad \left\{ \sum_{k=0}^{n} \binom{n}{k} = 2^n \right\}$$

→ Binomial coefficient $\binom{n}{k}$ → Binomial probabilities.

- $n \geq 1$ independent coin tosses: $P(H) = p \quad P(T) = 1-p$
- $P(HTTHHH) = p(1-p)(1-p)ppp = p^4(1-p)^2$
- $P(\text{particular sequence}) = p^{\#\text{heads}} (1-p)^{\#\text{tails}}$
- $P(\text{particular } k\text{-head sequence}) = p^k (1-p)^{n-k}$
- $P(k \text{ heads}) = p^k (1-p)^{n-k} \cdot (\#\, k\text{-head seq.})$
  $= p^k (1-p)^{n-k} \binom{n}{k}$

$$P(k \text{ heads}) = \binom{n}{k} p^k (1-p)^{n-k}$$

→ <u>A coin tossing problem</u>.

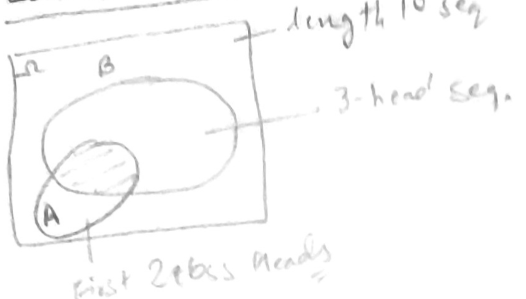
- Given that there were 3 heads in 10 tosses, what is the probability that the first 2 tosses were heads?
  - event A: the first 2 tosses were heads?
  - event B: 3 out of 10 tosses were heads?

Assumption:
- independence
- $P(H) = p$

First Solution

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{p^2 \cdot \binom{8}{1} p^1 (1-p)^7}{\binom{10}{3} p^3 (1-p)^7} = \frac{8}{\binom{10}{3}}$$

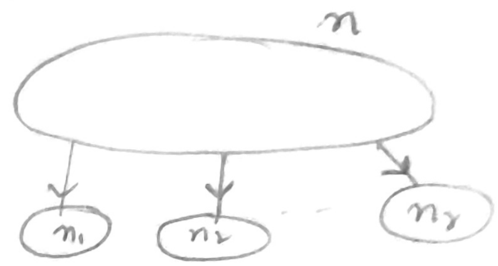
- **Second Solution:**



$$P = \frac{\#\text{ in } A \cap B}{\#\text{ in } B}$$

$$\boxed{P = \frac{8}{\binom{10}{3}}}$$

- **Partitions:** # of ways a given set can be partitioned on basis of given sizes.



$n_1 + n_2 + \dots + n_r = n$

- $n \geq 1$ distinct items; $r \geq 1$ persons
  given n_i items to person i
  - here $n_1, \dots, n_r$ are given nonnegative integers.
  - with $n_1 + \dots + n_r = n$

- Ordering n items: $n!$

  $c \cdot n_1! \, n_2! \dots n_r! = n!$

  c: # of ways to partition given set into subsets

$$\boxed{\text{number of partitions} = \frac{n!}{n_1! \, n_2! \dots n_r!}}$$ (multinomial coefficient)

Example: 52-card deck, dealt (fairly) to 4 players
Find P(each player gets an ace)

- Outcome: partition into 4 players (equally likely).
  - number of outcomes: $\frac{52!}{13!\,13!\,13!\,13!}$
  - Constructing an outcome with one ace for each person
    - distribute the aces — $4 \cdot 3 \cdot 2 \cdot 1 = 4!$] × product
    - distribute the remaining 48 cards ⇒ $\frac{48!}{12!\,12!\,12!\,12!}$

$$P = \frac{4! \times \frac{48!}{12!\,12!\,12!\,12!}}{\frac{52!}{13!\,13!\,13!\,13!}}$$

→ **Second Solution:**

Stack the deck, aces on top

$$\frac{39}{51} \times \frac{26}{50} \times \frac{13}{49} = 0.105$$

→ **The Multinomial Probabilities**

- balls of different colors $i = 1, \dots, r$
- probability of picking a ball of color i is p_i
- draw n balls, independently.
- given non negative numbers n_i with $n_1 + \dots + n_r = n$
- find $P(n_1$ balls of color 1, n_2 balls of color 2, …, n_r balls of color $r)$

$$P(\text{particular sequence of "type" } (n_1, n_2, \dots n_r)) = p_1^{n_1} p_2^{n_2} \dots p_r^{n_r}$$

sequence of type $(n_1, n_2 \dots n_r) \longleftrightarrow$ partition of $\{1 \dots n\}$ into subsets of sizes $n_1, n_2 \dots n_r$

$$\boxed{P(\text{get type } (n_1, n_2, \dots n_r)) = \frac{n!}{n_1!\, n_2! \dots n_r!} \, p_1^{n_1} p_2^{n_2} \dots p_r^{n_r}}$$

Lecture 5 – Discrete random variables:

- **Probability mass functions and expectations:**
  - Discrete random variables – takes value in finite or countable set.
- Random variable: numerical quantity that takes random values.

Random variable:

- A random variable ("r.v.") associates a value (a number) to every possible outcome.
- Mathematically: A function from the sample space Ω to the real numbers.
- It can take discrete or continuous values

- Probability mass function (PMF) of a discrete r.v X:

→ It is the "probability law" or "probability distribution" of X.

Ω a b c d X x 5 4 3

prob $= \frac{1}{4}$

- If we fix some x, then "$X = x$" is an event

$x = 5$ $\quad X = 5$ $\quad \{\omega : X(\omega) = 5\} = \{a, b\}$

$$p_X(5) = \frac{1}{2}$$

$$\boxed{p_X(x) = P(X = x) = P(\{\omega \in \Omega \text{ s.t } X(\omega) = x\})}$$

- $p_X(x) \geq 0$

- $\sum_x p_X(x) = 1$

- PMF calculation:

- Two rolls of a tetrahedral die:

- Let every possible outcome have probability 1/16. Find $p_Z(z)$ for all z

$$Z = X + Y$$

Y = second roll				
4	5	6	7	8
3	4	5	6	7
2	3	4	5	6
1	2	3	4	5
	1	2	3	4

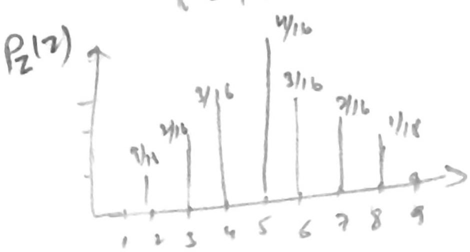
X = First roll

- repeat for all z:
- collect all possible outcomes for which Z is equal to z
- add their probabilities

$$p_Z(2) = P(Z = 2) = \frac{1}{16}$$

$$p_Z(3) = P(Z = 3) = \frac{2}{16}$$

$$p_Z(4) = P(Z = 4) = \frac{3}{16}$$



Examples of Random Variables:

① Bernoulli with parameter $p \in [0, 1]$

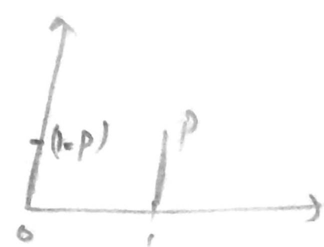
$$X = \begin{cases} 1 & \text{w.p } p \\ 0 & \text{w.p } (1-p) \end{cases}$$

given a value of 0 or 1 with a certain given probability.

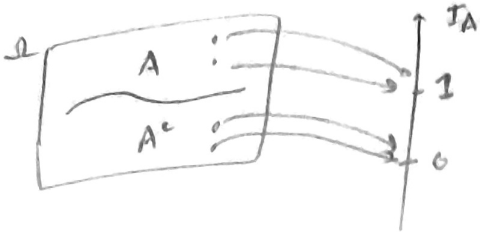
$p_X(0) = 1 - p$

$p_X(1) = p$

- Models a trail that results in success/failure.



- Indicator r.v. of an event A: $I_A = 1$ iff A occurs. Here, the indicator random variable is a Bernoulli r.v.



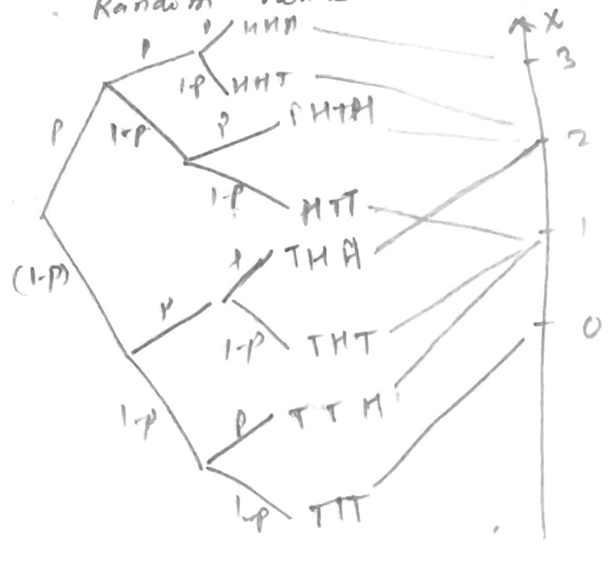
$P_{I_A}(1) = 1 = P(I_A = 1) = P(A)$

→ Discrete uniform random variable ; parameters a, b

- Parameters: a, b;
- Experiment : Pick one of $a, a+1, \dots b$ at random ; all equally likely.
- Sample Space : $\{a, a+1, \dots b\}$
- Random variable : $X : X(\omega) = \omega$.

→ Binomial random variable : parameters : positive integer n : $p \in [0,1]$

- Experiment : n independent tosses of a coin with $P(\text{Heads}) = p$
- Sample space : set of sequences of H & T of length n
- Random variable X : number of Heads observed.



$P_X(0) = P(X=0) = (1-p)^3$

$P_X(2) = P(X=2) = P(HHT) + P(HTH) + P(THH)$

$= 3p^2(1-p)$

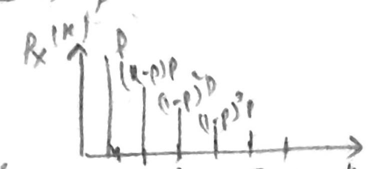
$= \binom{3}{2} p^2 (1-p)$

$$\boxed{P_X(k) = \binom{n}{k} p^k (1-p)^{n-k}} \quad \text{for } k = 0, 1, \dots n$$

→ Geometric random variable : parameter p: $0 < p \le 1$

Experiment : infinitely many independent tosses of a coin. $P(\text{Heads}) = p$

- Sample space : Set of infinite sequence of H & T.
- Random Variable X : number of tosses until the first Heads.
- Model of : waiting times, number of trials until a success.



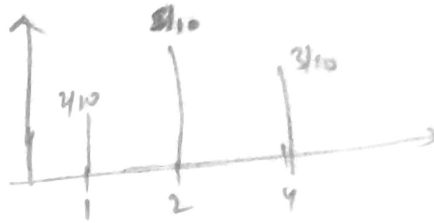
$P_X(k) = P(X = k) = P(\underbrace{T \dots T}_{k-1} H) = (1-p)^{k-1} p \quad , k = 1, 2, 3 \dots$

$P(\text{no Heads ever}) \le P(\underbrace{T \ldots T}_{k}) = (1-p)^k \xrightarrow{k \to \infty} 0$,

$P(\text{no Heads ever}) = 0$

→ **Expectation / mean of a random variable:**

- Motivation: Play a game 1000 times
  Random gain at each play described by $X = \begin{cases} 1 & \text{wp } 2/10 \\ 2 & \text{wp } 5/10 \\ 4 & \text{wp } 3/10 \end{cases}$



- Average gain in 1000 games

$$\frac{1 \cdot 200 + 2 \cdot 500 + 4 \cdot 300}{1000}$$

$$= 1 \cdot \frac{2}{10} + 2 \cdot \frac{5}{10} + 4 \cdot \frac{3}{10}$$

Definition $\quad E[X] = \sum_x x\, p_X(x)$

- Interpretation: Average in large number of independent repetitions of the experiment
- Caution: If we have an infinite sum, it needs to be well-defined.
  We assume $\sum_x |x|\, p_X(x) < \infty$

- **Expectation of a Bernoulli r.v**

$X = \begin{cases} 1 & \text{wp. } p \\ 0 & \text{wp } 1-p \end{cases} \qquad E[X] = 1 \cdot p + 0 \cdot (1-p) = p$

If X is the indicator of an event A, $X = I_A$
$X = 1$ iff A occurs $\qquad p = P(A)$

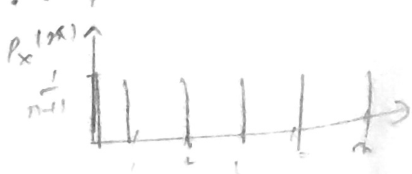
$$E[I_A] = P(A)$$

→ **Expectation of a uniform r.v.**

- Uniform on $0, 1, \ldots, n$

$$E[X] = 0 \cdot \frac{1}{n+1} + 1 \times \frac{1}{n+1} + 2 \cdot \frac{1}{n+1} + \ldots + n \frac{1}{n+1}$$

$$= \frac{1}{n+1}(0 + 1 + \ldots + n) = \frac{1}{n+1} \cdot \frac{n(n+1)}{2} = \frac{n}{2}$$



$$\boxed{E[X] = \frac{n}{2}}$$

→ Expectation as a population average:

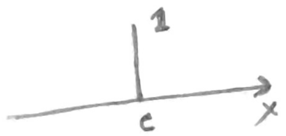
- n students
- Weight of ith student: x_i
- Experiment: pick a student at random, all equally likely
- Random variable X: weight of selected student
  - assume the x_i are distinct

$p_X(x_i) = \frac{1}{n}$ $E[X] = \sum_i x_i \frac{1}{n} = \underbrace{\frac{1}{n} \sum x_i}_{\text{average over all student}}$

Elementary Properties of Expectation:

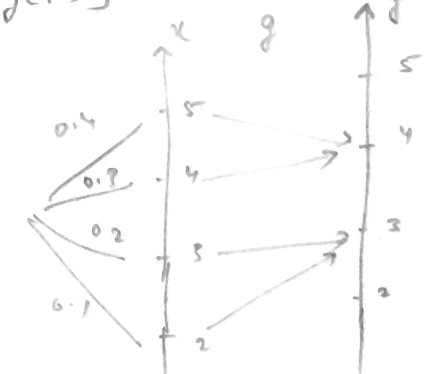
Definition $E[X] = \sum_x x\, p_X(x)$

- If $X \geq 0$, then $E[X] \geq 0$
  for all w: $X(w) \geq 0$
- If $a \leq X \leq b$ then $a \leq E[X] \leq b$
  for all w: $a \leq X(w) \leq b$, $E[X] = \sum_x x\, p_X(x) \geq \sum_x a\, p_X(x)$
  $= a \sum_x p_X(x)$ $= 1$
  $\geq a$
- If c is a constant, $\boxed{E[c] = c}$



→ The expected value rule, for calculating $E[g(X)]$

- Let X be a r.v & $Y = g(X)$
- Averaging over y:
  $E[Y] = \sum_y y\, p_Y(y)$
  $= 3 \cdot (0.1 + 0.2) + 4\,(0.4 + 0.3)$
  $= 3 \times 0.1 + 3 \cdot 0.2 + 4 \times 0.3 + 4 \times 0.4$



- Averaging over x:

$$\boxed{E[Y] = E[g(X)] = \sum_x g(x)\, p_X(x)}$$