Probablity

Applications
- History
- Finance
- Gambling
- Life

→ **Sample Space**: Set of all possible outcomes of an experiment.

→ Event: Is the subset of sample space

→ Naive Definition of Probablity: $P(A) = \frac{\text{No. favorable outcomes to event A}}{\text{No. of Total possible outcome}}$

* Assumes that all outcomes are equally likely & finite Sample Space

Basic principles of Counting

Multiplication Rule: if a experiment can have n_1 possible outcomes and for each output of first experiment there are n_2 possible outcomes ..... & n_r outcomes possible for r^{th} experiment. Then overall there are $n_1 n_2 \dots n_r$ overall possible outcomes.

Combination: $^nC_r = \frac{n!}{(n-r)!\,r!}$

Sampling Table: Choose k objects out of n

	order matters	order doesn't matter
replacement	n^k	$(n+k-1)C_k$ ★★★
without replacement	nP_k	nC_k

choose k objects from a set of n (order doesn't matter with replacement)

$(n+k-1)C_k$

Extreme case

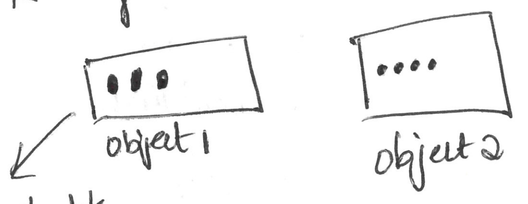
$k=0 \Rightarrow 0$ } trivial cases

$k=1 \Rightarrow n$

$\boxed{n=2}$ simplest non trivial example. $\Rightarrow {}^{k+1}C_k = {}^{k+1}C_1 = k+1$

When $n=2$

$k=$ any no.



No. of dots $\in \{0,1,2,\dots k\}$

• → choosen

order does'nt matter

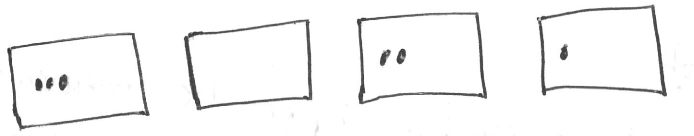
replacement → one object can be choosen multiple times

sum of all dots $=k$

Equalent problem. How many ways to put k indistinguishable particles into n distinguishable boxes → $(n+k-1)C_k$

★ Order does'nt matter therefore dots are completely indistinguishable & swapable

eg:- $n=4$



$n=4$, $k=6$

graphical encoding → •••| |••|•

k dots — •

n-1 separators — |

n — boxes. — □

- Equalent problem: Arrange k dots & $n-1$ separators

$$=\frac{(n+k-1)!}{(n-1)!\,k!} = (n+k-1)C_k$$

- Equalent problem: Total $n+k-1$ positions arrange k dots in them

$$=(n+k-1)C_k$$

Story proof: Proof by Interpretation

1) $nC_k = nC_{n-k}$

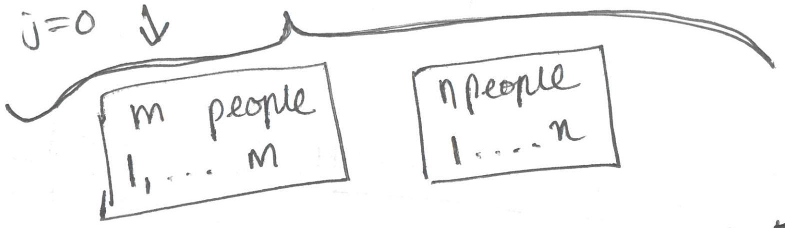
2) $n\,(n-1)C_{k-1} = k\,nC_k$ → pick k people out of n with 1 people designated as leader

↓

pick a ~~[illegible]~~ leader from n people and then attach $k-1$ people from remaining $(n-1)$ people

3) $^{m+n}C_k = \sum_{j=0}^{k} {}^{m}C_j \cdot {}^{n}C_{k-j}$ — Vandermonde's identity

↓ picking k out of m+n



Total = k

No of ways = (No of ways to pick 0 form people ✱ No of ways to pick k from n people) +

(No·of ways to pick 1 from m people ✱ No·of ways to pick k-1 from n people)

Non. Naive definition of plobablity

A plobablity space consists of S & P.
where S is sample space; Set of all outcomes of an experiement
and P is a function/mapping : which takes any evenet 'A' ∈ S as input gives P(A) which is a $\boxed{\text{no. between } 0 \,\&\, 1.}$ as the output, such that

Rule 0

Rule No. 1 $P(\phi) = 0, \; P(S) = 1$

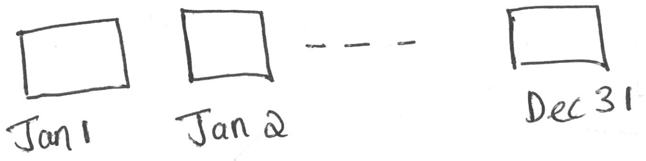
Rule No. 2 $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$ if $A_1, A_2 \dots A_n$ are disjoint

Axioms only 2

Birthday Problem : K people, Find the plobability that atleast 2 have the same birthday.

Assumption :- exlude leap day (365 days a year)
:- 365 days are equally likely
:- Assume independence of birthdays

If $k > 365$ then $P = 1$



(pigeon hole principle)

if $K<365$

$$P(\text{no match}) = \frac{365 \cdot 364 \cdot 363 \cdots (365-K+1)}{365^K} = \frac{365!}{(365-k)! \cdot 365^K} = \frac{{}_{365}P_K}{365^K}$$

$$P(\text{match}) = 1 - \frac{{}_{365}P_K}{365^K} = \begin{cases} 50.7\%, \text{ if } k=23 \\ 97\%, \text{ if } k=50 \\ 99.999\% \text{ if } k=100 \end{cases}$$

Consequens of Fundamental Axioms of Probablity

1) $P(A^c) = 1 - P(A)$
2) If $A \subseteq B$, then $P(A) \leq P(B)$
3) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4) $P(A \cup B \cup C) = P(A) + P(B) + P(C) - [P(A \cap B) + P(A \cap C) + P(B \cap C)] + P(A \cap B \cap C)$
5) $P(A_1 \cup A_2 \cup A_3 \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k)$

$---- + (-1)^{n+1} P(A_1 \cap A_2 \cap \cdots A_n)$

---

demontmort's Problem / Matching problem

N cards labelled 1 through N

N positions 1 to N

if cards are shuffled to a random sequence and arranged then what is the probablity that from position 1 to N atleast one card card label matches with card position.

→ Let A_j be the event that j^{th} card label matches with position j
ie card in j^{th} position is labelled j

$P(A_1 \cup A_2 \cup \dots A_n)$ = Probablity that atleast one card label mathes with possition.

$P(A_j) = \frac{1}{n}$ Since all positions are equally likely for a card labelled

$$P(A_1 \cap A_2) = \underset{i=j}{P(A_i \cap A_j)} = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$$

$$P(A_1 \cap A_2 \dots A_k) = \frac{(n-k)!}{n!} = \frac{1}{n(n-1)(n-2)\dots(n-k+1)}$$

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = n\frac{1}{n} - \left(nC_2 \frac{1}{n(n-1)}\right) + \left(nC_3 \frac{1}{n(n-2)(n-1)}\right) + \dots$$

$$P(A_1 \cup A_2 \dots A_n) = n \cdot \frac{1}{n} - \left(\frac{n(n-1)}{2!} \cdot \frac{1}{n(n-1)}\right) + \left(\frac{n(n-1)n-2}{3!} \cdot \frac{1}{n(n-1)(n-2)}\right)$$

$$P(A_1 \cup A_2 \cup \dots A_n) = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \frac{1}{5!} + \dots \dots (-1)^{n+1}\frac{1}{n}$$

$P(A_1 \cup A_2 \cup \dots A_n) \approx 1 - \frac{1}{e}$ for $n > 10$

∴ Probablity that no card matches = $1/e$

Independence Definition

Events A & B are independent ⇒ $P(A \cap B) = P(A) \cdot P(B)$

Note: Completely different from disjointness.

→ If A, B, C are independent if

$$\left.\begin{aligned} P(A \cap B) &= P(A) \cdot P(B) \\ P(A \cap C) &= P(A) \cdot P(C) \\ P(B \cap C) &= P(B)\, P(C) \\ P(A \cap B \cap C) &= P(A) P(B) P(C) \end{aligned}\right\} \text{All are required}$$

Similarly for events $A_1 \dots A_n$

Newton-Pepys Problem

Q) Fair dice, which is more likely amoung the following 3.

A) At least one 6 with 6 dice ← Truth

B) At least 2 6's with 12 dice

C) At least 3 6's with 18 dice ← Pepys Intuition

$$P(A) = 1 - \left(\frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6}\right) = 0.665$$

$$P(B) = 1 - \underbrace{\left(\frac{5}{6}\right)^{12}}_{P(\text{NO SIX})} - \underbrace{\left[\frac{1}{6} \cdot \left(\frac{5}{6}\right)^{11} \cdot 12\right]}_{P(\text{exactly one six})} = 0.619$$

$$P(C) = 1 - \left(\frac{5}{6}\right)^{18} - \left[\frac{1}{6} \left(\frac{5}{6}\right)^{17} \cdot 18\right] - \left[\frac{1}{36} \cdot \left(\frac{5}{6}\right)^{16} \cdot {}^{18}C_2\right]$$

$$P(C) = 1 - \sum_{k=0}^{2} \left({}^{18}C_k \left(\frac{1}{6}\right)^k \cdot \left(\frac{5}{6}\right)^{18-k} \right) \approx 0.597$$

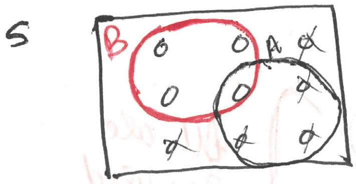
Conditional Probablity

How should you update prob/beliefs/uncertainity based on new evidence.

Definition

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, \text{ if } P(B) > 0$$

Intuition 1. Pebble world

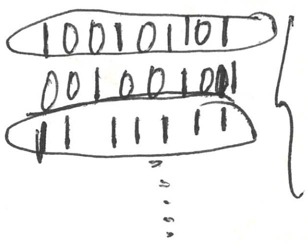


9 pebbles, total mass = 1

$P(A|B)$ → get rid of B^c

they are irrelevant since given B occured.

Intuition 2. Frequentist World : Repeat many times



- Circles reps where B occured
- Among those what fraction of time A also occured.

Theorm 1

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

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

Theorm 2

$$P(A_1 \cap A_2 \cap A_3 \ldots A_n) = P(A_1)\, P(A_2/A_1)\, P(A_3/A_2 \cap A_1$$

Theorm 3.

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

Baye's Rule

Law of Total Probablity

Thinking Conditionally

How to solve a problem

1) Try simple & extreme cases
2) break down to smaller problems.
3) Condition on first step.

$A_1 \ldots . A_n$ are partitions (they are disjoint & their union is the whole space)

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \ldots P(A_n \cap B)$$

$$\boxed{P(B) = P(B|A_1)P(A_1) + P(B|A_2)PA_2 \ldots P(B/A_n)P}$$

Law of Total probality.

Example: Get random 2 cards from a standard deck.

i) Find P(both are aces/one of them is ace)
ii) Find P(both aces | one of them is ace of spades)

i) $$\frac{P(\text{Both aces} \cap \text{Ace}, \text{one of them})}{P(\text{have ace})} = \frac{{}^4C_2 / {}^{52}C_2}{1 - \frac{{}^{48}C_2}{{}^{52}C_2}} = 1/33$$

ii) $$\frac{P(\text{both aces} \cap \text{Ace of spades})}{P(\text{have ace of spades})} = \frac{3}{51} = \frac{1}{17}$$

$\boxed{A_S}$ $\boxed{?}$

Ex 2.

Patients gets tested for a disease. And tests positive. Test is advertised as 95% accurating

Disease affects 1% of population

Assume that this means $P(T/D) = 95\% = P(T^c/D^c)$

Let D be the event that patient has disease

T is event that patient tests +ve

We need to find $P(D/T) = \frac{P(T/D).P(D)}{P(T)}$ for patient. Baye's Rule

$$P(D/T) = \frac{P(T/D).P(D)}{P(T/D)P.D + P(T/D^c)\,P(D^c)} \approx 0.16$$

$$\boxed{P(A/B) = 1 - P(A'/B)}$$

Biohazards: Common Mistakes.

1) Confusing $P(B/A)$ with $P(A/B)$ —— ["Prosecutors fallacy]

Ex) Sally clark's case,

---

2) Prior, Posterior.

P(Prior) = Before Evidence

P(Posterior) = After evidence.

$$\overset{\downarrow \text{Posterior}}{P(A/B)} = \frac{P(B/A).P(A)}{P(B)} \leftarrow \text{Prior}$$

---

3) Confusing independence & conditional independence.

Definition. Events A, B are conditionally independent given C

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

1) Does cond. Independance given C ⇒ independence? No

Ex. chess opponent of unknown strength

Game outcomes may be conditionally independent give strength of opponent

Game outcome not independent unconditionally because earlier trails give data that affects the predictable outcome of ~~be~~ later trails.

2) Does independence imply conditional indep. given C? No.

Counter example:

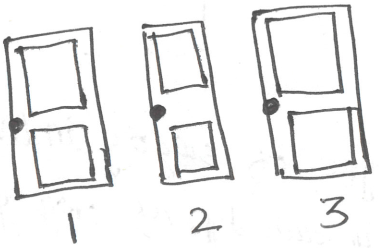
FT: Suppose a fire alarm goes off.
Caused by : F Fire
: C Pop-Corn

} But if alarm goes off they become dependent.

F & C are independent unconditionally

Mon~~t~~ Hall Problem / 3 doors Problem

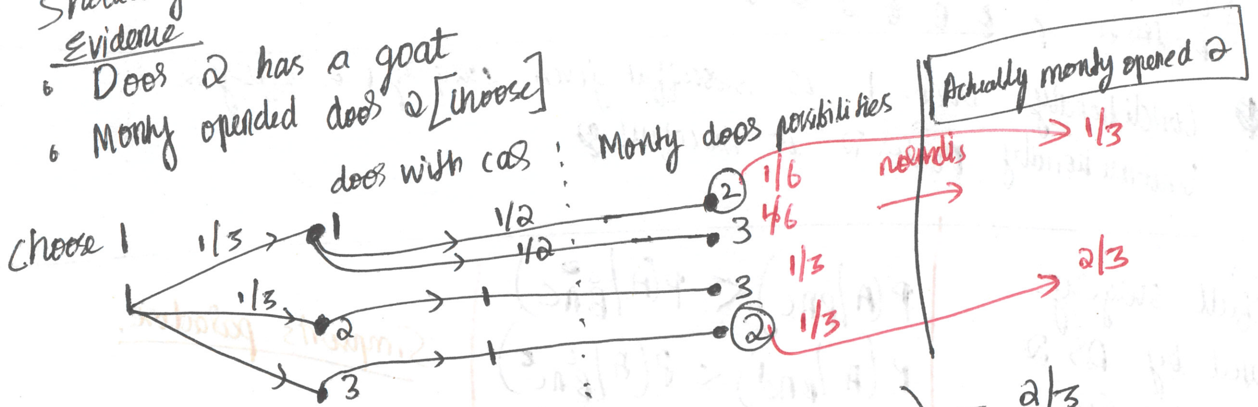
Problem:



1) 1 door has a car behind it
2) Others 2 has ~~goat~~ goats behind
3) Mon~~t~~ hall knowns which one has the car.
4) You want car and choose a door

5) Monty Reveals ~~an~~ door ~~2~~ ~~a~~ and shows a goat. (If monty has choice he choses equally likely)
6) Monty allows you to switch
7) Should you switch.

Evidence
- Door 2 has a goat
- Monty opended door 2 [choose]



p (Success if switching / Monty opens door 2) = 2/3

Intuition:
1/3 of the time your initial guess is ~~wro~~ correct & Monty ~~ch~~ eliminates randomly
2/3 of the time your initial guess is wrong & Monty eliminates the wrong one for you

Approach using law of total probability.

We wish we know where the car is.

S: Succeed using switch strategy.

D_j : Door j has car $\left(j \in \{1,2,3\}\right)$

$$P(S) = P(S|D_1)\,\tfrac{1}{3} + P(S|D_2)\,\tfrac{1}{3} + P(S|D_3)\,\tfrac{1}{3}$$

$$P(S) = 0 + \tfrac{1}{3} + \tfrac{1}{3} = \tfrac{2}{3}$$

$$P(S/M_2) = 2/3$$

Controversy

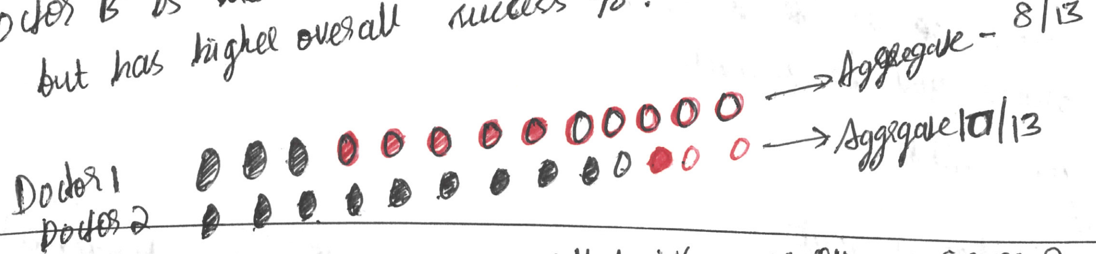
Solved with arguing the 1/2-1/2 fallacy in case of 100000 doors.

Simpson's Paradox

- Two doctors A & B

Doctor A is has success % in each type of surgery individually.

Doctor B is has lower success % in each type of surgery individually but has higher overall success %.



● Conditionally Doctor 1 is successful given surgery 1 & surgery 2

Unconditionally Doctor 2 is successful ●

A: Successful surgery

B: treated by Dr 2

C: Heart Surgery

C': Bandage Removal

$$\boxed{\begin{aligned} &P(A|B\cap C) < P(A|B^c\cap C) \\ &P(A|B\cap C') < P(A|B^c\cap C^c) \\ &P(A|B) > P(A|B^c) \end{aligned}}$$

Simpson's paradox.

C is called a confounder or control

$$P(A|B) = P(A|B\cap C)\,\underbrace{P(C|B)} + P(A|B\cap C^c)\cdot\underbrace{P(C^c|B)}$$

Weights

Probability of Heart Surgery given Doc 2

Probability of Bandage Removal of Doc 2

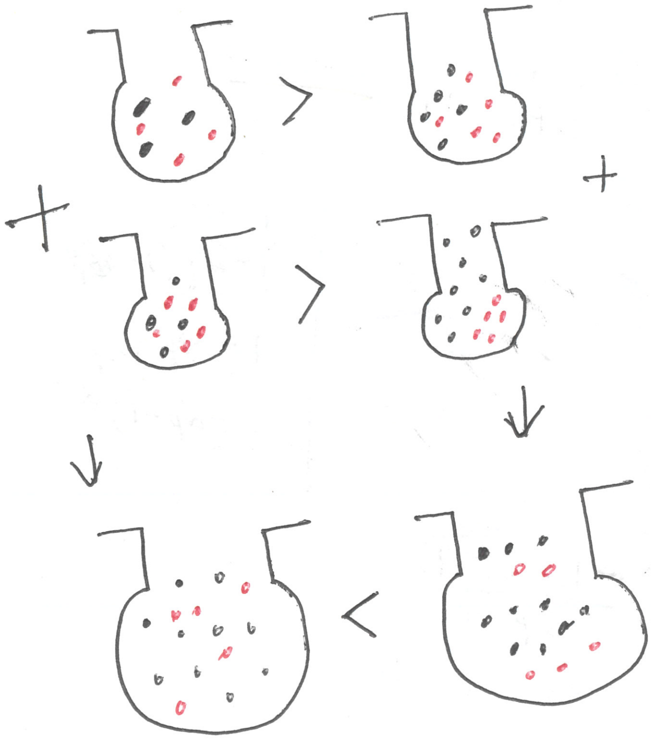
Weighted addition → Not direct additional.

Eg2.

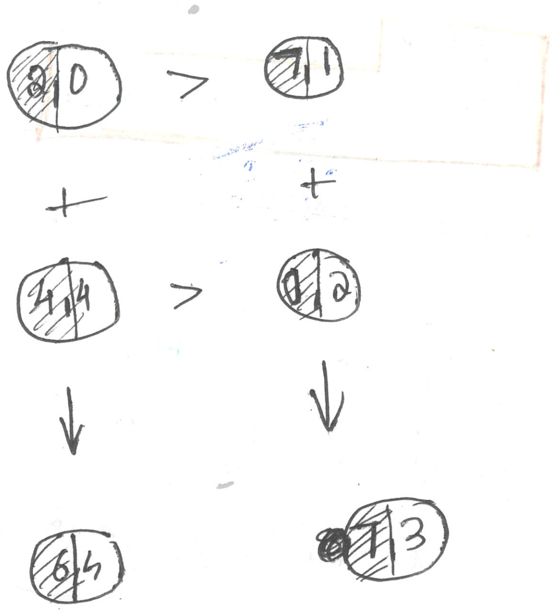
2 Jars

2 Jelly bean Flavours ● ●

only one flavour is favorite



Total Same.



1) Conditioning: The soul of Statistics

2) Random Variables & their distribution

Gambler's Ruin

Two gamblers A & B. Take turns taking bets and trades money back & forth. They bet 1$ each time. Game ends when ~~or~~ any player goes bankrupt.

$p = P$(A wins a certain round)

$q = P$(B wins a round)

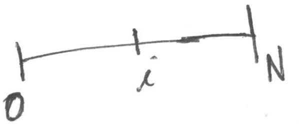
$p+q = 1$

P(A wins game) = ?

Assuming that A starts with $i$$ ~~with~~.

& B starts with $(N-i)$$

Random walk:



particle at i initially

P(it moves to left) → ~~p~~ q

P(it moves to right) → ~~q~~ p

$p+q = 1$

absorbing states @ 0,N

Strategy: Condition ~~of~~ on the first step.

Let P_i – P(A wins game/A starts with i dollars)

LOTP – Law of Total Prob.

$$\boxed{P_i = p(P_{i+1}) + q(P_{i-1})}, \quad 1 \le i \le (N-1) \quad \begin{cases} P_0 = 0 \\ P_N = 1 \end{cases}$$ Boundary.

difference equation

$1-4pq$

$1-4p(1-p)$

$1-4p+4p^2$

$=(2p-1)^2$

Guess
$P_i = x^i$

$\Rightarrow x^i = p\,x^{(i+1)} + q\,x^{(i-1)}$

$x = px^2 + q \rightarrow (px^2 - x + q = 0)$

$$x = \frac{1 \pm \sqrt{1-4pq}}{2p} = \frac{1 \pm (2p-1)}{2p}$$

$x = 1, \frac{2-2p}{2p}$

$x = 1, q/p$

General solution $P_i = A1^i + B(q/p)^i$ $\quad p \ne q$ to avoid repeated root.

$\Rightarrow$ $P_0 = 0 \Rightarrow A + B$

$\Rightarrow A = -B$

$P_N = 1 \Rightarrow A - A(q/p)^N$

$\Rightarrow A\left(1 - \left(\frac{q}{p}\right)^N\right) = 1$

$\Rightarrow A = \dfrac{1}{1-\left(\frac{q}{p}\right)^N}$

$$P_i = \begin{cases} \dfrac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)^N} & \text{if } p \ne q \\ i/N & \text{if } p = q \end{cases}$$

---

for repeated root case $p=q$

$x = \frac{q}{p}$

$$\lim_{x \to 1} \left(\frac{1-x^i}{1-x^N}\right) \overset{\text{LH rule}}{=} \frac{+ix^{i-1}}{-Nx^{N-1}} = \frac{i}{N}$$

$P_i + P_{N-i} = 1$ $\Big\}$ ⇒ There is zero probability that game goes on for ever

Prob that A wins + Prob that B wins

Random Variables

what is a RV? is a function from set $S \to \mathbb{R}$

It's a function from the sample space S to R $f: S \to \mathbb{R}$

- It is a ~~nu~~ numerical "summary of some/certain aspect" of experiment

eg:-

Definition of Bernoulli distribution: A random variable X is said to have Bernoulli distribution if X has two possible values 0 & 1

$P(X=1) = p$ — event — event $\{s : X(s) = 1\}$

$P(X=0) = 1-p$

Such a variable is called indicator random variable.

Binomial(n,p) The distribution of the no. of succes in n independent bernoullli trails. it's distribution is given by.

$$\boxed{P(X=k) = {}^nC_k \, p^k q^{n-k}}$$

$X \sim$ Binomial(n,p) $\Big\}$ independent
$Y \sim$ Binomial(m,p)

Then $X+Y \sim$ Binomial$(n+m, p)$

To add random Variables, they must have same domain.

Proff: Consider n trails & m more trails

1) Story: X is the no. of success in n independent Bernouli(p) trails

2) (Sum of Indicator random variable) $X = X_1 + X_2 + X_3 + \dots + X_n$, $X_j = \begin{cases} 1, & \text{if the } j^{th} \text{ trail a success} \\ 0, & \text{otherwise} \end{cases}$

$X_1, X_2 \dots X_n$ are independent bernouli trails.

Here both RVs X & Y operate on the domain

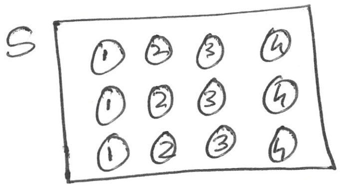
$S = \{$set of all possible outcomes of n ~~ba~~ bernoulli trails followed by m bernoulli trails$\}$

i.i.d → independent & identically distributed.

3) PMF: $P(X=k)$ ${}^nC_k p^k q^{n-k}$

RV may be functioned however we want. it need not take all trails into consideration

R.Vs



X=2, denotes an event
X≤3, denotes an event

C.D.F

$F(x) = P(X \leq x)$

F(x) is the CDF or Cumulative distribution function.

P.M.F (for discrete random variable)

PMF $p(X=a_j) \ \forall j$

$\boxed{0 \leq p_j \leq 1}$ $\boxed{\Sigma p_j = 1}$

~~Discrete means takes up finite~~

~~1st~~ Discrete is countable infinity or finite

Continous is uncountable infinite or finite.

$X \sim Binomial(n,p)$ } independent
$Y \sim Binomial(m,p)$

$X+Y \sim Binomial(n+m,p)$

$P(X+Y=k)$

$= \sum_{j=0}^{K} P(X+Y=k/X=j) \cdot P(X=j)$

$= \sum_{j=0}^{K} P(Y=k-j\cancel{/X=j}) \cdot P(X=j) = \sum_{j=0}^{K} P(Y=k-j)\, P(X=j)$

Y independent of x
∴ conditioning on x gives no info on Y.

$= \sum_{j=0}^{k} \left(mC_{k-j}\, p^{k-j} q^{m+j-k} \cdot nC_j\, p^j q^{n-j}\right)$

$= p^k q^{m+n-k} \sum_{j=0}^{k} \left[mC_{k-j}\, nC_j\right]$ — Vandermonde.

$= {}^{m+n}C_k\, p^k q^{m+n-k}$

= PMF of Binomial(m+n, k)

Ex. 5 card hand, Find the distribution of no. of Aces in the hand.

Let X represent the no of aces

Find $P(X=k) = \begin{cases} 0, \text{ except } k=0,1,2,3,4 \end{cases}$

$$P(X=k) = \frac{{}^{4}C_k \; {}^{48}C_{5-k}}{{}^{52}C_5} \quad k \in (0,1,2,3,4)$$

Q) Suppose we have jar full of marbles, where b of them are black & w of them are white. Pick a random sample of size n. What is the distribution of the no. of white marble in the sample. (sampling without replacement)

Let X → no. of white marbles in sample.

$$P(X=k) = \begin{cases} \dfrac{{}^{w}C_k \cdot {}^{b}C_{n-k}}{{}^{w+b}C_n}, & k \in (0,1,2 \ldots n) \\ & 0 \leq k \leq w, \\ & 0 \leq n-k \leq b \\ 0, \text{ otherwise} & \end{cases}$$

hypergeometric distribution

With replacement → binomial
large values of w & b, w+b ⇒
replacement/no replacement has little effect
eg 10,000 marbles & 6 samples.
∴ hypergeometric tends to binomial.

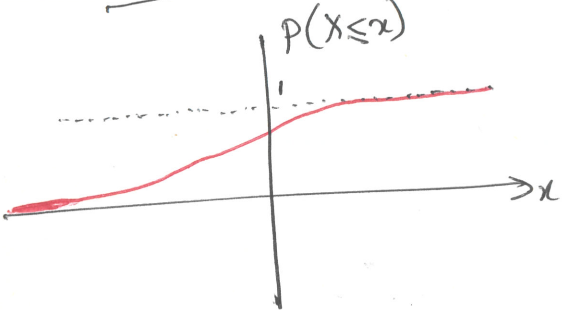
check for valid PMF

$$\sum_{k=0}^{n} \frac{{}^{w}C_k \cdot {}^{b}C_{n-k}}{{}^{w+b}C_n} \quad \text{Vandermond.}$$

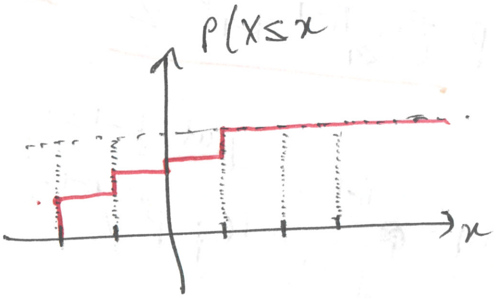
$$= \frac{{}^{w+b}C_n}{{}^{w+b}C_n} = 1$$

CDF $P(X \leq x)$

Continous



Discrete



Properties of CDF

1) increasing
2) right continous
3) $F(x) \to 0$ as $x \to -\infty$
   $F(x) \to 1$ as $x \to \infty$

Independance of R.V

X, Y are independent R.Vs if

$$P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y) \quad \forall\, x, y$$

Discret case

$$P(X=x, Y=y) = P(X=x) \cdot P(Y=y) \quad \forall\, x, y$$

Averages (Mean, Expectance) $E(X)$

$$\boxed{E(X) = \sum_{\forall x_i} P(X=x_i) \cdot x_i}$$

eg: $X \sim \text{Bernoulli}(p)$

$$E(X) = 1 \cdot P(X=1) + 0 \cdot P(X=0)$$

$$E(X) = p$$

but $p = P(A)$

$$\Rightarrow E(X) = P(A)$$

$$X = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{otherwise} \end{cases}$$

indicator random variable.

eg: Binomial(n, p)

$$E(X) = \sum_{k=1}^{n} k \cdot {}^{n}C_k \cdot p^k q^{n-k}$$

$$E(X) = \sum_{k=1}^{n} n \cdot {}^{n-1}C_{k-1} \cdot p^k \cdot q^{n-k}$$

$$E(X) = np \underline{\sum_{k=1}^{n} {}^{n-1}C_{k-1} p^{k-1} q^{n-k}} = (p+q)^n \text{ according binomial theorm.}$$

put $k-1=j$

$$E(X) = np \cdot \sum_{j=0}^{n-1} \left({}^{n-1}C_j \, p^j q^{n-1-j} \right)$$

$$\underline{E(X) = np}$$

Linearity.

$$\boxed{\begin{array}{c} E(X+Y) = E(X)+E(Y) \\ \text{even if } X \text{ \& } Y \text{ are dependent} \end{array}}$$

$E(X) = E(X_1+X_2+X_3+X_4+..X_n)$

$E(X) = E(X_1)+E(X_2)+E(X_3)+....$

$E(X) = p+p+p+\cdots p$

$E(X) = \underline{np}$

X_i — ...riable for the

√ X_i represen... on th... Sampl...

...ting n indicator variables

Ex.)

8

Le...

E(...

Otherwise

$$PMF = P(X=x) = \begin{cases} \dfrac{4C_x \cdot 48C_{5-x}}{52C_5} & x \in \{0,1,2,3,4\} \\ 0, & \text{otherwise.} \end{cases}$$

$E(X) = \sum_{x=0}^{4} x \cdot P(X=x)$

$E(X) = \sum_{x=0}^{4} x \cdot \dfrac{4C_x \cdot 48C_{5-x}}{52C_5}$

$E(X) = \sum_{x=1}^{4} 4 \cdot \dfrac{3C_{x-1} \cdot 48C_{5-x}}{52C_5}$

put $x-1=u$ $\quad x=1 \Rightarrow u=0 \quad x=4 \Rightarrow u=3$

$u = x+1$

$5-x = 5-(u+1)$
$= 4-u$

$\Rightarrow E(x) = \dfrac{4}{52C_5} \sum_{u=0}^{3} 3C_u \cdot 48C_{4-u}$

$E(x) = \dfrac{4}{52C_5} \cdot 51C_4$

$E(x) = \dfrac{4 \cdot 5!}{52 \cdot 4!} = \dfrac{4\times5}{52}$

$E(x) = \dfrac{5}{13}$

Expected value of hypergeometric = n... successful.

Geometric Distribution (p); ~~Geometric distributi~~

Independent bernoulli trails, bernoulli(p)
Each trail has the same probablity p of success } $q=1-p$

$X \rightarrow$ no. of failures before a success. $\quad X \sim geom(p)$

$P.M.F = P(X=k) = q^k \cdot p \quad k \in \{0,1,2,3....\}$

FFFFFS
$P(X=5) = q^5 p$

Valid Since.

★ $\sum_{k=0}^{\infty} pq^k = p \cdot \sum_{k=0}^{\infty} q^k = p \cdot \dfrac{1}{(1-q)} = \dfrac{p}{p} = 1$

$X \sim Geometric(p)$

$E(X) = \sum_{k=0}^{\infty} k \cdot pq^k \qquad p \sum_{k=1}^{\infty} kq^k$

Linearity.

$$\boxed{\begin{array}{l} E(X+Y) = E(X) + E(Y) \\ \text{even if } X \,\&\, Y \text{ are dependent} \end{array}}$$

$$E(X) = E(X_1 + X_2 + X_3 + X_4 + \dots X_n)$$
$$E(X) = E(X_1) + E(X_2) + E(X_3) + \dots$$
$$E(X) = p + p + p + \dots \; p$$
$$E(X) = \underline{np}$$

adding n indicator variables

$X_i \rightarrow$ indicator variable for the i^{th} bernouli trail.

$\sqrt{\;}$ xi opperates on this

Sample space = {all possible out comes of n bernoulli trail}

Ex.) 5 card hand, $X = (\#\text{aces})$, Find $E(X)$

randomly drawn from standard deck.

Let X_j be an indicator of j^{th} card being an ace. $\Big\{ \; 1 \le j \le 5$, $j \in (1, 2, 3, 4, 5)$

$$E(X) = E(X_1 + X_2 + X_3 + X_4 + X_5) \quad \text{(Linearity) } \sout{\text{Symmetry}}$$
$$E(X) = E(X_1) + E(X_2) + E(X_3) + E(X_4) + E(X_5)$$
$$E(X) = 5 \cdot E(X_1) \text{ — (due to symmetry)}$$
$$E(X) = 5 \cdot \left(\frac{4}{52}\right) = \underline{\underline{\frac{5}{13}}}$$

, even though X_j are dependent ★★★★

Expected value of hypergeometric $= n * \big|$ probablity of individual indicator being succesfull.

Geometric Distribution (p) ; ~~Geometric distributi~~

Independent bernouli trails, bernoulli(p)
Each trail has the same probablity p of success $\Big\} \; q = 1-p$

$X \rightarrow$ no. of failures before a success. $X \sim \text{Geom}(p)$

$$\text{P.M.F} = P(X=k) = q^k \cdot p \qquad k \in \{0, 1, 2, 3 \dots\}$$

FFFFFS
$$P(X=5) = q^5 p$$

Valid Since.

$$★ \; \sum_{k=0}^{\infty} pq^k = p \cdot \sum_{k=0}^{\infty} q^k = p \cdot \frac{1}{(1-q)} = \frac{p}{p} = 1 \,//$$

$X \sim \text{Geometric}(p)$

$$E(X) = \sum_{k=0}^{\infty} k \cdot pq^k \qquad p \sum_{k=1}^{\infty} kq^k$$

$$\Rightarrow E(X) = p \sum_{k=1}^{\infty} k q^{k}$$

$$E(X) = \frac{pq}{p^2} = \underline{\underline{q/p}}$$

$$\boxed{E(X) = q/p}$$

$$\sum_{k=0}^{\infty} q^{k} = \frac{1}{1-q}$$

take derivative w.r.t q

$$\sum_{k=1}^{\infty} k q^{k-1} = \frac{-1}{(1-q)^2} \cdot -1$$

$$\Rightarrow \sum_{k=1}^{\infty} k q^{k} = \frac{q}{(1-q)^2}$$

Story proof: Let $c = E(X)$

$$c = 0 \cdot p + (1+c)\,q$$

since coin is memory less.

L.O.T.P conditioning on the first trail.

very important technique. L.O.T.P conditioning on the first trail for expectance variable.

$$c = (1+c)\,q$$

$$c - cq = q$$

$$c = \frac{q}{(1-q)} = \frac{q}{p} = E(X)$$

Linearity of Expectance Proof

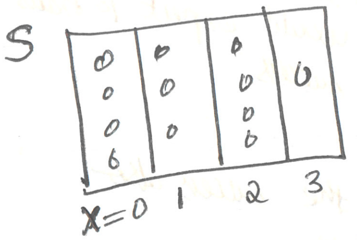
$$\sum_{t} t P(T=t) \overset{?}{=} \sum_{x} x P(X=x) + \sum_{y} y P(X=y)$$

Let $T = X+Y$

Show $E(T) = E(X+Y)$

$E(T) = E(X) + E(Y)$

$$P(T=t) = \sum_{x} P(T=t / X=x)\, P(X=x)$$



$$E(X) = \underbrace{\sum_x x\, P(X=x)}_{\text{grouped}} = \underbrace{\sum_s X(s)\, P(\{s\})}_{\text{ungrouped}}$$

Proff of Linearity (discrete case)

$$E(T) = \sum_s (X+Y)(s)\, P(\{s\})$$

$$E(T) = \sum_s (X(s)+Y(s)) + P(\{s\})$$

$$E(T) = \sum_s X(s)\, P(\{s\}) + \sum_s Y(s)\, P(\{s\})$$

$$E(T) = E(X) + E(Y)$$

Negative Binomial: parameters r, p

Story: Independent Bernouli (p) trails. No. of failures before the r^{th} succes.

PMF: $$\boxed{P(X=n) = p^r (1-p)^n \, {}^{(n+r-1)}C_{r-1}}$$

for $n = 0, 1, 2$ ---- so on

eg.
SFFFSFFSFFFFSFF

there would be r succes and n failures
the sequence would end with a succes.
before this succes we [illegible] have $n+r-1$ place holders
out of this we select $r-1$ placeholders for the remaing 'r-1' succes
& the rest is for failure.

$$E(X) = E(X_1 + X_2 + x \ldots X_r)$$

X_j = No. of failures between the $j-1^{th}$ & j^{th} success

$X_j \sim \text{geom}(p)$

$$E(X) = E(X_1) + E(X_2) + \text{---} \; E(X_r)$$

$$\boxed{E(X) = rq/p}$$

$X \sim FS(p)$ time until 1st success, including success

If there is 1 in 10 chance of success then you would expect 10 trails to see a success

Let $Y = X - 1$

$Y \sim geom(p)$

$$E(X) = E(Y) + 1 = q/p + 1 = \underline{\underline{1/p}}$$

↓ counting the success also intuitive

Putnam Example

Random permutation of integers $1, 2, \ldots, n$ where $n \geq 2$

Find the expected no. of local maxima

Let I_j be the indicator of position j being the local maxima

eg ③, 2, 1, 4, ⑦, 5, ⑥

greater than it's neighbours ⇒ local maxima

$$E(I_1 + I_2 + \ldots I_n)$$

$$= E(I_1) + E(I_2) + E(I_3)$$

$$= \frac{n-2}{3} + \frac{2}{2} = \frac{n+1}{3}$$

St. Petersburg Paradox.

Q) You get 2^x \$ where X is the no. of flips of coin until (& including) it lands heads. How much should be willing to be payed to play this game.

What price of game would make this a fair game.

Solution

$Y = 2^X$

$$E(Y) = \sum_{K=1}^{\infty} 2^K P(X=K) = \sum_{K=1}^{\infty} 2^k \frac{1}{2^K} = \sum_{K=1}^{\infty} 1 = \infty$$

but if the no. of turns is upper bounded or the upper limit of money is bounded then

eg 1 trillion \$ $= 2^{40}$ \$

$$\sum_{K=1}^{40} 2^k \cdot \frac{1}{2^k} = \sum_{K=1}^{40} (1) = 40\$$$

Which means

$$E(2^X) \neq 2^{E(X)}$$

Sympathetic Magic

Random Variable vs distribution Confusion.

Q) Sum of two random variables $(X+Y)$

$$PMF(X+Y) \neq PMF(X) + P.MF.Y$$

adding random variable is not the same as adding their distribution

"Word is not the thing, map is not the teritory"

Poisson Distribution

$$P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!} \quad k \in \{0,1,2,\ldots\}$$

λ is the "rate" parameter [+ve constant real no#]

check valid PMF.

$$\sum_{k=0}^{\infty}\left[\frac{e^{-\lambda}\lambda^k}{k!}\right] = e^{-\lambda}e^{\lambda} = 1$$

*Often used for applications where counting # of success where we have a large no. of trials but the probablity of success ~~of~~ ~~trail~~ each trail is very small.

eg:- # of emails that you get in a hour

$$E(X) = \sum_{k=0}^{\infty} k\frac{e^{-\lambda}\lambda^k}{k!}$$

$$E(X) = e^{-\lambda}\sum_{k=0}^{\infty}\left(\frac{k\lambda^k}{k!}\right)$$

$$E(X) = e^{-\lambda}\sum_{k=1}^{\infty}\left(\frac{k\lambda^k}{k!}\right)$$

$$E(X) = e^{-\lambda}\sum_{k=1}^{\infty}\left(\frac{\lambda^k}{(k-1)!}\right)$$

$$E(X) = e^{-\lambda}\cdot\lambda\sum_{k=1}^{\infty}\left(\frac{\lambda^{(k-1)}}{(k-1)!}\right)$$

$$E(X) = e^{-\lambda}\cdot\lambda\sum_{j=0}^{\infty}\left(\frac{\lambda^j}{j!}\right)$$

$$\boxed{E(X) = e^{-\lambda}\cdot\lambda e^{\lambda} = \underline{\lambda}}$$

Integrator random variables $X_1, X_2, X_3 \ldots X_n$ — bernouli trails.

X be the poisson R.V $= X_1 + X_2 + X_3 \ldots\ldots X_n$

$$\left.\boxed{\begin{array}{l} n \rightarrow \text{large} \\ P(X_i) = P_i \rightarrow \text{very small} \end{array}}\right\} \Rightarrow X \text{ follows poison}(\lambda)$$

where $\lambda = np$

$$\therefore PMF(X=x) = \text{poison}(\lambda) = \frac{e^{-x}\lambda^x}{x!}$$

$X \sim$ Binomial (n, p)

Let $n \to \infty$
$p \to 0$

Let $\lim_{\substack{n \to \infty \\ p \to 0}} (np) = \lambda$ (is held constant for every trail)

ie $np = \lambda$

$p = \lambda / n$

$$P(X=k) = {}^nC_k \, p^k (1-p)^{n-k}$$

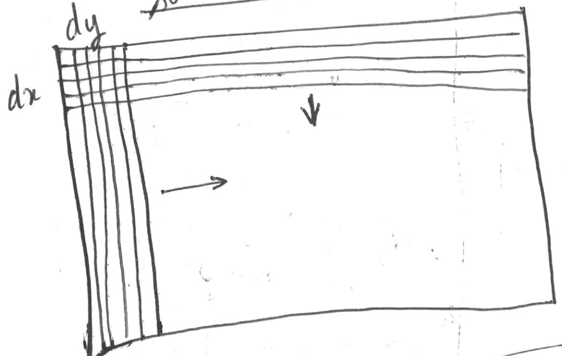
$$P(X=k) = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!} \cdot \left(\frac{\lambda^k}{n^k}\right)\left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$P(X=k) = \frac{\lambda^k}{k!} \frac{n}{n} \frac{(n-1)}{n} \frac{(n-2)}{n} \cdots \frac{n-k+1}{n} \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-k}$$

$$\left.\begin{matrix} Lt \\ n \to \infty \\ p \to 0 \end{matrix}\right\} \Rightarrow P(X=k) = \frac{\lambda^k}{k!} \cdot 1 \cdot 1 \cdot 1 \ldots 1 \cdot e^{-\lambda} \cdot 1$$

$$\Rightarrow P(X=k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

Raindrops



$dx, dy \to 0$

no. of raindrop hitting paper in one minute.

Eg:- n people, P(3 people with same birthday) Find the approximate $P(A)$

$A =$

${}^nC_3$ - triplets of people.

An indicator random variable for 3 people with same birthday $\{ I_{ijk}, \ i<j<k$

$E(\#\text{triple matches}) = {}^nC_3 \, P(I_{ijk})$

$= {}^nC_3 \cdot \frac{1}{365^2}$

$$P(I_{ijk}) = \frac{1}{365^2}$$

$\#\text{triple matches} = I = \sum_{ijk} I_{ijk}$

let I follow poison(λ)
Reasonable approximation

where $\lambda = E(I) = E(\#\text{triple matches})$

$\lambda = {}^nC_3 \frac{1}{365^2}$

I_{ijk}: large no. of indicator random variable $({}^nC_3)$

p very small $\frac{1}{(365)^2}$

I_{123} loosely dependent on I_{124}

$$P(I \geq 1) = 1 - P(X=0) = 1 - \frac{e^{-\lambda}\lambda^{0}}{0!} = 1 - e^{-\lambda}$$

where $\lambda = {}^{n}C_{3} \frac{1}{(365)^{2}}$

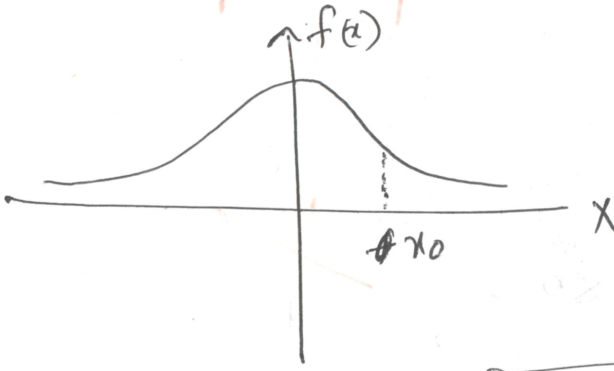
Continous Distribution

discrete	continous
X	X
PMF: $P(X=x)$	P.D.F $f_X(x)$
CDF: $P(X \leq x)$	CDF $P(X \leq x)$
$E(X) = \sum_{x} x P(X=x)$	$E(X) = \int_{-\infty}^{\infty} x f(x) \, dx$
$Var(X) = E(X^2) - E^2(X)$	$Var(X) = E(X^2) - E^2(X)$
LOTUS $\boxed{E(g(X)) = \sum_{x} g(x) P(X=x)}$	$\boxed{E(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f(x) \cdot dx}$

$\boxed{P(X=x) = 0}$

PDF (Probablity Density Function)

- Defn: A random variable X has PDF $f(x)$ if
$$P(a \leq X \leq b) = \int_{a}^{b} f(x) \, dx \quad \forall a, b$$
- To be valid $f(x) \geq 0$, $\int_{-\infty}^{\infty} f(x) dx = 1$.



$$f(x_0)\varepsilon \rightarrow P\left(X \in \left(x_0 - \frac{\varepsilon}{2}, x_0 + \frac{\varepsilon}{2}\right)\right)$$

(for ε very small)

as $\varepsilon \rightarrow 0$

$\varepsilon > 0$

If X has P.D.F f.

then C.D.F $P(X \leq x_0) = \int_{-\infty}^{x_0} f(x) \, dx$

If X has CDF: F

then P.D.F $f(x) = \left. \frac{dF(t)}{dt} \right|_{t=x}$

$$P(a<X<b) = \int_a^b f(x)\,dx = F(b)-F(a)$$

$F(x)$ is continous & differentiable
$f(x)$ is continous

Variance

: On average how far is X from its mean

$$\boxed{\begin{aligned} Var(X) &\overset{?}{=} E(X-E(X)) \\ Var(x) &\overset{?}{=} E(X) = E\,E(X) \\ Var(x) &\overset{?}{=} E(X) - E(X) \\ Var(x) &\overset{?}{=} 0 \end{aligned}}$$

not used therefore

$$\boxed{E(X-E(X))^2 = Var(X)}$$

$$S.D(X) = \sqrt{Var(X)}$$

Another way to express Variance.

$$Var(X) = E(X^2 - 2XE(X) + E^2(X))$$
$$Var(X) = E(X)^2 - 2E(X)E(X) + E^2(x)$$
$$Var(X) = E(X)^2 - E^2(X)$$
$$Var(X) = E(X^2) - E^2(X)$$

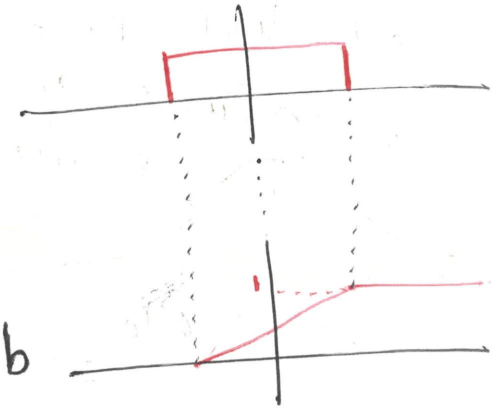
$Var(X) \geq 0$
$Var(X) = 0$ for X constant

Uniform distribution

$Unif(a,b)$

Completely random point in $[a,b]$

$$P.D.F \quad f(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$



$$F(x) = \begin{cases} 0, & x<a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x>b \end{cases}$$

$$E(X) = \int_a^b \frac{x}{(b-a)}\,dx = \frac{x^2}{2(b-a)}\Bigg|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$$

$$\boxed{E(X) = \frac{a+b}{2}}$$

Variance of uniform distribution

$E(X2) = ?$ Let $Y = X^2$ function of RV is an RV

$E(Y)$ need PDF Y

$$E(X2) = \int_{-\infty}^{\infty} x^2 f(x).dx$$

Law of the unconscious statistician (LOTUS)

will be proved later.

$$\boxed{E(g(x)) = \int_{-\infty}^{\infty} g(x) f(x) \cdot dx}$$

eg · $X \sim$ uniform$(0,1)$

$E(X) = 1/2$

$$E(X2) = \int_0^1 x^2 . 1\, du = \underline{\underline{1/3}}$$

$$V(X) = E(X2) - E(X)^2$$
$$= 1/3 - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$$

Uniform is Universal

$U \sim$ uniform$(0,1)$ Let F be a CDF (assume that F is strictly increasing & continous)

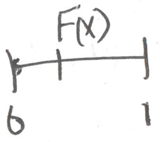
For Simulating Complex distribution

Theorm $\boxed{\text{Let } X = F^{-1}(U). \text{ Then } X \sim F}$

X follows CDF(F)   Function of RV is an RV

Proof:

$$P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x))$$



F(X) is a probablity

$$P(U \leq F(x)) = F(x)$$

$\boxed{\text{Also: If } X \sim F, \text{ then } F(X) \sim \text{Uniform}(0,1)}$ → To confirm the model of a CDF.

taken as a function & not a C.D.F

eg:- Let $F(x) = 1-e^{-x}$, $x > 0$, $u \sim$ Uniform $(0,1)$

Simulat $X \sim F$, $F^{-1}(u) = -\ln(1-u)$

$\therefore -\ln(1-u) \sim F$

$\ln\left(\frac{1}{1-u}\right) \sim F$

$1-u \sim$ Unif $(0,1)$

$\therefore \ln\left(\frac{1}{1-u}\right) = \ln\left(\frac{1}{u}\right)$

$u = 1-e^{-x}$

$1-u = e^{-x}$

$\ln(1-u) = -x$

$x = -\ln(1-u)$

$F^{-1}(u) = x = -\ln(1-u)$

$1-u \sim$ Uniform $(0,1)$

$a+bu \sim$ Uniform (someth..) Linear Transformation

*Non linear leads to non uniform.

Independence of R.V.

$\forall (X_1, X_2, X_3 \text{------} X_n)$

Defn They are independent if $P(X_1 < x_1, X_2 < x_2, \dots, X_n < x_n)$

$= P(X_1 < x_1)\, P(X_2 < x_2) \dots\dots P(X_n < x_n)$

$\forall (x_1, x_2 \dots x_n)$

Discrete case. $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P(X_1 = x_1) \cdot P(X_2 = x_2) \dots, P(X_n = x_n)$

$\forall (x_1, x_2, x_3 \dots x_n)$

Ex $X_1, X_2 \sim$ Bernoulli$(1/2)$, $X_3 = \begin{cases} 1 & \text{if } X_1 = X_2 \\ 0 & \text{otherwise} \end{cases}$

(X_1, X_2, X_3) These are pairwise independent but not independent

$X_1\ X_2$ — independent

$X_1\ X_3$ — independent

$X_2\ X_3$ — independent

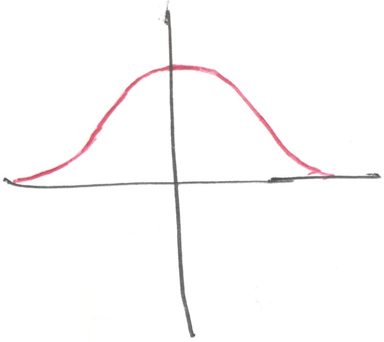
} X_3 depends on X_1 & X_2

Normal Distribution (Centre Limits Theorm)

Adding up a bunch of IID RVs give the normal RV

* Sum of a lot i.i.d. r.v.s looks normal

IID RV
independent & identically distributed Random variable.



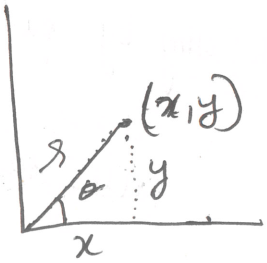
Standard normal = N(0,1) P.D.F =

$$f(z) = \underbrace{\frac{1}{\sqrt{2\pi}}}_{C} e^{\frac{-z^2}{2}}$$

To find C

$$I = \int_{-\infty}^{\infty} e^{-z^2/2} \cdot dz \quad . \quad \text{Taking} \int_{-\infty}^{\infty} e^{-z^2/2} \cdot dz \int_{-\infty}^{\infty} e^{-t^2/2} dt = I^2$$

$$I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2} \, dx \, dy$$



$r^2 = x^2 + y^2$

$$I^2 = \int_0^{2\pi}\int_0^{\infty} e^{-r^2/2} \, r dr d\theta = \int_0^{2\pi}\int_0^{\infty} e^{-u} \cdot du \cdot d\theta$$

$$I^2 = \int_0^{2\pi} \left[-e^{-u}\right]_0^{\infty} \cdot d\theta$$

$$I^2 = \int_0^{2\pi} d\theta = 2\pi$$

$$\boxed{\Rightarrow I = \sqrt{2\pi} \Rightarrow c = \frac{1}{\sqrt{2\pi}}}$$

Mean = 0

Variance = 1

Let $Z \sim N(0,1)$

$$E(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-\frac{z^2}{2}} dz = 0 \text{ (symmetry)}$$

$$\int_{-\infty}^{\infty} \text{odd} \cdot \text{even} = 0.$$

Variance of (z)

$$= E(z^2) - E^2(z) = E(z^2)$$

using <u>LOTUS</u> (Law of the unconscious statistician)

$$E(z^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{z^2 e^{-z^2/2}}_{\text{even}} dz = \frac{2 \cdot 1}{\sqrt{2\pi}} \int_0^{\infty} z^2 e^{-\frac{z^2}{2}} dz = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} \underbrace{z}_{u} \cdot \underbrace{z e^{-\frac{z^2}{2}} dz}_{dv}$$

integrating by parts

$u = z \qquad dv = z e^{-z^2/2}$

$du = dz \qquad v = -e^{-z^2/2}$

$$\Rightarrow \frac{2}{\sqrt{2\pi}} \left((uv) \Big|_0^{\infty} - \int_0^{\infty} v\, du \right)$$

$$= \frac{2}{\sqrt{2\pi}} \left[-z e^{-z^2/2} \Big|_0^{\infty} + \int_0^{\infty} e^{-z^2/2} dz \right]$$

$$\Rightarrow \left(\frac{2}{\sqrt{2\pi}} \left(0 + \frac{\sqrt{2\pi}}{2} \right) \right.$$

$$= 1 //$$

$$\Rightarrow V(z) = 1 = E(z^2)$$

Notation Φ is the standard normal CDF

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

$$\boxed{\Phi(-z) = 1 - \Phi(z)}$$

Summary

$Z \sim N(0,1)$

CDF Φ

$E(Z)=0$ 1st moment

$Var(Z)=E(Z^2)=1$ 2nd moment

$E(Z^3)=0$ 3rd moment

$$\boxed{\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^3 e^{-\frac{z^2}{2}}\,dz=0}$$

LOTUS

All odd moments are zero

$-Z \sim N(0,1)$ —— Symmetry.

Let $\boxed{X=\mu+\sigma Z}$ function of RV also an RV

$\mu \in \mathbb{R}$ (mean, location)

$\sigma > 0$ (S.D, $=\sqrt{Var}$)

We say $X \sim N(\mu,\sigma^2)$

$E(X)=E(\mu+\sigma Z)=\underline{\underline{\mu}}$

$Var(\mu+\sigma Z)\overset{?}{=}\sigma^2 Var(Z)=\underline{\underline{\sigma^2}}$

} confirmed

Variance

$Var(X)=E([X-E(X)]^2)=E(X^2)-E^2(X)$

1) $Var(X+c)=Var(X)$
2) $Var(cX)=c^2\,Var(X)$
3) $Var(c)=0$
4) $Var(X+Y)\neq Var(X)+Var(Y)$
5) $Var(X+Y)=Var(X)+Var(Y)$
   ↳ if X & Y independent
6) $Var(a+cX)=c^2\,Var(X)$

$$\boxed{Z=\frac{X-\mu}{\sigma}}$$

Standardisation

Assuming we know P.DF of normal
How to get P.D.F of Normal.

C.D.F: $P(X\leq x)=P\left(\frac{X-\mu}{\sigma}\leq\frac{x-\mu}{\sigma}\right)$

CDF: $P\left(Z\leq\frac{x-\mu}{\sigma}\right)$

CDF $=\Phi\left(\frac{x-\mu}{\sigma}\right)$

P.DF is derivative of C.DF by chain rule.

P.D.F $=\frac{1}{\sigma\sqrt{2\pi}}e^{-\left(\frac{x-\mu}{\sigma}\right)^2/2}$

$-X=-\mu+\sigma(-Z)$

$-X=-\mu+\sigma Z \sim Normal(-\mu,\sigma^2)$

$X \sim N(\mu,\sigma)$

$-X \sim N(-\mu,\sigma)$

Later we will show if $X_j \sim N(\mu_j,\sigma_j^2)$ independent.

then $\boxed{X_1+X_2 \sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)}$

$X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$ if X_1 & X_2 are independent.

$E(X_1) + E(-X_2) = \mu_1 - \mu_2$

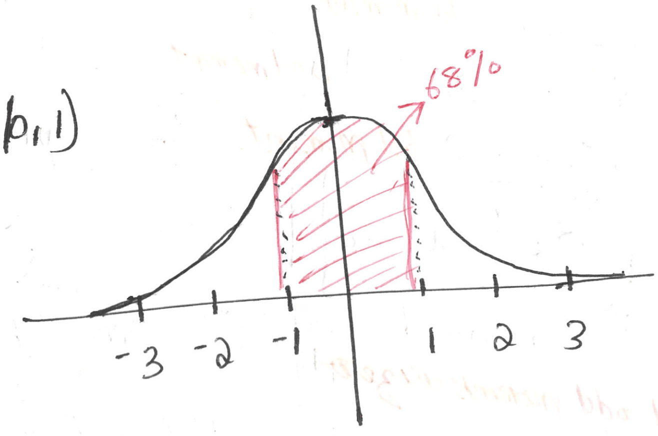
$V(X_1) + V(-X_2) = \sigma_1^2 + \sigma_2^2$

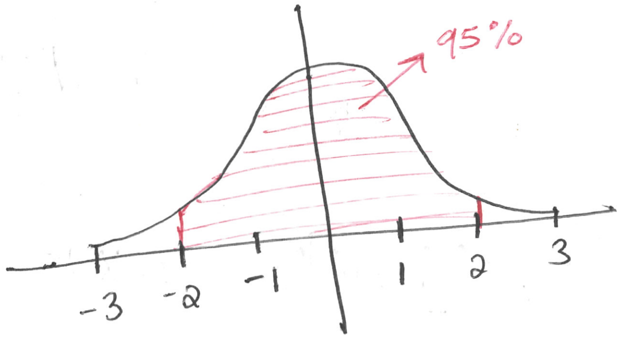
68-95-99.7% Rule $X \sim N(0,1)$

$P(|X-\mu| \leq \sigma) = 68\%$

$P(|X-\mu| \leq 2\sigma) \approx 95\%$

$P(|X-\mu| \leq 3\sigma) \approx 99.7\%$





Variance of Poisson distribution

$X : 0, 1, 2, 3, \text{-----}$

$P(X=x) : P(0)\ P(1) \ldots\ldots$

$x^2 : 0, 1, 4, 9, \text{-----}$

$P(X^2=x) : P(0)\ P(1)\ P(2)$

$E(X) = \sum_x x P(X=x)$

$E(X^2) = \sum_x x^2 P(X=x)$

even if $X \to X^2$ is ont one to one but many to one.

$\therefore E(X^2)$ for poisson $= \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda}\lambda^x}{x!}$

$E(X^2) = e^{-\lambda} \sum_{x=1}^{\infty} \frac{x^2 \lambda . \lambda^{x-1}}{x!}$

~~$E(X)^2$~~ $E(X^2) = e^{-\lambda} \cdot \lambda \sum_{x=1}^{\infty} \frac{x^2 \cdot \lambda^{x-1}}{k!}$

$E(X^2) = e^{-\lambda} \cdot \lambda \cdot e^{\lambda}(\lambda+1)$

$E(X^2) = \lambda(\lambda+1)$

$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda} \ \forall \lambda$

$\sum_{k=1}^{\infty} \frac{k \lambda^{k-1}}{k!} = e^{\lambda}$

$\sum_{k=1}^{\infty} \frac{k \lambda^k}{k!} = \lambda e^{\lambda}$

$\sum_{k=1}^{\infty} \frac{k^2 \lambda^{k-1}}{k!} = \lambda e^{\lambda} + e^{\lambda}$

$\sum_{k=1}^{\infty} \frac{k^2 \lambda^{k-1}}{k!} = e^{\lambda}(\lambda+1)$

$\boxed{Var = E(X^2) - E^2(X) = \lambda}$

Variance of Binomial

$X \sim Bin(n,p)$

Binomial can thought of sum of n independent bernoulip.

$$X = X_1+X_2+X_3 - - - - + X_n$$

$Var(X) = Var(X_1+X_2+X_3+ - - - X_n)$ where $X_i \sim$ bernouli(p)

$$X^2 = X_1^2+X_2^2+X_3^2+ - - - X_n^2 + 2X_1X_2+2X_1X_3+ - - - - 2X_{n-1}X_n$$

$$X^2 = \sum_i X_i^2 + \sum_{\substack{i<j \\ i\neq j}} (2X_iX_j)$$

$I_1^2 = I_1$

I_1I_2 = indicates of success on both.

$$E(X^2) = nE(I_1^2) + nC_2 \cdot 2E[I_1I_2]$$

$$E(X^2) = nE(I_1) + n(n-1)E[I_1I_2]$$

$$E(X^2) = np + n(n-1)p^2$$

$$E(X^2) = np + n^2p^2 - np^2$$

$$E^2(X) = n^2p^2$$

$$Var(X) = E(X^2) - E^2(X) = np - np^2 = np[1-p] = \boxed{npq = Var(X) \text{ Binomial}}$$

Prove LOTUS for discrete sample space

$$E(g(X)) = \sum_x g(x) = P(X=x)$$

$$E(g(x)) = \sum_x g(x) \underbrace{P(X=x)}_{\text{grouped}} = \sum_{s\in S} g(X(s)) \cdot \underbrace{P(\{s\})}_{\text{ungrouped}}$$

$$E(g(x)) = \sum_x \sum_{s:(X(s)=x)} g(X(s)) \cdot P\{s\} =$$

$$E(g(x)) = \sum_x \sum_{s:X(s)=x} g(x) \cdot P\{s\} = \sum_x g(x) \cdot \sum_{s:X(s)=x} P\{s\}$$

$$E(g(x)) = \sum_x g(x) \cdot P(X=x)$$

Review

Problem

Coupon Collector (toy collector)   n toytypes equally likely

Q Find the expected time to collect full set.

ie No of toys until you have a full set.

\# toys = T

$T = T_1 + T_2 + T_3 + T_4 + ---- T_N$

T_1 = time until first new toy = 1

T_2 = time additional until second new toy

T_j = ~~time~~ additional time until n^{th} new toy.

$T_1 = 1$

$T_2 - 1 \sim geo\left(\frac{n-1}{n}\right)$

$$\boxed{T_j - 1 \sim geo\left(\frac{n-(j-1)}{n}\right)}$$

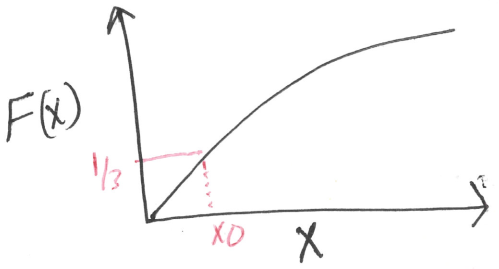
$E(T) = E(T_1) + E(T_2) + -- E(T_N)$

$E(T) = 1 + \frac{n}{(n-1)} + \frac{n}{(n-2)} + \frac{n}{(n-3)} + --- n$

$$\boxed{E(T) = n\left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + --- \frac{1}{n}\right)}$$

$E(T) \approxeq n\log n$ for large n

Universality of the Uniform distribution



$F(x_0) = 1/3$

$= P(F(X) < 1/3)$

$= P(X < x_0)$

$= F(x_0) = 1/3$

$F(X) \sim$ Uniform $(0,1)$

$X \sim F$ C.D.F

eg of application

Logistic distribution: $F(x) = \frac{e^x}{1+e^x} = u$

Let u = Uniform $(0,1)$

$F'(u) = g(u) = \log\left(\frac{u}{1-u}\right)$

Symmetry

X, Y, Z be iid positive Random variable. Find $E\left(\frac{X}{X+Y+Z}\right)$

$$\left.\begin{array}{l} E\left(\frac{X}{X+Y+Z}\right) = E\left(\frac{Y}{X+Y+Z}\right) = E\left(\frac{Z}{X+Y+Z}\right) \\ E\left(\frac{X}{X+Y+Z}\right) + E\left(\frac{Y}{X+Y+Z}\right) + E\left(\frac{Z}{X+Y+Z}\right) = E(1) \end{array}\right\} \Rightarrow E\left(\frac{X}{X+Y+Z}\right) = \underline{\underline{1/3}}.$$

$3I = 1$

LOTUS

$$\left.\begin{array}{l} U \sim \text{uniform}(0,1) \\ X = U^2 \\ Y = e^X \end{array}\right\}$$ Find $E(Y)$ as an integral

$$E(Y) = \int_0^1 e^X \cdot P.DF(x)$$

$$E(Y) = \int_0^1 e^{u^2} \cdot du$$

CDF of X

$P(X \le x)$

$= P(U^2 \le x)$

$= P(U \le \sqrt{x})$

$P(X \le x) = \sqrt{x}$ since $U \sim$ Uniform$(0,1)$

PDF $= f(x) = P(X = x) = \frac{1}{2} x^{-1/2} = \underline{\underline{\frac{1}{2\sqrt{x}}}}$

---

practice:

$X \sim Bin(n,p)$, then find distribution of $n - X$

$$P(n-X=k) = P(X=n-k) = {}^nC_{(n-k)}\, p^{n-k} q^{k} = {}^nC_k\, q^k p^{n-k}$$

$\Rightarrow (n-X) \sim Bin(n,q)$

* Story − $(n-X)$ is $Bin(n,q)$ by interchanging success & failure.

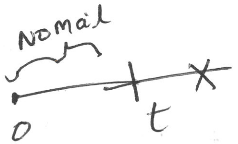
if X → #success

then n−X → #failure

Ex)

#emails I get in time t is Poison (λt)

Find P.D.F of T time of first email.



Solution:

$P(T>t) = P(N_t=0)$

1 - CDF

$\text{C.D.F} = P(T \leq t)$
↓
definition of C.D.F

with $N_t = \#$ of emails in interval $[0,t]$

$$P(N_t) = \frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t}$$

$\Rightarrow \text{CDF} = 1 - e^{-\lambda t}$

Exponential Distribution, rate parameter λ

$X \sim \text{Expo}(\lambda)$ has P.D.F $\begin{cases} \lambda e^{-\lambda x} & , x>0 \\ 0 & , \text{otherwise} \end{cases}$

$$\text{C.D.F} \int_0^x \lambda e^{-\lambda t} dt = \lambda \frac{e^{-\lambda t}}{-\lambda}\Big|_0^x = \begin{cases} \underline{\underline{1-e^{-\lambda x}}} & , (x>0) \\ 0 & , \text{otherwise} \end{cases}$$

Let $Y = \lambda X$ if $X \sim \text{Expo}(\lambda)$

then $Y \sim \text{Expo}(1)$ Standard exponential.

$\text{CDF of } Y = P(Y \leq y)$

$= P(\lambda X \leq y)$

$= P(X \leq \frac{y}{\lambda}) = \underline{\underline{1-e^{-y}}}$

~~$\text{PDF of } Y = P(Y=y)$~~

~~$= P(\lambda X = y)$~~

~~$= P(X = y/\lambda)$~~

Let $Y \sim \text{Expo}(1)$, Find $E(Y)$, $Var(Y)$

$$E(Y) = \int_0^\infty y e^{-y} dy = \left(y(-e)^{-y} + \int_0^\infty e^{-y} dy\right)\Big|_0^\infty$$

$$E(Y) = e^{-y}\Big|_0^\infty = \underline{\underline{+1}}$$

$Var(X) = E(X^2) - E^2(X)$

$E(Y^2) = \int_0^\infty y^2 e^{-y} dy = y^2 \cdot -e^{-y} - \int 2y \cdot -e^{(-y)}$

$= y^2 \cdot -e^{-y} + 2\int_0^\infty y e^{-y}$

$= -y^2 e^{-y} - 2y e^{-y} + 2\int_0^\infty e^{-y} dy = 2$

$= \left| y^2 \cdot -e^{-y} - 2y \cdot (e^{-y}) + 2 \cdot -e^{-y} \right|_0^\infty$

$= 2$

$E(Y^2) = 2 \Rightarrow Var(X) = 2 - 1 = \underline{\underline{1}}$

$Y \sim Expo(1)$

$E(Y) = 1$

$V(Y) = 1$

$Y = \lambda X$

$\Rightarrow X = \left(\frac{Y}{\lambda}\right)$

$\Rightarrow E(X) = 1/\lambda$

$Var(X) = 1/\lambda^2$

$\rightarrow$
$$\boxed{\begin{array}{l} X \sim Expo(\lambda) \\ E(X) = 1/\lambda \\ V(X) = 1/\lambda^2 \end{array}}$$

<u>Memoryless property</u>

$P(X \geq s+t \mid X \geq s) = P(X \geq t)$

Let $X \sim Expo(\lambda)$

$P(X \geq s) = 1 - P(X \leq s)$

$= 1 - (1 - e^{-\lambda s}) = e^{-\lambda s}$

$$P(X \geq s+t / X \geq s) = \frac{P(X \geq s+t, \overset{\substack{\text{Now redundant} \\ \uparrow}}{X \geq s})}{P(X \geq s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = \underline{\underline{e^{-\lambda t}}}$$

$X \sim Expo\ \lambda$

$E(X / X > a) = a + E((X-a)/X > a)$

$= a + 1/\lambda$ by memorylessness.

~~Exponential MGF~~ , ~~find MGF, moments~~

~~$M(t) = E(e^{tx})$~~

Theorm

If X is a continuous random variable (positive) with the memoryless property, then $X \sim \text{Expo}(\lambda)$ for some λ.

Proof:

Let F be the CDF as usual, $G(x) = P(X>x) = 1-F(x)$

memory less property says

$$\boxed{G(s+t) = G(s)G(t)}$$

$\Downarrow$

Solving for G

~~$G(s+t)$~~

~~$= P(X>s+t)(X>s)$~~

~~$= \frac{P(X>s+t \cap X>s)}{P(X>s)}$~~

$$\boxed{\begin{aligned} G(s+t) &= P(X>s+t) \\ G(s+t) &= P(X>s+t \mid X>s) \cdot P(X>s) \\ G(s+t) &= \overset{\text{Memoryless.}}{P(X>t)} \cdot P(X>s) \\ G(s+t) &= G(t) \cdot G(s) \end{aligned}}$$

Consequences

$s=t \Rightarrow G(2t) = G(t)^2$

$s=(n-1)t \Rightarrow G(nt) = G(t)^n$

$t \to \frac{t}{m} \Rightarrow G(\frac{n}{m}t) = G(t)^{n/m}$

$\frac{n}{m} = x \Rightarrow G(xt) = G(t)^x$

$t=1 \Rightarrow G(x) = G(1)^x = e^{x \log G(1)}$

$\log G(1) < 0$ as G is a probablity $\in (0,1)$

$\Rightarrow G(x) = e^{x -\lambda}$

where $-\lambda = \log G(1)$

$G(x) = e^{-\lambda x}$

3 C.DF $= 1 - G(x)$

C.DF $= 1 - e^{-\lambda x}$

P.DF $= \underline{\underline{\lambda e^{-\lambda x}}}$

Moment generating Function (MGF)

Definition. A R.V X has MGF $M(t) = E(e^{tX})$, as a function of t, if $M(t)$ is finite on $\in (-a, a)$ about 0; $a > 0$

→ Why moment "generating"?

$$E(e^{tx}) = \underbrace{E\left(\sum_{n=0}^{\infty}\left(\frac{x^n t^n}{n!}\right)\right) = \sum_{n=0}^{\infty}\left[\frac{E(x^n)\, t^n}{n!}\right]}_{\text{due to linearity } \Sigma \,\&\, E \text{ is swapped}}$$

$\xrightarrow{}$ n^{th} moment

→ Why is MGF important. Let X has M.G.F $M(t)$

(1) The n^{th} moment $E(x^n)$ is the coefficient of $\frac{t^n}{n!}$ in taylor series of M,
ie, $\boxed{M^n(0) = E(x^n)}$

(2) M.G.F determines the distribution
if X & Y have same M.G.F then they have same distribution

(3) If X has M.G.F M_X, Y has M.G.F M_Y & if they are independent, the MGF of $X+Y$ is $E(e^{t(x+y)}) = E(e^{tx})E(e^{ty})$

False if not independent

$E[XY] = E(X)\,E(Y)$
$E[f(x)\cdot g(y)] = E(f(x))\cdot E(g(y))$

$\rightarrow \boxed{MGF_{X+Y} = M_X(t)\cdot M_Y(t)}$

Ex) $X \sim Bern(p)$, $M(t) = E(e^{tx}) = pe^t + q$

$X \sim Bin(n,p)$ $M(t) = (pe^t + q)^n$

by considering
$X = X_1 + X_2 + X_3 \ldots$
& Rule 3 of MGF

$Z \sim N(0,1)$

$$\text{M.G.F, } M(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz - z^2/2} \cdot dz$$

$$M(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-t)^2 + \frac{t^2}{2}} dz$$

$$M(t) = \frac{e^{t^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-1/2(z-t)^2} \cdot dz$$

$$M(t) = \frac{e^{t^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-t)^2} dz$$

put $z - t = m$
$dz = dm$

$$M(t) = e^{t^2/2} \cdot \underline{\left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-1m^2}{2}} dm \right)}$$

Gausian integral

$$M(t) = e^{t^2/2}$$

Laplaces rule of Succesion

$X_1, X_2, X_3 \dots\dots X_n$ Bern(p) — i.i.d

given P iid → X_i → indicator random variable representing sun rising on i^{th} day.

if p is unknown

Baseian approach is to treat p as R.V update prior based on base rule.

Let $p \sim$ Uniform $(0,1)$ as prior

Let $S_n = X_1 + X_2 + \dots\dots X_n$

So $(S_n / p) \sim$ Binomial(n, p), $p \sim$ Uniform$(0,1)$

Find the posterior distribution of $P \rightarrow P/S_n$ and $P(X_{n+1} = 1 / S_n = n)$

probablity that sun will rise tmrw

$$f(p \mid S_n = k) = \frac{P(S_n = k/p)\, f(p)}{P(S_n = k)}$$

→ prior is uniform distribution $u(0,1) = 1 \ \forall p$

Bayes rule for P.D.F not for probability.

$P(S_n=k) = \int_0^1 P(S_n=k/p)\, f(p)\,dp$ is a constant a doesn't depend on p.

$P(S_n=k/p) = {}^nC_k\, p^k q^{(n-k)}$ —— since $S_n=k/p \sim Bin(n,p)$

$f(p|S_n=k) \underline{\underline{=}} \frac{{}^nC_k\, p^k q^{(n-k)} \cdot u(0,1)}{\text{constant}}$

Special case

$f(p|S_n=n) = \frac{{}^nC_n\, p^n \cdot 1}{\text{Constant}}$

$\therefore f(p|S_n=n) = p^n \cdot c$

$\therefore \boxed{f(p|S_n=n) = (n+1)p^n}$

where c = some constant to normalise.

$\int_0^1 f(p/S_n=n)\,dp = 1$

$\Rightarrow \int_0^1 cp^n\,dp = 1$

$= \left[\frac{cp^n}{(n+1)}\right]_0^1 = \frac{c}{(n+1)} = 1$

$\Rightarrow c = n+1$

$P(X_{n+1}=1/S_n=n) = E(p/S_n=n)$

$= \int_0^1 (n+1)\, p \cdot p^n\,dp = \frac{n+1}{n+2}$

∴ The probability that the sun will rise the next day after it has consistently rose for n days is $\frac{n+1}{n+2}$

Exponential MGF $X \sim Expo(1)$

$M(t) = M_X = E(e^{tX})$

$M_X = \int_0^\infty e^{tx} e^{-x}\,dx$

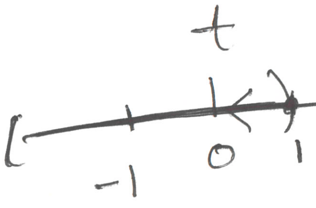
$M_X(t) = \int_0^\infty e^{-x(1-t)}\,dx$

$M_X(t) = \frac{1}{(1-t)} \quad i(t<1)$

$M'(0) = E(X)$

$M''(0) = E(X^2)$

$M'''(0) = E(X^3)$



$\frac{1}{1-t}$ = geometric series

$\frac{1}{1-t} = 1+t+t^2+t^3+t^4+\dots$

$\frac{1}{1-t} = \sum_{n=0}^{\infty} t^n \quad (t<1)$

$$M_X(t) = \sum_{n=0}^{\infty} \frac{E(X^n)\, t^n}{n!} = \sum_{n=0}^{\infty} t^n$$

$$M_X(t) = \sum_{n=0}^{\infty} \frac{E(X^n)\, t^n}{n!} = \sum_{n=0}^{\infty} \frac{n!\, t^n}{n!}$$

$$\Rightarrow \boxed{E(X^n) = n! \quad \forall n}$$

$Y \sim$ Exponential (λ)

$X = \lambda Y \sim$ Expo(1)

so $Y^n = \frac{X^n}{\lambda^n}$

$$\Rightarrow \boxed{E(Y^n) = \frac{n!}{\lambda^n}}$$

Let $Z \sim N(0,1)$, Find all moments.

$E(Z^n) = 0$ for n odd

$E(Z) = 0$

$E(Z^2) = 1$

$M_Z(t) = E(e^{tZ})$

$M_Z(t) = e^{t^2/2}$ — previous result.

$$e^{t^2/2} = \sum_{n=0}^{\infty} \frac{(t^2/2)^n}{n!} = \sum_{n=0}^{\infty} \left(\frac{t^{2n}}{2^n\, n!} \right)$$

$$M(t) = \sum_{n=0}^{\infty} \frac{E(\cancel{X}^n)\, t^n}{n!} = e^{t^2/2} = \sum_{n=0}^{\infty} \frac{(2n)!\, t^{2n}}{2^n\, n! \cdot 2n!}$$

$$M(t) = \sum_{n=0}^{\infty} \frac{E(\cancel{X}^{2n})\, t^{2n}}{(2n)!}$$

$$\Rightarrow E(\cancel{X}^{2n}) = \frac{(2n)!}{2^n\, n!}$$

$$E(Z^2) = \frac{2}{2} = \underline{\underline{1}}$$

$$E(Z^4) = \frac{24}{4 \cdot 2} = \underline{\underline{3}}$$

$$E(Z^6) = \frac{6!}{8 \cdot 3!} = \underline{\underline{15}}$$

M.G.F of poison distribution

$X \sim \text{Poisson}(\lambda)$ $\quad E(e^{tX}) = \sum_{k=0}^{\infty} e^{tk}\cdot\frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{e^{tk}\lambda^k}{k!}$

$$M(t) = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(e^t\lambda)^k}{k!} = e^{-\lambda}\cdot e^{e^t\lambda} = \underline{\underline{e^{\lambda(e^t-1)}}}$$

Poison M.G.F. $M(t) = e^{\lambda(e^t-1)}$

Let $Y \sim$ poisson(μ)

Let X & Y are independent. Find Distribution of $X+Y$.

From Rule 3 of MGF just multiply

$$M_{X+Y}(t) = M_X(t)\cdot M_Y(t) = e^{\lambda(e^t-1)}\cdot e^{\mu(e^t-1)} = e^{(e^t-1)(\mu+\lambda)}$$

$M_{X+Y} =$ poison $(\lambda+\mu)$

$\Rightarrow$ $X+Y \sim$ poisson$(\lambda+\mu)$

Joint Distribution

X,Y are bernouli may be dependent/independent.

		$Y=0$	$Y=1$
$X=0$	$\frac{1}{2}$	$\frac{2}{6}$	$\frac{1}{6}$
$X=1$	$\frac{1}{2}$	$2/6$	$1/6$
		$2/3$	$1/3$

Independent

joint PDF (continous)

$f(x,y)$ such that

$$\boxed{P((x,y)\in B) = \iint_B f(x,y)\,dx\,dy}$$

X,Y R.Vs

joint CDF

$F(x,y) = P(X\le x, Y\le y)$

joint PMF

$P(X=x, Y=y)$

} Discreate.

Marginal distribution

$P(X\le x)$ is the marginal.

Independence

if $F(x,y) = F_X(x)\,F_Y(y)$

joint CDF = product of marginal C.D.F

Equivalent

$\Rightarrow$ $P(X=x, Y=y) = P(X=x)\cdot P(Y=y)$

$\Rightarrow$ $f(x,y) = f_X(x)\,f_Y(y)$ $\forall x, y$

→ Getting marginal from joint.

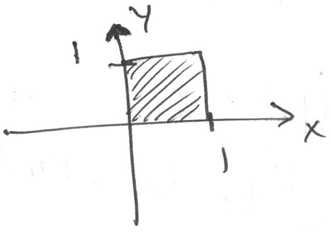
$$P(X=x) = \sum_{y} P(X=x, Y=y).$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\, dx$$

	Y=0	Y=1	
X=0	1/2	0	1/2
X=1	1/4	1/4	1/2
	3/4	1/4	

dependent

Eg.) Uniform on square $\{(x,y) : x,y \in [0,1]\}$



$$\text{P.D.F} = \begin{cases} c, & x \,\&\, y \in [0,1] \\ 0, & \text{otherwise} \end{cases}$$

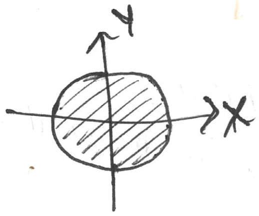
joint P.D.F is a constant on the square & zero outside the square.

where $c = \frac{1}{\text{Area of Region}}$

X, Y independt

$$\text{Marginal } f_X(x) = \int_0^1 c\, dx = \frac{1}{\text{length}}$$

Ex Uniform in disc $x^2+y^2 \leq 1$



$$\text{joint pdf} = \begin{cases} \frac{1}{\pi}, & x^2+y^2 \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

X & Y are dependent

given $X=x$ $\quad y \in \left(-\sqrt{1-x^2}, \sqrt{1-x^2}\right)$

Joint, Conditional, Marginal Distributions

joint CDF $F(x,y) = P(X \le x, Y < y)$

Continuous case:

joint PDF : $f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$

$$P(X,Y \in A) = \iint_A f(x,y)\,dx\,dy$$

Marginal P.DF. $x = \int_{-\infty}^{\infty} f(x,y)\,dy$

Marginal P.D.F $y = \int_{-\infty}^{\infty} f(x,y)\,dx$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$$

Conditional P,DF

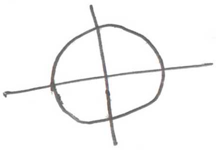
P.D.F of Y/x

$$f(y/x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

$$f(y/x) = \frac{f(x/y)\cdot f(y)}{f_X(x)}$$

---

eg:



$$f(x,y) = \begin{cases} \frac{1}{\pi}, & x^2+y^2=1 \\ 0, & \text{outside} \end{cases}$$

$$f_X(x) = \int_{y=-\sqrt{1-x^2}}^{y=+\sqrt{1-x^2}} \frac{1}{\pi}\cdot dy = \frac{1}{\pi}\left(\sqrt{1-x^2}+\sqrt{1-x^2}\right) = \underline{\underline{\frac{2\cdot\sqrt{1-x^2}}{\pi}}}, \quad x \in (-1,1$$

$$f(y/x) = \begin{cases} \dfrac{\frac{1}{\pi}}{\frac{2}{\pi}\sqrt{1-x^2}}, & y \in (-\sqrt{1-x^2}, \sqrt{1-x^2}) \\ 0, & \text{otherwise.} \end{cases} = \begin{cases} \frac{1}{2\sqrt{1-x^2}}, & y \in (-\sqrt{1-x^2}, \sqrt{1-x^2}) \\ 0, & \text{otherwise} \end{cases}$$

$f(y/x) = \text{Uniform}(-\sqrt{1-x^2}, \sqrt{1-x^2})$

$f(y/X=x) \sim \text{Uniform}(-\sqrt{1-x^2}, \sqrt{1-x^2})$

$f(X/Y=y) = \text{Uniform}(-\sqrt{1-y^2}, \sqrt{1-y^2})$

$f(X=x, Y=y) \neq f_X(x)\cdot f_Y(y)$

2-D Lotus

Let (X,Y) have a joint P.D.F: $f(x,y)$ & let $g(x,y)$ be any real function X,Y. Then $E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f(x,y)\, dx\,dy$

---

eg.) If X & Y are independent $E(XY) = E(X)E(Y)$

("Independence implies uncorrelated")

Proof in continuous case.

2D LOTUS — $E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)\, dx\,dy$

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_X(x) f_Y(y)\, dx\,dy = \int_{-\infty}^{\infty} y f_Y(y)\,dy \int_{-\infty}^{\infty} x f_X(x)\,dx$$

$$E(XY) = E(X)\,E(Y)$$

---

eg. Let $X,Y \sim$ i.i.d Uniform $(0,1)$ find $E[|X-Y|]$

independent & identically distributed.

LOTUS $\int_0^1\int_0^1 |x-y|\, dx\,dy = \iint_{x>y} (x-y)\cdot dx\,dy + \iint_{x<y} y-x\cdot dy\,dx$

$$= 2\cancel{x}\cdot \int_0^1\int_y^1 (x-y)\, dx\,dy = 2\int_0^1 \left[\frac{x^2}{2} - yx\right]_y^1 dy$$

$$= 2\cdot\int_0^1 \frac{1}{2} - y - \left(\frac{y^2}{2} - y^2\right) dy = 2\int_0^1 \left(\frac{1}{2} - y + \frac{y^2}{2}\right) dy$$

$$= 2\cdot\left[\frac{1}{2}y - \frac{y^2}{2} + \frac{y^3}{6}\right]_0^1$$

$$= 2\cdot\left(\frac{1}{2} - \frac{1}{2} + \frac{1}{6}\right) = \underline{\underline{1/3}}$$

Let $M = \text{Max}(X,Y)$

Let $L = \text{Min}(X,Y)$

$M-L = |X-Y|$

$E(M-L) = E(|X-Y|)$

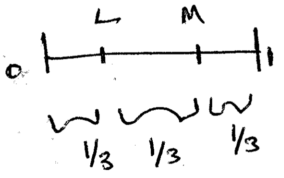
$$\left.\begin{aligned} E(M) - E(L) &= 1/3 \\ E(M)+E(L) &= 1 \end{aligned}\right\}$$

$E(M+L) = E(X+Y)$

$E(M)+E(L) = E(X)+E(Y)$

$E(M) = 2/3 ; \ E(L) = 1/3$

ON AVERAGE



Chicken & Egg problem

N eggs

N is random ~ Poison(λ)

each hatches with probablity p

and independt on others

Let X = #hatches

$(X/N=n) \sim \text{Binomial}(n,p)$

Let Y = # don't hatch

$X+Y = N$

Q) joint PMF of X & Y?

Q) Are they independent

Solution

$$P(X=i, Y=j) = \sum_{n=0}^{\infty} P(X=i, Y=j/N=n) \cdot P(N=n) \quad \text{L.O.T.P}$$

$$P(X=i, Y=j) = P(X=i, Y=j / N=i+j) \cdot P(N=i+j)$$

$$P(X=i, Y=j) = P(X=i/N=i+j) \cdot P(N=i+j)$$

$$P(X=i, Y=j) = {}^{i+j}C_i \, p^i q^j \cdot \frac{e^{-\lambda} \lambda^{(i+j)}}{(i+j)!}$$

$$P(X=i, Y=j) = \frac{(i+j)!}{i!\, j!} \frac{p^i q^j e^{-\lambda} \lambda^{(i+j)}}{(i+j)!} = \frac{p^i q^j e^{-\lambda} \lambda^{i+j}}{i!\, j!}$$

$$P(X=i, Y=j) = \frac{(p\lambda)^i}{i!} \cdot \frac{(q\lambda)^j}{j!} \cdot e^{-\lambda}$$

$$P(X=i, Y=j) = \frac{e^{-\lambda p}(\lambda p)^i}{i!} \cdot \frac{e^{-\lambda q}(\lambda q)^j}{j!}$$

$P(X=i, Y=j) \Rightarrow$ factored $\Rightarrow$ They are independent

$X \sim \text{poisson}(\lambda p)$

$Y \sim \text{poisson}(\lambda q)$

Eg:- Find $E|z_1-z_2|$ iid standard normal z_1 & z_2

Theorm: $\left.\begin{matrix} X\sim N(\mu_1,\sigma_1^2) \\ Y\sim N(\mu_2,\sigma_2^2)\end{matrix}\right\}$ independent then $X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$

Proof Use M.G.F M.G.F of $(X+Y) = MGF_X \cdot MGF_Y$

$$MGF_{X+Y} = e^{\mu_1 t+\frac{1}{2}\sigma_1^2 t^2} e^{\mu_2 t+\frac{1}{2}\sigma_2^2 t^2}$$

$$MGF_{X+Y} = e^{(\mu_1+\mu_2)t+\frac{1}{2}(\sigma_1^2+\sigma_2^2)t^2}$$

$$MGF_{(X+Y)} = MGF \text{ of } N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$$

$$\Rightarrow X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$$

$E|z_1-z_2|$

$z_1-z_2\sim N(\mu_1-\mu_2,\sigma_1^2+\sigma_2^2)$

for normal z_1 & z_2 standard·

$z_1-z_2\sim N(0,2)$ Note $z_1-z_2=\sqrt{2}z$

$E|z_1-z_2| = E|\sqrt{2}z|$ ~~$\sim N(0,2)$~~ where $z_1,z_2,z\sim N(0,1)$

$$E|z_1-z_2| = \sqrt{2}\,E|z|$$

$$= \sqrt{2}\int_{-\infty}^{\infty}\underbrace{|z|\frac{1}{\sqrt{2\pi}}e^{\frac{-z^2}{2}}}_{\text{even function}}dz = 2\sqrt{2}\cdot\int_0^{\infty} z\frac{1}{\sqrt{2\pi}}e^{\frac{-z^2}{2}}dz$$

put $\frac{z^2}{2}=u$

$dz\cdot z = du$

$$E|z_1-z_2| = \frac{2\sqrt{2}}{\sqrt{2\pi}}\int_0^{\infty}e^{-u}du = \frac{2\sqrt{2}}{\sqrt{2\pi}}\left[e^{-0}-e^{-\infty}\right] = \frac{2}{\sqrt{\pi}} \text{ or } \frac{\sqrt{2}}{\sqrt{\pi}}$$

Multinomial [$\text{Mult}(n,\vec{p})$]

$\vec{p}$ is a vectos

$\vec{p} = (p_1, p_2 \text{----} p_k)$

probablity vectos.

$\sum_k p_k = 1 ; \; p_k \geq 0$

Denifition & story

$\vec{X} \sim \text{Mult}(n,\vec{p})$

$\vec{X} = (X_1, X_2, \ldots X_k)$

we have n independent trials.

they are in put in k categories independently

P_j = probablity ~~of the~~ of any trial becoming category ~~k~~ j

X_j = No. of occurance of category j

23331111222

Joint PMF

$$p(X_1=n_1, X_2=n_2 \ldots X_k=n_k) = \frac{n!}{n_1! n_2! n_k!} p_1^{n_1} p_2^{n_2} p_3^{n_4}$$

if $n_1+n_2+ \ldots n_k = n$

0, otherwise

Q) $\vec{X} \sim \text{Mult}_k(n,\vec{p})$. Find

Marginal distribution of X_j

then $X_j \sim \text{Binomial}(n, p_j)$

$E(X_j) = n p_j$

$Var(X_j) = n p_j (1-p_j)$

$Var(X_j) = n p_j \cdot \sum_{k \neq j} (p_k)$

Lumping property.

let $k=10$

$\vec{X} = (X_1, \ldots X_{10}) \sim \text{Mult}_{10}(n, (p_1 \ldots p_{10}))$

let $\vec{Y} = (X_1, X_2, (X_3 + \ldots X_{10}))$

then $\vec{Y} \sim \text{Mult}_{10}(n, (p_1, p_2, p_3 + \ldots p_{10}))$

Q) $\vec{X} \sim \text{Mult}(n,\vec{p})$. Then given $X_1 = n_1$, P.MF of $(X_2, \ldots X_k)$

~~$(X_2, X_3 \ldots X_k) / X_1 = n_1 = \text{Mult}_{k-1}(n - n_1)$~~

$$(X_2, X_3, X_4, \ldots X_k / X_1 = n_1) \sim \text{Mult}_{k-1}[n - n_1, (p_2', \ldots p_k')]$$

with $p_2' = P$(being in category 2 / given not in 1)

$$p_2' = \frac{p_2}{(1-p_1)} = \frac{p_2}{p_2+p_3+\ldots p_k}$$

(renormalised to add up to one)

Cauchy Interview Problem

The Cauchy distribution is the distribution of $\frac{X}{Y}$ with $X,Y \overset{iid}{\sim} N(0,1)$

Q) Find the PDF of $T = \frac{X}{Y}$! (Cauchy's distribution)

$$P\left(\frac{X}{Y} \leq k\right) = P\left(\frac{X}{|Y|} \leq t\right) = P(X \leq t|Y|) = F(t)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{t|y|} \frac{e^{-y^2/2}\, e^{-x^2/2}}{\sqrt{2\pi}\,\sqrt{2\pi}}\, dx\, dy = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/2} \int_{-\infty}^{t|y|} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx\, dy$$

$$\boxed{C.D.F = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/2}\, \Phi(t|y|)\, dy} = \sqrt{\frac{2}{\pi}}\int_{0}^{\infty} e^{-y^2/2}\, \Phi(ty) \cdot dy$$

$$P.DF : F'(t) = \sqrt{\frac{2}{\pi}}\int_0^{\infty} e^{-y^2/2} \cdot y \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2y^2}{2}}\, dy$$

$$\boxed{F'(t) = \frac{dF(t)}{dt}}$$

$$PDF = F'(t) = \frac{1}{\pi}\int_0^{\infty} y e^{-\left(\frac{y^2}{2} + \frac{t^2y^2}{2}\right)} dy = \frac{1}{\pi}\int_0^{\infty} y e^{\frac{-y^2}{2}(1+t^2)}\, dy$$

$$P.D.F = \frac{1}{\pi}\int_0^{\infty} e^{-u(1+t^2)}\, du = \frac{1}{\pi(1+t^2)}\left[e^{-0} - e^{-\infty}\right] = \underline{\underline{\frac{1}{(1+t^2)\pi}}}$$

$$P(X \leq t|Y|) = \int_{-\infty}^{\infty} P\left[X \leq t|Y| \,/\, (Y=y)\right] \varphi(y)\, dy$$

→ PDF standard normal = PDF ~ N(0,1)

Using LOTP

$$P(X \leq t|Y|) = \int_{-\infty}^{\infty} P(X \leq t|y|)\, \varphi(y)\, dy$$

X independ of Y

$$\boxed{P(X \leq t|Y|) = \int_{-\infty}^{\infty} \Phi(t|y|)\, \varphi(y)\, dy}$$

Covariance

Definition : $Cov(X,Y) = \boxed{E[(X - E(X))(Y - E(Y))]}$

Properties

1. $\boxed{Cov(X,X) = Var(X)}$

$E[(X - E(X)](X - E(X))]$

$E[(X - E(X))^2]$

$E[X^2 + E(X)^2 - 2XE(X)]$

$= E(X^2) + E(X)^2 - 2E(X)E(X)$

$= E(X^2) - E(X)^2 = Var(X)$

also $Var, X = E[(X - E(X))^2]$

2. $\boxed{Cov(X,Y) = Cov(Y,X)}$

3. ~~ll~~ $\boxed{E(XY) - E(X)E(Y) = Cov(X,Y)}$

Proof $E[(X - E(X))(Y - E(Y))]$

$= E[XY - E(X)Y + E(X)E(Y) - XE(Y)]$

$= -E(X)E(Y) + E(XY) + E(X)E(Y) - E(X)E(Y)$

$= E(XY) - E(X)E(Y)$

4) $\boxed{Cov(X,C) = 0 \text{ if } C \text{ is a constant}}$

5) $\boxed{Cov(CX, Y) = c\,Cov(X,Y)}$

6) $Cov(X, Y+Z) = E(X(Y+Z)) - E(X)E(Y+Z)$

$= E(XY) + E(XZ) - (E(X)E(Y) + E(X)E(Z))$

$\boxed{Cov(X, Y+Z) = Cov(X,Y) + Cov(X,Z)}$ bilinearity

7) $Cov(X+Y, Z+W) = Cov(X+Y, Z) + Cov(X+Y, W)$

$= Cov(X,Z) + Cov(Y,Z) + Cov(X,W) + Cov(Y,W)$

$$\boxed{Cov\left(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_j Y_j\right) = \sum_{ij} a_i b_j Cov(X_i, Y_j)}$$

8) $Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$

Proof $Var(X_1 + X_2) = Cov(X_1 + X_2, X_1 + X_2)$

$= Cov(X_1, X_1) + Cov(X_2, X_2) + 2Cov(X_1, X_2)$

$\boxed{Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)}$

9. $\boxed{Var(X_1 + X_2 + X_3 + \cdots X_n) = Var(X_1) + Var(X_2) + \cdots Var(X_n) + 2\sum_{i<j} Cov(X_i, X_j)}$

Theorm

$$\boxed{\text{If } X,Y \text{ are independent, then they are uncorrelated, ie, } Cov(X,Y)=0 \quad \text{Converse is false.}}$$

Example: $Z \sim Normal(0,1)$

$X = Z$

$Y = Z^2$

$Cov(X,Y) = E(XY) - E(X)E(Y)$

$Cov(X,Y) = E(Z^3) - E(Z)E(Z^2)$ — odd moment zero for standard normal.

$Cov(X,Y) = 0$

$$\boxed{X,Y \text{ are extremely dependent } Y = f(X) \text{ even then their covariance } Cov(X,Y) = 0}$$

Definition: Correlation

$$Cor(X,Y) = \frac{Cov(X,Y)}{SD(X)\,SD(Y)}$$

$SD(X) = \sigma_X$

$SD(Y) = \sigma_Y$

$$\boxed{Cor(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}} = Cov\left(\frac{X-E(X)}{SD(X)}, \frac{Y-E(Y)}{SD(Y)}\right)$$

standardised version of Covariance.

Theorm $-1 \leq Cor(X,Y) \leq 1$ (form of Cauchy-Schwarz)

Proof

WLOG (without loss of generality) Assume X,Y are standardised.

$$0 \leq Var(X+Y) = Var X + Var Y + 2Cov(X,Y) = 2+2\rho$$

Let $Cor(X,Y) = \rho$

$$0 \leq Var(X-Y) = Var X + Var Y - 2Cov(X,Y) = 2-2\rho$$

because if standardised $Var(X)=1$

$Var(Y)=1$

$Cov(X,Y) = Cor(X,Y)$

$$so \Rightarrow \left.\begin{array}{l} 2+2\rho \geq 0 \\ 2-2\rho \geq 0 \end{array}\right\} \begin{array}{l} \Rightarrow \rho \geq -1 \\ \Rightarrow \rho < 1 \end{array} \Bigg\} \Rightarrow \underline{-1 \leq \rho \leq 1}$$

Multinomial Covariane

Ex $(X_1, \dots, X_K) \sim \text{Mult}(n, \vec{p})$ $\quad \vec{p} = (p_1, p_2, p_3 \cdots p_K)$

$\text{Cov}(X_i, X_j)\ \forall\ i, j$

if $i = j$, $\text{Cov}(X_i, X_i) = \text{Var}(X_i) = np_i(1-p_i)$

$\text{Cov}(X_1, X_2) = c$

$$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) + 2c$$

$$n(p_1+p_2)[1-(p_1+p_2)] = np_1(1-p_1) + np_2(1-p_2) + 2c$$

$$c = \frac{n(p_1+p_2)(1-p_1-p_2) - np_1(1-p_1) - np_2(1-p_2)}{2}$$

$$\Rightarrow c = \frac{n}{2}\left[p_1 + p_2 - (p_1+p_2)^2 - p_1 + p_1^2 - p_2 + p_2^2\right]$$

$$c = \frac{n}{2}\left(\cancel{p_1} + \cancel{p_2} - \cancel{p_1^2} - \cancel{p_2^2} - 2p_1p_2 \cancel{-p_1} + \cancel{p_1^2} \cancel{-p_2} + \cancel{p_2^2}\right]$$

$$c = \frac{n}{2}(-2p_1p_2) = \underline{\underline{-np_1p_2}}$$

Generally.

$$\boxed{\text{Cov}(X_i, X_j) = \begin{cases} -np_ip_j & i \neq j \\ np_i(1-p_i) & i = j \end{cases}}$$

Variance of Binomial $(n, p) = npq$

Ex $X \sim \text{Binomial}(n, p)$ write as $X = X_1 + X_2 + \dots X_n$ where

X_j are i.i.d Bern(p

$\text{Var}(X_j)$

$= E(X_j^2) - E(X_j$

$= E(X_j) - E(X_j)$

$= p - p^2$

$= p(1-p)$

$= pq$

Let I_A is the indicator R.V of event (A)

$$\boxed{\begin{aligned} I_A^2 &= I_A \\ I_A^3 &= I_A \\ I_A I_B &= I_{A \cap B} \end{aligned}}$$

$\mathrm{Var}(X) = npq$ since $\mathrm{Cov}(X_i, X_j) = 0$

for X_i, X_j are independent

$$\Rightarrow \mathrm{Var}(X_1+X_2+X_n) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \mathrm{Var}\, X_n) = \underline{npq}$$

<u>Example</u> Hypergeometric without replacement

$X \sim \mathrm{Hypergeom}(w, b, n)$

$X = X_1 + X_2 + \dots X_n$ $X_j = \begin{cases} 1 \text{ if the } j^{th} \text{ ball is white} \\ 0 \text{ otherwise.} \end{cases}$

X_j are not independent

$$\mathrm{Var}(X) = \mathrm{Var}(X_1 + X_2 + \dots X_n)$$

$$\mathrm{Var}(X) = n\,\mathrm{Var}(X_1) + 2\,{}^nC_2\,\mathrm{Cov}(X_1, X_2)$$

$$P(1 \cap 2) = P(1) \cdot P(2/1)$$

$$\mathrm{Cov}(X_1, X_2) = E(X_1 X_2) - E(X_1)E(X_2)$$

$$= E(X_1 X_2) - \left(\frac{W}{W+b}\right)^2$$

$$= E(X_{1\cap 2}) - \left(\frac{W}{W+b}\right)^2 = \left(\frac{W}{W+b}\right)\left(\frac{W-1}{W+b-1}\right) - \left(\frac{W}{W+b}\right)^2$$

$$\mathrm{Var}(X_1) = pq = \left(\frac{W}{W+b}\right)\left(\frac{b}{W+b}\right)$$

$$\Rightarrow \mathrm{Var}(X) = n \cdot \left(\frac{W}{W+b} \cdot \frac{b}{W+b}\right) + 2\,{}^nC_2 \cdot \left[\left(\frac{W}{W+b}\right) \cdot \left(\frac{W-1}{W+b-1}\right) - \left(\frac{W}{W+b}\right)^2\right]$$

$$= \frac{W}{W+b}\left[\frac{nb}{W+b} + 2\,{}^nC_2 \cdot \left[\frac{W-1}{W+b-1} - \frac{W}{W+b}\right]\right]$$

$$= \frac{W}{W+b}\left[\frac{nb}{W+b} + n(n-1)\left[\frac{(W-1)(W+b) - W(W+b-1)}{(W+b-1)(W+b)}\right]\right]$$

$$= \frac{nW}{W+b}\left[\frac{b}{W+b} + (n-1)\left(\frac{W^2+Wb-W-b-(W^2+Wb-W)}{(W+b)^2-(W+b)}\right)\right]$$

$$= \frac{nW}{W+b}\left[\frac{b}{W+b} + (n-1)\frac{-b}{(W+b)^2-(W+b)}\right]$$

$$= \frac{nWb}{(W+b)^2}\left[1 + \frac{(n-1)(-1)}{(W+b)-1}\right] = \frac{nWb}{(W+b)^2}\left[\frac{W+b-1-n+1}{(W+b)-1}\right]$$

$p = \frac{W}{W+b}$

$q = \frac{b}{W+b}$

$N = W+b$

$$= \frac{nW}{W+b} \cdot \frac{b}{W+b} \cdot \frac{W+b-n}{W+b-1} = \underline{npq \cdot \left(\frac{N-n}{N-1}\right)}$$

$\frac{N-n}{N-1}$ = finite population correction

Extreme case
$n=1$ with & without replacement are equal

$N >> n$

the $\frac{N-n}{N-1} \simeq 1$

then $V(x) \simeq V(Y)$

$X \sim$ Hypergeometric

$Y \sim$ Binomial

Transformations.

Theorm. Let X be a continuous random variable with PDF f_X, $Y = g(X)$, where g is differentiable & continous. also that g is strictly increasing.

- P.D.F $f_y(y) = f_X(x) \frac{dx}{dy}$ where $y = g(x)$, $x = g^{-1}(y)$
  and this is written in terms of y.

- Also $\frac{dx}{dy} = \left(\frac{dy}{dx}\right)^{-1}$

Proof

C.D.F of Y is $P(Y \leq y) = P(g(x) \leq y)$
$= P(X \leq g^{-1}(y))$
$= F_X(g^{-1}(y))$

derivative of C.D.F = P.D.F.

F_X = C.D.F of X
$g^{-1}(y) = x$

$$\text{C.D.F} \rightarrow P(Y \leq y) = F_X(x)$$
$$\downarrow$$
$$\text{P.D.F} \rightarrow f_y(y) = f_X(x) \frac{dy}{dx}$$ ...... chain rule

---

Example $Y = e^Z$ $Z \sim$ Normal $(0,1)$

log Normal

where $g(Z) = e^Z$ $\quad \frac{dy}{dz} = e^z$

P.D.F $= f_y(y) = f_X(x) \cdot \frac{dy}{dz}$

$y = g(z) = e^z$

$\cancel{f_x(y)} = \cancel{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\log y)^2} \cdot e^{\log y}}$

$z = \log y$

$Z = \log Y$

$\because \quad f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{\frac{-(\log y)^{2}}{2}} \cdot \frac{dz}{dy}$

$e^{z} = y$

$y = e^{z}$

$\frac{dy}{dz} = e^{z} = y$

$\frac{dz}{dy} = \frac{1}{y}$

$$\boxed{f_Y(y) = \frac{1}{y} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\log y)^2}}$$

transformation (in) $\mathbb{R}^n$

$\vec{Y} = g(\vec{X}) \,, \; g : \mathbb{R}^n \longrightarrow \mathbb{R}^n$

$\vec{X} = (X_1, X_2, X_3 \dots X_n)$ $\qquad \vec{Y} = (Y_1, Y_2, Y_3 \dots Y_n)$

Continous Random Vetos

joint P.D.F of $\vec{Y}$ is $f_{\vec{Y}}(\vec{y}) = f_{\vec{X}}(\vec{x}) \left| \frac{d\vec{x}}{d\vec{y}} \right| \longrightarrow$ Jacobian

abs(determinat of matrix) ↓

$$\left| \frac{d\vec{x}}{d\vec{y}} \right| = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & & & \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{bmatrix}$$

$$\left| \frac{d\vec{y}}{d\vec{x}} \right|^{-1} = \left| \frac{d\vec{x}}{d\vec{y}} \right|$$

Convolution (sums)

Let $T = X + Y$ $\quad X$ & Y independent

$P\left(\overset{\text{discrete}}{T=t}\right) = \sum_{x} P(X=x)\, P(Y=T-x)$

continous case

P.D.F $\boxed{f_T(t) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(t-x)\, dx}$

since $F_Y(t) = P(T \le t)$

$= \int_{-\infty}^{\infty} P(X+Y \le t / X=x)\, f_X(x)\, dx$

$= \int_{-\infty}^{\infty} P(Y \le (t-x))\, f_X(x)\, dx$

$F_T(t) = \int_{-\infty}^{\infty} F_Y(t-x) f_X(x)\,dx$ → Taking derivative w.r.t t on both sides

$$\Rightarrow \boxed{f_T(t) = \int_{-\infty}^{\infty} f_Y(t-x) f_x(x)\,dx}$$ Hence proved

---

Idea : prove existence of objects with (desired properties) using probablity.
$\xleftarrow{\quad} A \xrightarrow{\quad}$

show $P(A)>0$ for a random object

eg:- suppose each object has a number associated with it (A score)

~~P~~(A) = there is an object with a "good" score

Theorm
~~Theorem~~ There is an object with score is at least the average score (that is $E(X)$)

where X is the score of random variable

Example 100 people, 15 committees of 20 people, a person can be on 3 committees

Q show that there exists 2 committees with overlap ≥ 3

Idea: is to find average overlap of 2 random committees.

$E(\text{overlap of 2 random comittees}) = 100\left(\frac{3C_2}{15C_2}\right) = \frac{300\cdot 2}{15\cdot 14} = \frac{40}{14} = \underline{\underline{\frac{20}{7}}}$

(indicates ~~of~~ for each person whether overlaps in both or not ~~for~~ for all person)

↑ P(a person being in 2 comittes)

⇒ There exist a pair of committees with an overlap $\geq \lceil \frac{20}{7} \rceil$

⇒ There exist a pair of committees with an overlap ≥ 3 (Since overlap has to be integer.)

Beta distribution   Beta(a,b)   a>0, b>0

P.D.F $\boxed{f(x) = cx^{a-1}(1-x)^{b-1}}$, $0<x<1$

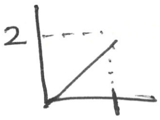
$0<X<1$

Features

- Flexible family of continuous distributions on (0,1)
- Usefull modelling tool
- often used as a prior for a parameter in (0,1)
- "Conjugate prior to Binomial"
- Probablity of probablities

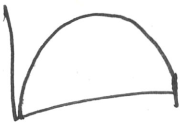
ie, a=b=1 } ⇒ uniform

a=2, b=1 ⇒



a=b=½

a=b=2



Conjugate prior for Binomial

$X/p \sim \text{Binomial}(n,p)$

But p is unknown

Applying Baes' rule

Let $p \sim \text{Beta}(a,b)$ → prior

Find posterior distribution P/X

$$f(P/X=k) = \frac{P(X=k/p)\cdot f(p)}{P(X=k)} = \frac{nC_k\, p^k (1-p)^{n-k} \cdot \left[c\, p^{a-1}(1-p)^{b-1}\right]}{P(X=k)}$$

→ does not depend on p

— constant

$$\propto p^{a+k-1}(1-p)^{n+b-k-1} \Rightarrow P/X \sim \text{Beta}(a+k, b+n-k)$$

$$\Rightarrow P/X \sim \text{Beta}(a+X, b+n-X)$$

both prior & posterior P & P/X are distributed in family of Beta

⇒ Conjugate prior to Binomial

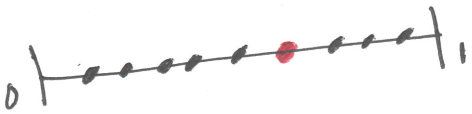
a & b in Beta(a,b) may not be integers

Case a & b are integers

Find $\int_0^1 nC_k x^k (1-x)^{n-k} \cdot dx$ without using calculus, using a story.

Baye's Billards

1) Take n+1 billard balls, all white, paint on of them pink, then throw them on (0,1) independently.



2) Take (n+1) billard balls, first throw then randomly paint one of them.

Let X = #balls to the left of •ball $X \in (0,n)$

$$P(X=k) = \int_0^1 P(X=k/p)\, f(p)\, dp$$

p → position of pink ball.

f(p) = 1 → ball has equal chance of being in (0,1)

p ~ uniform(0,1)

$$P(X=k) = \int_0^1 nC_k p^k (1-p)^{n-k} dp$$

→ According to second story $X \in (0,n)$

→ Any ball is picked equally likely & painted pink

⇒ X=(0,n) equally likely

$P(X=k) = \frac{1}{n+1}$ } because n+1 equally likely events 0,1,2...n

⇒

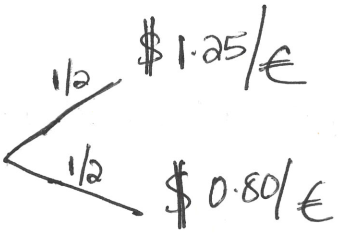
$$\boxed{\int_0^1 nC_k p^k (1-p)^{n-k} dp = \frac{1}{n+1}}$$

Foreign exchange Fx

Brain teasers

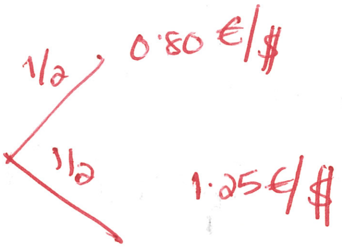
1) Let $1€ = 1\$$ initially

1. year's time



$E(€) = 1.025\ \$$

$E(\$) = 0.9756\ €$



$E(\$) = €1.025$

2) TARP

Warrants ≡ call option

US government paid $450M to GS

US gov had the right to buy 10M GS shares for $125 in 10 years time.

on Oct 2008, GS shares were 95$ during this warrant.

If in 10 years time GS share becomes $150, government gets 25$ worth profit/share.

If in 10 years time GS becomes $100, goverment does not get anything worth

$$g\{S_T\} = \max\{S_T - K, 0\}$$

$$\text{fair Price of warrant} = \int_0^\infty \max\{S_T - K, 0\} f(S_T)\, dS_T = \text{Expected Worth of Warrant}$$

Puzzle

What is the next # in sequence : 0, 1, 2,

Ans is 720!

if 0, 1, 2, 720!
next in seq = $(((4!)!)!)!$

Sterling's formula

$$\boxed{n! \simeq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n}$$

Gamma function & gamma distribution

*making factorial continous.

$$\Gamma(a) = \int_0^\infty x^a e^{-x} \cdot \frac{dx}{x} \quad , \text{ for any real } a>0$$

$\Gamma(n) = (n-1)!$ for n a positive integer

$\Gamma(x+1) = x\Gamma(x)$

$\Gamma(1/2) = \sqrt{\pi} \Rightarrow \Gamma(3/2) = \frac{\sqrt{\pi}}{2}$

$\Gamma(5/2) = \frac{3\sqrt{\pi}}{4}$

$$\Gamma(5/2) = \Gamma(1+3/2)$$
$$= \frac{3}{2}\Gamma(3/2)$$
$$= \frac{3}{2}\frac{\sqrt{\pi}}{2} = \frac{3\sqrt{\pi}}{4}$$

Back to gamma distribution

$$1 = \frac{\int_0^\infty x^a e^{-x} \cdot \frac{dx}{x}}{\Gamma(a)} \Rightarrow \boxed{\frac{1}{\Gamma(a)} x^a \cdot e^{-x} \cdot \frac{1}{x}}$$

Gamma (a, 1) P.D.F
a>0

$Y \sim$ Gamma $(a, \lambda) \longrightarrow$ Let $Y = \frac{X}{\lambda}$, $X \sim$ Gamma $(a, 1)$

$$f_Y(y) = f_X(x)\frac{dx}{dy} =$$

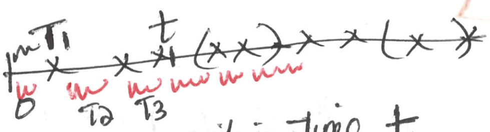
$Y = \frac{X}{\lambda}$ $x = \lambda y$
$dx = \lambda dy$
$\frac{dx}{dy} = \lambda$

$$f_Y(y) = \frac{(\lambda y)^a}{\Gamma(a)} \frac{e^{-\lambda y}}{\lambda y} = \frac{(\lambda y)^a e^{-\lambda y}}{\Gamma(a) \cdot y} \quad (y>0)$$

$Y \sim \text{Gamma}(a, \lambda) \sim \frac{(\lambda y)^a e^{-\lambda y}}{\Gamma(a) \cdot y}$

P.DF. $\boxed{f_Y(y) = \frac{(\lambda y)^a e^{-\lambda y}}{\Gamma(a) y}}$ $y > 0$

Gamma, Exponential Connection



→ Inter arrival times are i.i.d exponential λ

N_t = #emails in time t
~ Poson (λt)

#arrival in disjount equally lengthed intervals are constant.

X_n = time of n^{th} arrival

$P(T_1 > t) = P(N_t = 0) = \frac{e^{-\lambda t} \cdot (\lambda t)^0}{0!} = e^{-\lambda t}$

$X_n = T_1 + T_2 + \dots T_n$

$X_n = \sum_{j=1}^{n} T_j$ where $T_n \sim$ iid-exponential

n = integer
then $X_n \sim \text{Gamma}(n, \lambda)$ ie sum of iid exponential ~ gamma

negative Binomial = Sum of geometric (Discrete Domain)

gamma = Sum of exponentia (Continous Domain)

Gamma distribution is continous analogue of negative Binomial

---

Proof that $X_n = \sum_{j=1}^{n} T_j$, T_j = iid exponential expo(1) = Gamma(n,1) ($\lambda = 1$ here)

MGF. $T_j = T_1$ (Since iid) $= \frac{1}{1-t}$, $t < 1$

$\Rightarrow$ MGF of X_n is $\left(\frac{1}{1-t}\right)^n$, $t < 1$ $\Rightarrow$ MGF of $(X_n) = \left(\frac{1}{1-t}\right)^n$

$\boxed{MGF(nT) = MGF(T)^n}$

Let $Y \sim \text{gamma}(n,1)$

$$E(e^{tY}) = \int_0^\infty e^{ty} \frac{1}{\Gamma(n)} \frac{y^n e^{-y}}{y} dy \quad \text{[LOTUS]}$$

$$= \frac{1}{\Gamma(n)} \int_0^\infty y^n e^{-(1-t)y} \frac{dy}{y}$$

Let $X = (1-t)Y$

$x = (1-t)y$

$dx = (1-t)dy$

$dy = \frac{dx}{(1-t)}$

$y = \frac{x}{(1-t)}$

$$= \frac{1}{\Gamma(n)} \int_0^\infty \left[\frac{x}{(1-t)}\right]^n e^{-x} \frac{dx}{\cancel{(1-t)}} \cdot \frac{\cancel{(1-t)}}{x}$$

$$= \frac{1}{\Gamma(n)} \int_0^\infty x^n e^{-x} \cdot \frac{dx}{x} \cdot \left(\frac{1}{1-t}\right)^n = \frac{\cancel{\Gamma(n)}}{\cancel{\Gamma(n)}} \cdot \left(\frac{1}{1-t}\right)^n$$

$\boxed{\text{MGF of } (X_n) = \text{MGF of gamma}(n,1)}$

Hence proved.

$$\Rightarrow \text{MGF}(\text{gamma}(n,1)) = \underline{\underline{\left(\frac{1}{1-t}\right)^n}}$$

Let $X \sim \text{gamma}(a,1)$, Find $E(X^c)$

$$E(X^c) = \int_0^\infty x^c \cdot \frac{1}{\Gamma(a)} \cdot x^a e^{-x} \frac{dx}{x} = \frac{1}{\Gamma(a)} \cdot \int_0^\infty x^{a+c} \cdot e^{-x} \frac{dx}{x} = \frac{\Gamma(a+c)}{\Gamma(a)} \text{ if } (a+c > 0)$$

$$\mu = E(X) = \frac{\Gamma(a+1)}{\Gamma(a)} = \frac{a\Gamma(a)}{\Gamma(a)} = a$$

$$E(X^2) = \frac{\Gamma(a+2)}{\Gamma(a)} = \frac{\Gamma(a+1)(a+1)}{\Gamma(a)} = a(a+1)\frac{\Gamma(a)}{\Gamma(a)} = a(a+1)$$

$$\boxed{E(X^n) = a(a+1)(a+2)\cdots(a+(n-1))}$$

$$\text{Variance} = E(X^2) - E(X)^2 = a(a+1) - a^2 = a$$

for gamma(a,λ) has

$\mu = \frac{a}{\lambda}$	$\sigma^2 = \frac{a}{\lambda}$
mean	Variance

Bank- Post office Example

$X \sim \text{gamma}(a, \lambda)$;
$Y \sim \text{gamma}(b, \lambda)$

a= ~~no#~~ of people in line@Bank

b = # people in line @ P.O.

X ∼ wait time @ Bank
Y ∼ wait time @ post office.

Q) Find distribution of Total wait time $T = X + Y \sim \text{gamma}(a+b, \lambda)$ if

Q) $W = \frac{X}{X+Y}$

Q Joint of T,W

Let $\lambda = 1$ to simplify

(Joint P.DF) $f_{T,W}(t,w) = f_{X,Y}(x,y) \left| \frac{\partial(x,y)}{\partial(t,w)} \right|$

$$f_{T.W} = \frac{1}{\Gamma(a)\Gamma(b)} x^{a} e^{-x} y^{b} e^{-y} \frac{1}{xy} \cdot \left| \frac{\partial(x,y)}{\partial(t,w)} \right|$$

$x + y = t$
$\frac{x}{x+y} = w$
$\frac{x}{t} = w$
$x = tw$
$y = t(1-w)$

$$J = \begin{vmatrix} w & t \\ 1-w & -t \end{vmatrix}$$

$$|w(-t) - (1-w)t|$$
$$|-t| = t$$

$$\Rightarrow f_{T.W}(t,w) = \frac{1}{\Gamma(a)\Gamma(b)} x^{a} e^{-x} y^{b} e^{-y} \frac{1}{xy} t = \frac{1}{\Gamma(a)\Gamma(b)} (tw)^{a} e^{-tw} \, {}^{b}t(1-w)^{b} e^{-t(1-w)} \frac{t}{t^{2} w(1-w)}$$

$$f_{TW}(t,w) = \frac{1}{\Gamma(a)\Gamma(b)} t^{a+b} e^{-t} w^{a} (1-w)^{b} \cancel{e^{-tw}} \frac{t}{t^{2} w(1-w)}$$

$$= \frac{1}{\Gamma(a)\Gamma(b)} t^{a+b} e^{-t} w^{a-1} (1-w)^{b-1} \frac{1}{t} \times \frac{\Gamma(a+b)}{\Gamma(a+b)}$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \underbrace{\frac{1}{\Gamma(a+b)} t^{a+b} e^{-t} \cdot \frac{1}{t}}_{} \cdot w^{a-1} \cdot (1-w)^{b-1}$$

P.DF
$$\boxed{f_T(t) = f(t) = \text{gamma}(a+b, 1)}$$

⇒ T, W are independent.

Marginal P.D.F's

$$f_W(w) = \int_{-\infty}^{\infty} f_{T,W}(t,w)\,dt = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} w^{a-1} \times \frac{(1-w)^{b-1}}{\Gamma(a+b)} \int_{-\infty}^{\infty} t^{a+b} e^{-t} \frac{dt}{t}$$

$$\boxed{f_W(w) = \frac{\cancel{\Gamma(a+b)}}{\cancel{\Gamma(a+b)}} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} w^{a-1} (1-w)^{b-1}}$$

⇒ Therefore normalising constant of β is $\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$

∴ P.DF of β =

let $X \sim Beta(a,b)$ $a>0$, $b>0$, $0<X<1$

P.D.F of X $$\boxed{f_X(x) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} , & 0\leq x \leq 1 \\ 0 , & \text{otherwise} \end{cases}}$$

$T \sim gamma(a+b, 1)$ } independent
$W \sim Beta(a,b)$

Find $E(W)$, $W \sim Beta(a,b)$ (i) LOTUS

$$E(W) = \int_0^1 x . \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} . x^{a-1} . (1-x)^{b-1} dx$$

$$E(W) = \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x . (1-x)^{b-1} dx$$

$$\Rightarrow \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} . \frac{\Gamma(a+b+1)}{\Gamma(a+1)\Gamma(b)} . \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} . x^{a+1-1} (1-x)^{b-1} dx$$

$$= \frac{\Gamma(a+b)\,\Gamma(a+1)\,\cancel{\Gamma(b)}}{\Gamma(a)\,\cancel{\Gamma(b)}\,\Gamma(a+b+1)} . \boxed{\int_0^1 Beta(a+1,b)dx} = 1$$

$$= \frac{\Gamma(a+1)}{\Gamma(a)} . \frac{\Gamma(a+b)}{\Gamma(a+b+1)}$$

$$= \frac{a\cancel{\Gamma(a)}}{\cancel{\Gamma(a)}} . \frac{\cancel{\Gamma(a+b)}}{(a+b)\cancel{\Gamma(a+b)}} = \frac{a}{a+b}$$

$$\Rightarrow \boxed{E(W) = \frac{a}{a+b}}$$

2^{nd} approach

$W = \left(\frac{X}{X+Y}\right)$ previous questions as distribution won't change.

$\Rightarrow E(W) = E\left(\frac{X}{X+Y}\right)$

$E(W) = \frac{E(X)}{E(X+Y)}$ [is True in this case of Bank, Post office because W independent T $\Rightarrow \frac{X}{X+Y}$ independent of $(X+Y)$

$\therefore \frac{X}{X+Y}$ & $(X+Y)$ are uncorrelated

independent $\Rightarrow$ uncorrelated.

$\Rightarrow E\left(\frac{X}{X+Y}\right) \cdot E(X+Y) = E(X)$

$\Rightarrow E\left(\frac{X}{X+Y}\right) = \frac{E(X)}{E(X+Y)}$

$$\boxed{E(W) = \frac{a}{a+b}}$$

Uncorrelated X, Y

$\Rightarrow E(X)E(Y) = E(XY)$

Order Statistics

Let $X_1 \dots X_n$ iid

The order statistics $X_1 < X_2 < X_3 \dots < X_n$, where $X_1 = \min$ & $X_n = \max$

eg. If n = odd then median $= X_{n+1/2}$

Get other quantiles/percentiles. (median = 50th percentile)

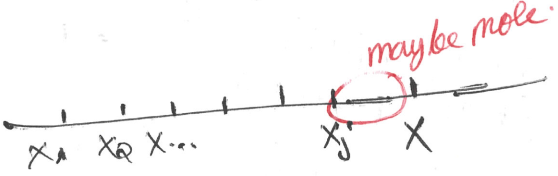
- → Difficult since they are dependent
- → Tricky in discrete case
- → In the continuous case there will not be repeatations since probablity is zero.

Let $X_1 \dots X_n$ be iid with PDF f, CDF F

Find C.D.F & P.D.F of $X_{(j)}$ j^{th} order statistics.

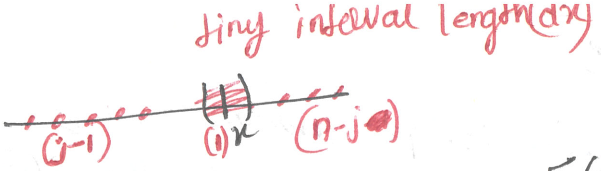
C.D.F $P(X_j < x)$ = P(atleast j of the X_i's are to the left of x)

$$= \sum_{k=j}^{n} \left[{}^{n}C_k \, F(x)^k \, (1-F(x))^{n-k} \right]$$



marginal

P.D.F



P.DF $f_{X_j}(x) = n\, f(x)\, dx \; {}^{n-1}C_{(j-1)}\, F(x)^{j-1} (1-F(x))^{n-j}$

$$\boxed{f_{X_j}(x) = n\; {}^{n-1}C_{j-1}\, F(x)^{j-1} (1-F(x))^{n-j} f(x)}$$

Example Uniform order statistics

$u_1, \dots u_n$ iid uniform $(0,1)$

$f_{u_j}(x) = n\, {}^{(n-1)}C_{j-1}\, x^{j-1} (1-x)^{n-j}$ $\qquad 0 \leq x \leq 1$

$\Rightarrow U_j \sim \text{Beta}(j, n-j+1)$

$F(x) = x$

$1 - F(x) = 1 - x$

$f(x) = 1 \quad 0 \leq x \leq 1$

---

if $n=2$

$E(u_2 - u_1)$

$u_1 \sim \text{Beta}(1,2)$ $\qquad$ mean of Beta $= \frac{a}{a+b}$

$u_2 \sim \text{Beta}(2,1)$

$E(u_1) = \frac{1}{3}$

$E(u_2) = 2/3$

Confirms earlier result.

$E(u_1 - u_2) = \underline{\underline{1/3}}$

$E(u_2 - u_1) = E(\max - \min)$
done previously

Conditional Expectance

$E(X/A)$

$E(X) = E(X/A) \cdot P(A) + E(X/A^c) \cdot P(A^c)$ $\qquad$ Law of Total probablity for conditional expectance.

Proof

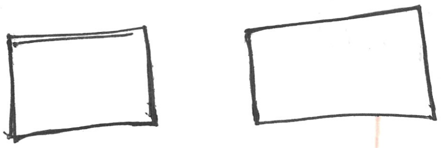
$E(X) = \sum_x x\, \underbrace{P(X=x)}_{}$

Expand using Law of TP.

$= \sum_x x \left[P(X=x/A) \cdot P(A) + P(X=x/A^c)\, P(A^c) \right]$

$= P(A) \sum_x x\, P(X=x/A) + \sum_x x\, P(X=x/A^c) P(A^c) = \left(\frac{E(X/A)}{P(A)}\right) + \left(\frac{E(X/A'}{P A}\right.$

Puzzle: Two Envolope Paradox



- One evalope contains exactly twice as money as other one
- Cannot be identified
- You get to pick one
- & open it & see the sum in it
- Now do you want to switch

Say you open & find X amount

the other envolope has $2x$ & $\frac{x}{2}$ amount.

$$E(\text{sum in other}) = \frac{1}{2}(2x) + \frac{1}{2}\left(\frac{x}{2}\right)$$

$$E(\text{sum in other}) = x + \frac{x}{4} = \boxed{\frac{5x}{4}}$$

What's wrong with argument 2:

Solution.

Argument 1: $E(X) = E(Y)$ By symmetry

Argument 2: $E(Y) = E(Y/Y=2X)\cdot P(Y=2X) + E(Y/Y=\frac{X}{2})\cdot P(Y=\frac{X}{2})$

$$E(Y) \overset{?}{=} E(2X)\cdot\frac{1}{2} + E\left(\frac{X}{2}\right)\cdot\frac{1}{2}$$

$$E(Y) = \frac{5}{4}E(X)$$

Correct version $E(2X/Y=2X)\,\frac{1}{2} + E\left(\frac{X}{2}/Y=X/2\right)\,1/2$

ie $E(Y/Y=2X) \neq E(2X)$

Let I be the indicator of $Y=2X$, the $E(Y/I \text{ happens}) \neq E(2X)$

Then X & I are dependent

Then Y & I are dependent

$E(Y/I) =$ when I unknown $E(X)=E(Y)$

when I is known $E(X) \neq E(Y)$

$\therefore$ I dependent on X

I dependent on Y

$E(A/B=k) = E(A)$ only if A & $B=k$ independent

$E(A/A=2B) = E(2B)$ only if A & $(A=2B)$ independent

Patterns in Coin flip : Reapeated fair Coin flips,

How many flips until HT? W_{HT} — RV

HH? W_{HH} — RV

$E(W_{HT})$

$E(W_{HH})$

$E(W_{HT}) \neq E(W_{HH})$ are false application symmetry.

Symmetry says, $E(W_{HH}) = E(W_{TT})$

, $E(W_{HT}) = E(W_{TH})$

- $E(W_{HT}) = E(W_1) + E(W_2) = 2+2 = 4$ linearity.

$$\underbrace{TTTTH}_{W_1}\underbrace{H}_{W_2}\boxed{HT}$$

- $E(W_{HH}) = E(W_{HH}/\text{1st toss} = \text{Head})\frac{1}{2} + E(W_{HH}/\text{1st toss is tails})\frac{1}{2}$

TTTTTT HT

- $E(W_{HH}) = \left[\underbrace{2\cdot\tfrac{1}{2}}_{HH} + \underbrace{\left[W_{HH}+2\right]\tfrac{1}{2}}_{HT}\right]\frac{1}{2} + \left[1 + E(W_{HH})\right]\frac{1}{2}$

let $E(W_{HH}) = x$

$$x = \left[1 + (x+2)\frac{1}{2}\right]\frac{1}{2} + \left[1+x\right]\frac{1}{2}$$

$$x = \frac{1}{2} + \frac{x+2}{4} + \frac{x+1}{2}$$

$$x = \frac{x+2+2(x+1)+2}{4} \Rightarrow 4x = 3x+6$$

$$\Rightarrow x = 6$$

$\Rightarrow E(W_{HH} = 6)$

$$E(Y/X=x) = \sum_{y} y\, P(Y/X=x)$$

discrete

$$E(Y/X=x) = \int_{-\infty}^{\infty} y \cdot f_{Y/X}(y/X=x)\, dy$$

~~$= \int_{-\infty}^{\infty} y \left[\int_{-\infty}^{\infty} f_{(X,Y)}(x,y)\, dx\right] \cdot dy$~~

$$= \int_{-\infty}^{\infty} y \cdot \frac{f_{X,Y}(x,y)}{\left[\int_{-\infty}^{\infty} f(x,y)\,dy\right]} \cdot dy = \int_{-\infty}^{\infty} y \cdot \frac{f_{X,Y}(x,y)}{f_X(x)} \cdot dy$$

$$f_{(Y/X)}(Y/X=x) = \frac{f_{(X,Y)}(x,y)}{f_X(x)}$$

as $$\boxed{f_{(X,Y)}(x,y) = f_{Y/X}(y/x=x) \cdot f_X(x)}$$

Let $g(x) = E(Y/X=x)$

The define $g(X) = g(X)$

evaluate $g(x)$ ~~after evaluating~~ and plug in X →

ie let $g(x) = x^2+2x$

then $g(X) = X^2+2X$

* don't plug in X before evaluating

$g(X)$ is a function of RV

∴ $g(X)$ is an R.V

Example: Let X,Y be iid poison(λ)

1) $E(X+Y/X) = E(X/X) + E(Y/X)$ • Linearity still holds.

$$= \underbrace{X}_{X \text{ defines } E(X)} + \underbrace{E(Y)}_{\text{independence}}$$

---

2) $E(X/X+Y)$ Let $T = (X+Y)$, find conditional p.m.f.

$E(X|T) = ?$

$$P(X=k/T=n) = \frac{P(T=n/X=k)\,P(X=k)}{P(T=n)}$$ Baye's Rule.

$$= \frac{P(Y=(n-k))\,P(X=k)}{P(T=n)}$$

$$= \frac{\frac{e^{-\lambda}\lambda^{n-k}}{(n-k)!}\cdot\frac{e^{-\lambda}\lambda^{k}}{k!}}{\frac{e^{-2\lambda}(2\lambda)^{n}}{n!}}$$

$$\boxed{T \sim \text{poisson}\,(2\lambda)}$$

$$= \frac{e^{-2\lambda}\lambda^{n}\;n!}{(n-k)!\,k!\;e^{-2\lambda}\lambda^{n}2^{n}} = nC_k\left(\frac{1}{2}\right)^n$$

$$\Rightarrow P(X=k/T=n) = \text{Binomial}\left(n, \frac{1}{2}\right)$$

back to problem

$$E(X/T=n) = \sum_{x} x\, P(X=k/T=n)$$

$$E(X/T=n) = \sum_{\substack{all\,k\\0}}^{n} k\cdot nC_k\left(\frac{1}{2}\right)^n = n/2$$

expectation of binomial $= n/2$

$$E(X/T) = \frac{T}{2}$$

Another approach

$$E(X/X+Y) = E(Y/X+Y)$$ because of symmetry ie iid

$$E(X/X+Y) + E(Y/X+Y) = E(X+Y/X+Y) = X+Y$$

$$\Rightarrow E(X/X+Y) = \frac{X+Y}{2} = \frac{T}{2}$$

Iterated Expectated ~~of~~ (Atoms Law)

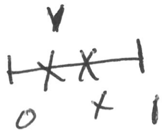
- $E(E(Y/X)) = E(Y)$

Example Let $X \sim N(0,1)$, $Y = X^{\alpha}$ Then $E(Y/X) = E(X^{a}/X) = X^{\alpha} = Y$

Example $E(X/Y) = E(X/X^2) = 0$

$x = \pm\sqrt{a}$ equally likely.

Example



- stick, break of a ~~tip~~ piece,
  - break of another piece.
- each break position is uniform within the limits

$Y/X \sim \text{Uniform}(0, X)$

~~EX~~ $E(Y/X=x) = \frac{x}{2}$ so $E(Y/X) = \frac{X}{2}$ = function of RV also a R.V

$E[E(Y/X)] = E\left(\frac{X}{2}\right) = \frac{1}{2}E(X) = \underline{\underline{\frac{1}{4}}}$ | but $X \sim \text{Uniform}(0,1)$ $E(X) = 1/2$

$E(Y) = \underline{\underline{\frac{1}{4}}}$ — from Atom's Law

---

Properties of Conditional Expectance

1) $E(h(X)Y/X) = h(X)E(Y/X)$ [taking out what's know]

2) $E(Y/X) = E(Y)$ if X & Y are independent

3) $E(E(Y/X)) = E(Y)$ Iterated Expectation / Atome Law

4) $E((Y - E(Y/X))h(X)) = 0$

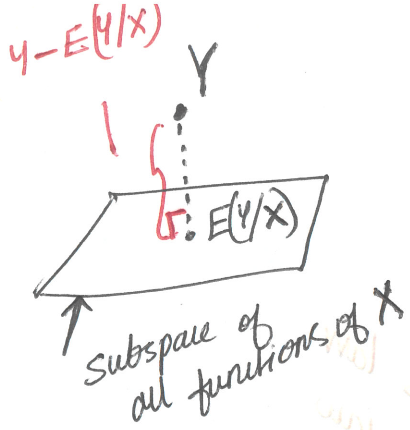
$(Y - E(Y/X))$ — is uncorrelated with any function of X

$\text{Cov}(Y - E(Y/X), h(X))$

$= E([Y - E(Y/X)]h(X)] - \underbrace{E(Y - E(Y/X))}_{} E(h(X))$

zero because of Atom Law & Linearity.

$\Rightarrow \text{Cov}([Y - E(Y/X)], h(X)) = 0$



inner product $<X,Y> = E(XY)$

Proof of 4

Linearity

$= E(Y\,h(X)) - E(E(Y/X)h(X))$ — Linearity.

$= E(Yh(X)) - E[E(h(X)Y/X)]$

$= E[Y\,h(X)] - E[Y\,h(X)]$ — Adam Law

$= 0$

Proof of 3 discrete case

$E(Y/X) = g(X)$ — function of X (R.V)

$E[g(X)] = \sum_x g(x)\,P(X=x)$ → LOTUS

$g(x) = E(Y/X=x)$

$E(gX) = \sum_x E(Y/X=x)\cdot P(X=x)$

$= \sum_X \sum_Y \left[y\,P(Y=y/X=x) \right] \cdot P(X=x)$

$= \sum_Y \sum_X y\,\underline{P(Y=y/X=x)\cdot P(X=x)}$

$\downarrow$

$\sum_Y \sum_X = y\,\underline{P(Y=y, X=x)}$ — joint P.M.F

$\sum_y y \cdot \sum_x P(Y=y, X=x) = \sum_y y \cdot P(Y=y)$ — marginal from joint.

Definition of Conditional Variance

$Var(Y/X) = E(Y^2/X) - E(Y/X)^2$ = Correct.

$Var(Y/X) = E\left((Y - E(Y/X))^2/X\right)$

Property of Conditional Variance

i) $Var(Y) = E(Var(Y/X)) + Var(E(Y/X))$ — EVE's law
EVVE law

Story
→ Assume there are 3 ~~po~~ groups of population in a community
X=1 represents 1'st group
X=2 represent 2'st group
X=3 represent 3rd group

Y = ~~population~~ height of a person selected at random from ~~a~~ the community.

E(Y) = average height of a ~~p~~ random person in the community.

E(Y/X) = averge height of a radom person in group X

Var(Y/X) = Variance of height in group X

$$Var(Y) = \underbrace{E(Var(Y/X))}_{\text{The average Value of Variance Within each group}} + \underbrace{Var(E(Y/X))}_{\text{The Variance of the average population between each group}}$$

* looking at Variance between population averages

Example: Pick a random city in a state.
Pick a random sample of n people in that city

X = # people with disease in the sample

RV = Φ = propostion of people infected in a city)

Φ between 0,1 because it is a probablity depend on city so RV.

Find mean(X) — E(X)
Variance(X) — Var(X)

Assume $\Phi \sim Beta(a,b)$

Intitutive sense:

$$E(X) = \mu\, E(N) \longrightarrow \text{expected \# customers. dim-less}$$

$E(X)$ ↓ profit \$

μ ↓ (average spent by a customer) \$

$$Var(X) = \sigma^2 E(N) + \mu^2 Var(X)$$

$Var(X)$ ↓ $\$^2$

σ^2 ↓ $\$^2$

$E(N)$ ↓ count dim-less

μ^2 ↓ $\$^2$

$Var(X)$ ↓ count^2 dim-less

Statistically Inequalities

1) Cauchy-Schwarz: $|E(XY)| \leq \sqrt{E(X)^2 E(Y)^2}$

* if X & Y are uncorrelated

then $E(X)E(Y) = E(XY)$

* if correlated then

$$|E(XY)| < \sqrt{E(X)^2 E(Y)^2}$$

If X,Y has mean zero then the correlation b/w X & Y

$$Cov(X,Y) = \left| \frac{E(XY)}{[E(X^2)E(Y^2)]^{1/2}} \right| \leq 1$$

} equivalent to Cauchy's-schwarz.

because if mean zero then $Var(X) = E(X^2)$, $Var(Y) = E(Y^2)$

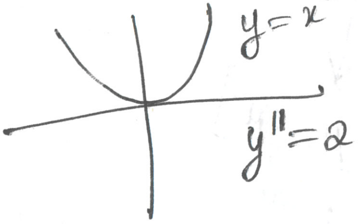
---

2) Jensen's inequality.

if g is a "convex" function, then

$$E[g(X)] \geq g\,E(X)$$

Convex function $g''(X) \geq 0$



If h is concave, $E[h(x)] \leq h(E(x))$ apply negative sign in Jensen's inequality $g(x) = -h(x)$

eg $E(x^2) \geq E(x)^2$

$g(x) = y = x^2$

---

eg:- $E\left(\frac{1}{x}\right)$ $x > 0$

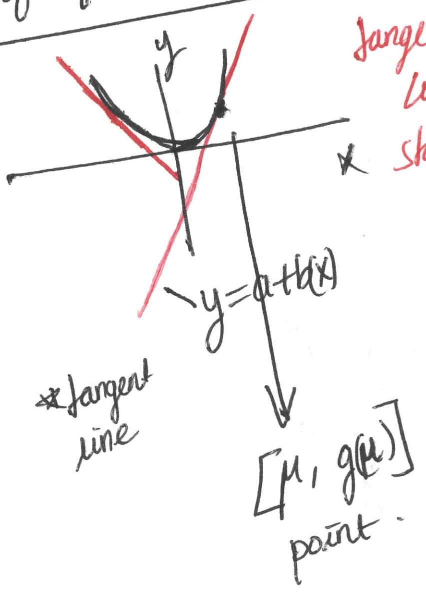
$$\left.\begin{array}{l} g(x) = \frac{1}{x} \\ g'(x) = \frac{-1}{x^2} \end{array}\right\} \Rightarrow \boxed{E\left(\frac{1}{X}\right) \geq \frac{1}{E(X)}}$$

$g''(x) = \frac{2}{x^3}$ convex $\forall x > 0$

eg. $E(\ln x) \leq \ln E(x)$

$$\left.\begin{array}{l} g(x) = \ln(x) \\ g'(x) = \frac{1}{x} \\ g''(x) = \frac{-1}{x^2} \text{ Concave} \end{array}\right\} \boxed{E(\ln(x)) \leq \ln E(x)}$$

Proof of Jensen's



tangent line of convex function stays below curve

$g(x) \geq a+bx$ — tangent line true for all x

$g(X) \geq a+bX$

$E(g(x)) \geq E(a+bX)$

$E(g(x)) \geq a+bE(X)$

$E(g(x)) \geq a+b\mu$

but $a+b\mu = g(\mu)$ since line passes through point $(\mu, g(\mu))$

$= E(g(x)) \geq g(\mu)$

$\Rightarrow E(g(x)) \geq g(E(x))$

---

*

③ Markov's Inequalities

$$P(|X| \geq a) \leq \frac{E|X|}{a}, \text{ for any } a > 0$$

Side note

$P(X=x) = E(I)$

Where I is the event (X=x) 0 if event does not occur
1 if event occurs

Since $E(I) = \sum_{i=0,1} i\,P(i)$

$E(I) = P(i=1)$

$E(I) = P(X=x)$

$\Rightarrow P(|X| \geq a) = E(I_{|X| \geq a})$

$a E(I_{|X| \geq a}) \leq E|X|$

To prove, we need to show

$a I_{|X| \geq a} \leq |X|$

Case one $I=0$

$\Rightarrow 0 \leq |X|$ always true

Case two $I=1$ if $|X| \geq a$

$\Rightarrow a \leq |X|$ which is always true if $I=1$

So we proved

$a I_{|X|>a} \leq |X|$

So $a E(I_{|X|>a}) \leq E(|X|)$

So $P(|X| \geq a) \leq \frac{E(|X|)}{a}$

Example for Intuition

Q) We have 100 people. is it possible atleast 95% of people are younger than average person in the group.

Ans Yes

Q) Is it possible that atleast 50% of people are older than 2x the average age.

Ans No because even if rest 50% are age 0 then

averge $= \frac{(2x + \text{older}) \times 50 + (0 \times 50)}{100} = x + \text{older}$! Contradicts

⇒ So We can have $\frac{1}{x^{th}}$ people be x times average

$$\underbrace{\left(\left(\frac{1}{x^{th}}\text{ people}\right) \times x \text{ times } \mu\right) + \left(\text{rest of people } 0\right)}_{\text{whole group } 1}$$

⇒ μ-toldes

---

4) Chebyshev's Inequality $\qquad \mu = E(X)$

$$P(|X-\mu| > a) \leq \frac{Var(X)}{a^2} \qquad a > 0$$

$$P(|X-\mu| > c * S.D(X)) \leq \frac{1}{c^2} \qquad c > 0$$

} Both equal
put $a = c \cdot SD(X)$

Proof $\quad P(|X-\mu| > a) = P((X-\mu)^2 \geq a^2)$ — equavelent

by Markov's inequality.

$$P((X-\mu)^2 > a^2) \leq \left(\frac{E(X-\mu)^2}{a^2} = \frac{Var\ X}{a^2}\right)$$

Let $X_1 X_2 \dots X_n$ be i.i.d with mean μ & Variance σ^2

let $\bar{X}_n = \frac{\sum_{j=1}^{n} X_j}{n}$ [Sample mean]

Law of LARGE NUMBERS (strong)

$\bar{X}_n$ converges to μ as $n \to \infty$ with probablity 1
(sample mean) (true mean)

Example let $X_j \sim$ Bernouli(p)
then $\frac{X_1 + X_2 + \dots X_n}{n} \to p$ with probablity 1

Weak law of large numbers

For any $c > 0$, $P(|\bar{X}_n - \mu| > c) \to 0$ as $n \to \infty$

Proof:

$P(|\bar{X}_n - \mu| > c) \leq \frac{\text{Var}\,\bar{X}_n}{c^2}$ — chebyshev from last class.

$\text{Var}\,\bar{X}_n = \frac{1}{n^2} = n$ ~~Var~~ $\text{Var}(X_j) = \frac{1}{n}\sigma^2$

$\Rightarrow P(|\bar{X}_n - \mu| > c) \leq \left(\frac{\sigma^2}{nc^2}\right)$ σ, c are constants so

$P|\,|\bar{X}_n - \mu| > c) < 0$ as $n \to \infty$

---

* $\bar{X}_n - \mu \to 0$ as $n \to \infty$ with probablity 1, but what does the distribution of what $\bar{X}_n$ look like.

$n^{1/2}\frac{(\bar{X}_n - \mu)}{\sigma} \rightsquigarrow N(0,1)$ in distribution as $n \to \infty$

CENTRAL LIMIT THEORM

Equality statement

$$\left(\frac{\sum_{j=1}^{n}(x_j) - n\mu}{\sqrt{n}\cdot\sigma}\right)$$

* We need to normalise
ie subtract the mean $\sum_{j=1}^{n} x_j = n\mu$ — linearity

* divid by $\overset{S.D}{}$ (

but variance $= n\sigma^2$
of $\sum_{j=1}^{n}(x_j)$

$SD = \sqrt{\text{Variane}}$
$SD = \sqrt{n\sigma^2}$
$SD = \sqrt{n}\,\sigma$

μ – mean of x_j } independent
σ – Variance of x_j } iid x_j's

ie $$\left(\frac{\sum_{j=1}^{n}(x_j) - n\mu}{\sqrt{n}\sigma}\right) \rightarrow \sim \text{Normal}(0,1)$$

intuition

$$T = \overset{\nearrow\infty}{n^{1/2}} \times \overset{\nearrow 0}{\left(\frac{\bar{X}_n - \mu}{\sigma}\right)} \sim \text{Normal}(0,1)$$

* This is approximately distributed as Normal $(0,1)$

Proof (MGF exists : assumption) But works even if MGF's doesnot converge Central limit theorm works.

We can assume that $\mu = 0$ & $\sigma = 1$ without using losing generality.

Let $S_n = \sum_{j=1}^{n} X_s$, We need to show $\text{MGF}\left(\frac{S_n}{\sqrt{n}}\right) \rightsquigarrow N(0,1)$ MGF as $n \to 0$

$$E\left(e^{tS_n/\sqrt{n}}\right) = E\left(e^{tx_1/\sqrt{n}}\right) \dots E\left(e^{tX_n/\sqrt{n}}\right)$$

$$E\left(e^{tS_n/\sqrt{n}}\right) = E\left(e^{\frac{tx_1}{\sqrt{n}}}\right)^n = \left[\text{MGF of } x_j \text{ evaluated at}\left(\frac{t}{\sqrt{2}}\right)\right]^n$$

$$\underset{n\to\infty}{Lt}\left[\text{MGF of } X_j \text{ at } \left(\frac{t}{\sqrt{2}}\right)\right]^n = \underset{n\to\infty}{Lt}\left(\text{MGF of } X_j \text{ evaluated at } 0\right)$$

(1^{∞}) — form Taking the log.

$$\underset{n\to\infty}{Lt}\left(n \cdot \ln\left[\text{MGF } x_j \text{ evaluated at } \frac{t}{\sqrt{n}}\right]\right)$$

$$\underset{n\to\infty}{Lt}\left[\frac{\ln\left(MGF_{X_j}\left(\frac{t}{\sqrt{n}}\right)\right)}{\frac{1}{n}}\right]$$

change of variables.

let $y = \frac{1}{\sqrt{n}}$

$$\underset{y\to 0}{Lt}\left[\frac{\ln\left(MGF_{X_j}(yt)\right)}{y^2}\right]$$

Apply L'Hopital's rule.

$$\underset{y\to 0}{Lt} = \frac{\frac{1}{MGF_{X_j}(yt)} \cdot M'(yt) \cdot t}{2y}$$

— chain rule.

$$\underset{y\to 0}{Lt} = \frac{t \cdot \overset{\to 0}{M'(yt)}}{\underset{0\uparrow}{2y} \cdot \underset{\to 1 \text{ so get rid of that.}}{M(yt)}}$$

Still $\frac{0}{0}$ form

$M(t) = E(e^{tX_i})$

$M(0) = E(e^0) = 1$

mean $= M'(0) = 0$

Variance $= M''(0) = 1$

(mean, variance) initially assumed to be 0,1

$$\frac{t}{2} \cdot \underset{y\to 0}{Lt}\left(\frac{M'(yt)}{y}\right)$$

— L'Hopital's rule

$$\frac{t}{2} \cdot \underset{y\to 0}{Lt}\left(\frac{\overset{\to 1}{M''(yt)} \cdot t}{1}\right) = \frac{t^2}{2}$$

$\Rightarrow$ reversing the log $MGF_{\frac{S_n}{\sqrt{n}}}(t) = e^{t^2/2}$

but $e^{t^2/2} = MGF_{N(0,1)}(t)$

$$\Rightarrow MGF_{\frac{S_n}{\sqrt{n}}}(t) = MGF_{N(0,1)}(t)$$

$$\Rightarrow \frac{S_n}{\sqrt{n}} \sim N(0,1) \quad \text{as } n \to \infty$$

Example

Binomial approximated by a Normal

Let $X \sim \text{Binomial}(n,p)$, think of $X = \sum_{j=1}^{n} X_j$, $X_j \sim$ iid Bernouli p

$$P(a \leq X \leq b) = P\left(\frac{a-np}{\sqrt{npq}} \leq \frac{X-np}{\sqrt{npq}} \leq \frac{b-np}{\sqrt{npq}}\right)$$

but $X = \sum_{j=1}^{n} (X_j)$

$$\Rightarrow \frac{\sum_{j=1}^{n}(X_j) - np}{\sqrt{npq}}$$

$p = \mu$

$pq = \sigma^2$ } for bernoulli (p)

now this is of the form $\frac{\sum_{j=1}^{n}(X_j) - n\mu}{\sqrt{n}\,\sigma} \sim N(0,1)$ as $n \to \infty$ according to CLT.

$Z \sim N(0,1)$

So $P\left(\frac{a-np}{\sqrt{npq}} \leq Z \leq \frac{b-np}{\sqrt{npq}}\right)$

$$= \int_{\frac{a-np}{\sqrt{npq}}}^{\frac{b-np}{\sqrt{npq}}} \text{Normal PDF} = \text{Normal CDF}\left(\frac{b-np}{\sqrt{npq}}\right) - \text{Normal CDF}\left(\frac{a-np}{\sqrt{npq}}\right)$$

therefore $\approx \Phi\left(\frac{b-np}{\sqrt{npq}}\right) - \Phi\left(\frac{a-np}{\sqrt{npq}}\right) \approx P(a \leq X \leq b)$

Contrast to Poisson approximation

n – large

p – small

$np = \lambda$ – moderate

Normal approximation

n – large

p close to 1/2

$\chi^2(n)$ [chi-squared] distribution

Let $V = Z_1^2 + Z_2^2 + \dots Z_n^2$ where Z_j i.i.d $N(0,1)$

Then by definition $V \sim \chi^2(n)$ — comes every in stat. because adding squared.

$\chi^2(1)$ is gamma $(\frac{1}{2}, \frac{1}{2})$ — H.W

Exercise :- idea →

So $\chi^2(n)$ is gamma $(n\frac{1}{2}, \frac{1}{2})$ — Property of Gamma

---

Student-t Distribution (gosset 1908)

Let $T = \frac{Z}{\sqrt{V/n}}$, with $\left\{ \begin{array}{l} Z \sim N(0,1) \\ V \sim \chi^2(n) \end{array} \right\}$ V & Z independent

Then, by defn $T \sim t_n$

n — degrees of freedom

Properties 1) Symmetric distribution (about 0)

2) $n=1 \Rightarrow$ Cauchy → no mean finite existance

3) $n \geq 2 \Rightarrow E(T)=0$

$T = \frac{Z_1}{|Z_2|} \sim$ Cauchy

$Z_1 Z_2$ independent

$E(T) = E(Z)\, \boxed{E\left(\frac{1}{\sqrt{V/n}}\right)}$ as Z & V are independent and they are uncorrilated.

but @ n=1 this part does'nt exist ∴ no mean for n=1

intuition (Symmetry)

$E(Z^2)=1$ (Variance), $E(Z^4)=3$ } skip factorial

$E(Z^5) = 3\times5$

$E(Z^{odd}) = 0$ — earlier proved

for even moments of Norma

$E(Z^{2n}) = E(Z^2)^n$

$Z^2 = \chi^2(1) = \text{Gamma}(\frac{1}{2}, \frac{1}{2})$

Use lotus to get moments of Gamma then easily get

(4) Heavier-tailed than normal distribution for (small n)

⑤ For large n, t_n looks very much like $N(0,1)$

($t_n \rightsquigarrow N(0,1)$ as $n \to \infty$)

Proof:

Let $t_n = \frac{Z}{\sqrt{V_n/n}}$ with $Z_1, Z_2, Z_3 \dots$ iid $(N(0,1))$

$V_n = Z_1^2 + \dots Z_n^2$

$Z \sim N(0,1)$ independent of Z_i's

Then $\frac{V_n}{n} \to 1$ as $n \to \infty$ with probability 1 by Law of Large Numbers

$\Rightarrow \underset{n\to\infty}{t_n} \quad \frac{Z}{\sqrt{V_n/n}} = Z$

so $T_n \to Z$ with probability 1

$\frac{V_n}{n} = \frac{Z_1^2 + Z_2^2 + \dots Z_n^2}{n} \to E\left(\frac{V_n}{n}\right)$

$= \frac{E\,n(Z_i^2)}{n} = E(Z_i^2)$

$E(Z_i^2) = Var(N(0,1))$ 2nd moment

$E(Z_i^2) = 1$

$\boxed{\text{* CLT is the normalised version of LLN}}$

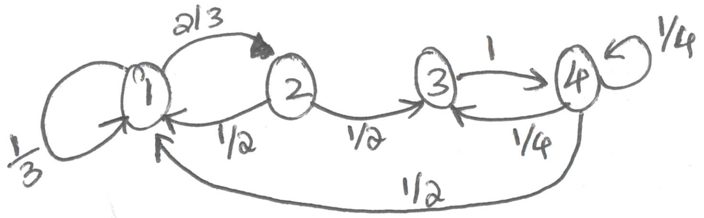
LLN gives distribution as $P(\text{event}) \to 1$ as $n \to \infty$

i.e the random variable without normalising → mean (sampling)

CLT gives distribution as $P\left(\frac{X_n - n\mu}{\sqrt{n}\sigma}\right) \sim N(0,1)$ as $n \to \infty$

i.e the random sampling variable after normalising → Normal.

Example:



Transition matrix $Q = q_{ij}$ — (transition probability from i^{th} state to j^{th} state)

$$Q = \begin{bmatrix} 1/3 & 2/3 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \\ 1/2 & 0 & 1/4 & 1/4 \end{bmatrix}$$

- Every row sums to 1
  Equally valid is Q^T, then columns sum to 1
- Non negative entries

Markov chain Monty Carlo MCMC — Lookup

Suppose at time n, X_n has distribution $\vec{s_n}$ (row vector) $1 \times m$ matrix
↳ p. ↳ (PMF of states) at n.

$p(X_{n+1} = j)$?
PMF @ n+1

Solution. $P(X_{n+1} = j) = \sum_i [P(X_{n+1} = j / X_n = i) \, P(X_n = i)$

$= \sum_i s_i q_{ij}$ = j^{th} entry of $\vec{S}_{ROW}[Q]$ ie SQ

$$S_{n+1} = \begin{bmatrix} s_1 & s_2 & s_3 & s_4 \end{bmatrix} \begin{bmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{32} & q_{34} \\ q_{41} & q_{42} & q_{43} & q_{44} \end{bmatrix} = S_n Q$$

$S_{1 \times m} Q_{m \times m}$

$SQ_{1 \times m}$

$$\Rightarrow \left.\begin{aligned} S_{n+1} &= S_n Q \\ S_{n+2} &= S_{n+1} Q \\ S_{n+3} &= S_{n+2} Q \end{aligned}\right\} \quad \begin{aligned} S_{n+1} &= S_n Q \\ S_{n+2} &= S_n Q^2 \\ S_{n+3} &= S_n Q^3 \end{aligned}$$

In general give $S \rightarrow$ distribution of present state. then distribution of X_{n+k}^{th} distribution is given by SQ^k

$P(X_{n+1} = j / X_n = i) = q_{ij}$

$P(X_{n+2} = j / X_n = i) = ?$

$= \sum_{k} P(X_{n+2} = j / X_{n+1} = k, X_n = i) \, P(X_{n+1} = k / X_n = i)$

$= \sum_{k} [q_{kj} \cdot q_{ik}] = [i^{th} \text{ row of } Q] \begin{bmatrix} j^{th} \\ \text{colum} \\ \text{of} \\ Q \end{bmatrix}$

$$\boxed{P(X_{n+2} = j / X_n = i) = ij^{th} \text{ entry in } Q^2}$$

* Extending generally.

$P(X_{n+k} = j / X_n = i) = ij^{th}$ entry in Q^K

Stationary Distribution (Steady state or Long run or equilibrium)

$\vec{S}$ is a probablity vector 1xM

$\vec{S}$ is a stationary distribution for the chain if $\vec{S}Q = \vec{S}$

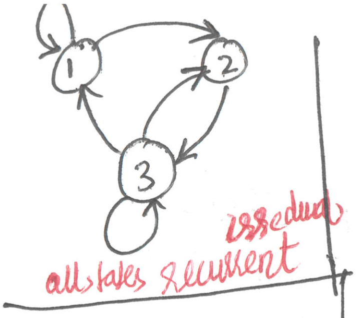
- Intuition if X_n is distributed as S then X_{n+1} is distributed as $\vec{S}Q = \vec{S}$

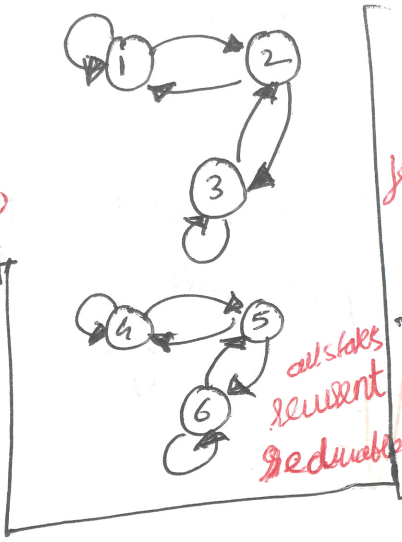
That is why it is ~~st~~ stationary.

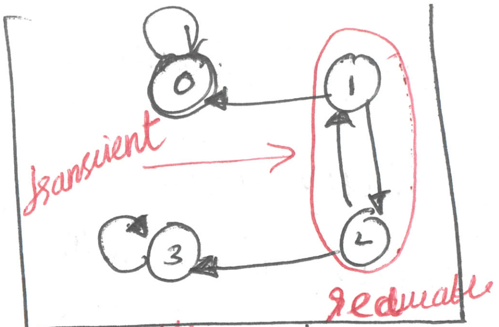
1) Does $\vec{S}$: stationary exist (all entries +ve and real)?
2) Is it unique?

} Mostly yes!

3) Does chain converge to $\vec{S}$ (Stability)?
4) How to compute it? → Elimination, but computationally difficult.







* chain is irreducible if possible with positive probablity to get from anywhere to anywhere

* A state in a chain is called recurrent if starting there has probablity 1 of returning, otherwise it is called transient.

irreducable ⇒ recurrent
recurrent ⇏ irreducable

Stationary distribution again

$\vec{s}$ (prob. row vector) is a stationary distribution if $\vec{s} = \vec{s}Q$.

Always assuming finite no of states

Actually an eigen value or eigen vector equation.
take transpose
$Q^T \vec{s}^T = \vec{s}^T$
$\lambda = 1$.
Q^T is Markov
$\lambda = 1$ always a eigen value
ref. 18:06

Theorm For any irreducable Markov chain's

1) A stationary distribution exists.
2) It is unique
3) $S_i = \frac{1}{r_i}$, where r_i is the average return time to return to ith state assuming starting from state i

true even for periodic case.

4) If Q^m is strictly positive for any m then, chain is irreducable, then X_n converges to s_i as $n \to \infty$ no matter the starting

$\Rightarrow tQ^n \to s$ as $n \to \infty$ } t can be any valid probablity vector

How to compute Stationary distribution

Reversable Markov Chain

Defn : Markov chains with transition matrix $Q=[q_{ij}]$ is reversible if there is a prob vectors $\vec{s}$ such that

$$\boxed{s_i q_{ij} = s_j q_{ji}}$$ for all states i,j ✓

$\vec{s} = [s_1\ s_2 \dots s_n]$

Thm

If reversible with respect to a $\vec{s}$ then that $\vec{s}$ is a stationary distribution.

Intuition reversible / Time reversible.

if we video tape the state machine and play it backwards, then we would not notice the diffrence ⟹ time reversible.

Proof

Let $\boxed{s_i q_{ij} \underset{\forall i,j}{=} s_j q_{ji}}$ We need to show $\vec{s}Q = \vec{s}$

$$\sum_i s_i q_{ij} = \sum_i s_j q_{ji}$$

⟹

j^{th} column of $\vec{s}Q = s_j \sum_i q_{ji}$

j^{th} column of $\vec{s}Q =$ j^{th} entry of $s \times 1$

$\sum_i q_{ji} = \sum_j q_{ij}$

= Sum of elements of i^{th} row $=1$ for Markov matrix

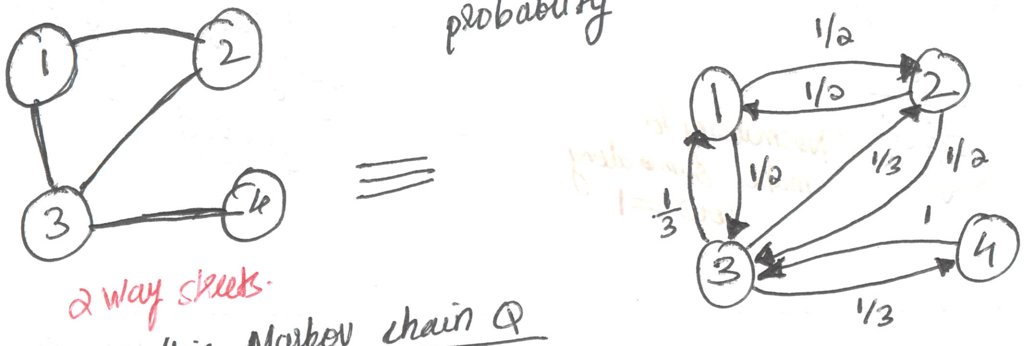
⟹ This is true $\forall j$

⟹ $\vec{s}Q = \vec{s}$

Example of Reversible Markov Chain

Random walk on an undirected (2way street) Network.

$\longrightarrow$ just says that we select one of the path avaliable at a particular state with ~~so~~ equal probablity



* At each state the paths to next state (if exist) is equally likely

~~matrix~~ Q for this Markov chain Q

$$Q = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Let d_i be the degree of the i^{th} state, degree = no. of path avalible at each node.

$$\boxed{d_i q_{ij} = d_j q_{ji} \quad \forall ij}$$
$i \neq j$

* q_{ij}, q_{ji} are both zero or both non zero

undireted
because it is undireted N/w
- either no path to & fro
- or path to & fro
- No case of oneway path

$\Rightarrow$
* if $i=j$ then LHS=RHS
* if $q_{ij}=0$ then q_{ji} also zero
* So only cases remaining to be checked are $q_{ij} \neq 0 \Rightarrow q_{ji} \neq 0$

* If there is an edge btw i & j
$q_{ij} = \frac{1}{d_i}$ (because all paths from i are equally likely)

but $q_{ij} = \frac{1}{d_i} \Rightarrow$ $\left.\begin{matrix} d_i q_{ij} = 1 \\ d_j q_{ji} = 1 \end{matrix}\right\}$ if $q_{ij} \neq 0$ & $q_{ji} \neq 0$ } They are both non zero at same time since undirected.

So proof

$$d_i q_{ij} = d_j q_{ji} \quad \forall ij$$

Case I : $i=j$ LHS=RHS

Case II : $q_{ij}=0 \Rightarrow q_{ji}=0$ (Undireted)

Case III $q_{ij} \neq 0 \Rightarrow q_{ji} \neq 0 \Rightarrow d_i q_{ij} = d_j q_{ji} = 1$

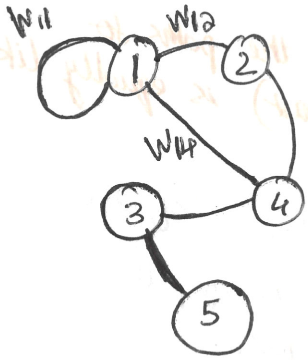
di's are non negative +ve integers. Just need to normalise

So with M nodes 1, ...M
and d_i = degree of node i
Then $S_i = \frac{d_i}{\sum_j d_j}$ Normalising to make sum along row = 1

$\Rightarrow S_i q_{ij} = S_j q_{ji}$

$\Rightarrow \frac{d_i}{\cancel{\sum_j d_j}} q_{ij} = \frac{d_j}{\cancel{\sum_j d_j}} q_{ji}$

proved



$W_{ij} > 0$ if there exists a edge b/w i^{th} & j^{th} node

$W_{ij} = 0$ if no edge

Assumption: $W_{ij} = W_{ji}$ ⇒ Random Walk

From state i go to j with propablity proportional to weight W_{ij}.

- This type of Random walk is still reversable.

$q_{ij} = \frac{W_{ij}}{\sum_k W_{ik}}$ } Normalising to make probablity total as 1

$$\left(\sum_k W_{ik}\right) q_{ij} = W_{ij} = W_{ji} = \sum_k W_{jk} (q_{ji})$$

$$\Rightarrow \left(\sum_k W_{ik}\right) q_{ij} = \left(\sum_k W_{jk}\right) q_{ji}$$

Let $S_i \propto \sum_k W_{ik}$

$S_i = \frac{\sum_k W_{ik}}{\sum_i \sum_k W_{ik}} = \frac{\sum_k W_{ik}}{\sum W}$

$\Rightarrow S_i q_{ij} = S_j q_{ji}$ ∀ i,j ⇒ reversible

* This is the Complete generalisation of Reversable Markov chain.

* If we have any markov chain, then it can be represented in this way.

Let, $W_{ij} = S_i q_{ij} = S_j q_{ji} = W_{ji}$

We need to prove that a network with weights Wij assigned in this manner has transition probablity qij. They be definition of assumption the matrix is a Markov chain and can always be expressed as a random walk along bidirectional weigted edges Wij from i to j & j to i.

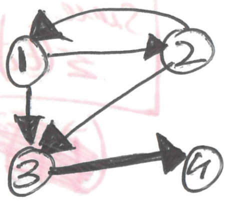
$$\text{ie, } P(X_{n+1} = j/X_n = i) = \frac{W_{ij}}{\sum_k W_{ik}} = \frac{S_i q_{ij}}{\sum_k S_i q_{ik}} = \frac{\cancel{S_i}\, q_{ij}}{\cancel{S_i} \sum_k q_{ik}}$$

$$P(X_{n+1} = j/X_n = i) = q_{ij}$$ Hence proved.

Non Reversible Chain

Example: **Google Page-Rank**

WWW with 4 pages



* importance of (Rank) a page should ~~to~~ not be just based on the ~~no~~ #pages linking to it but also based on those pages' rank themselves.

S_j – score of j^{th} page

$$S_j = \sum_i S_i q_{ij}$$

$$\Phi = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \Rightarrow \vec{S} = \vec{S}\Phi$$

which says that $\vec{S}$ is the stationary distribution of this random websurfing chain.

After normalsi vecto $\vec{S}$ it beco Stationa distribu

Stategy of solving S for large order matrixes without Gaussian elimination

$$G = \underbrace{\alpha\Phi}_{\text{follow link structure}} + \underbrace{(1-\alpha)\frac{J}{M}}_{\text{Teleportation}}$$

, where

$M = \#\text{pages}$

$$J = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 & 1 \\ 1 & 1 & \cdots & \cdots & 1 & 1 & 1 \\ \vdots & & & & & & \vdots \\ 1 & 1 & 1 & \cdots & 1 & 1 & 1 \end{bmatrix}$$ all ones.

$0 < \alpha \leq 1$ – probablity.

This equation says —

With probability $\alpha \rightarrow$ go ~~to~~ to a webpage determined by link structure distribution

With probability $1-\alpha \rightarrow$ Teleport to a completely random page with uniform ~~pag~~ distribution ignoring link structure & its distribution

Strategy to compute $\vec{S}$ without Gaussian elimination.

Run the chain for long time until it converges to the stationary distribution.

Consequence $\alpha = 0.85$ used by google.

$G \rightarrow$ Google chain

- * guarantees irreducability
- * no zeroes in transition matrix G
- * There exists a stationary distribution
- * It is unique
- * ~~It converges~~ Chain converges to stationary

Let $\vec{t}$ be initial probability vector [PMF @ time 0 written ~~at~~ time 0]

- $\vec{t}G$ after 1 step
- $\vec{t}G^2$ after 2 step

$\vec{t}\vec{J} = [1, 1, 1 \cdots]$

Since $\Sigma(t_i) = 1$

[illegible]

$$tG = \alpha \vec{t}\vec{Q} + (1-\alpha)\frac{\vec{t}J}{M} \Big\}$$ Not ~~complicated~~ complex

lot of zeros
Very Sparse matrix
(clever data storage & optimal multiplication)

$$\lim_{n\to\infty} tG^n = \vec{S}$$

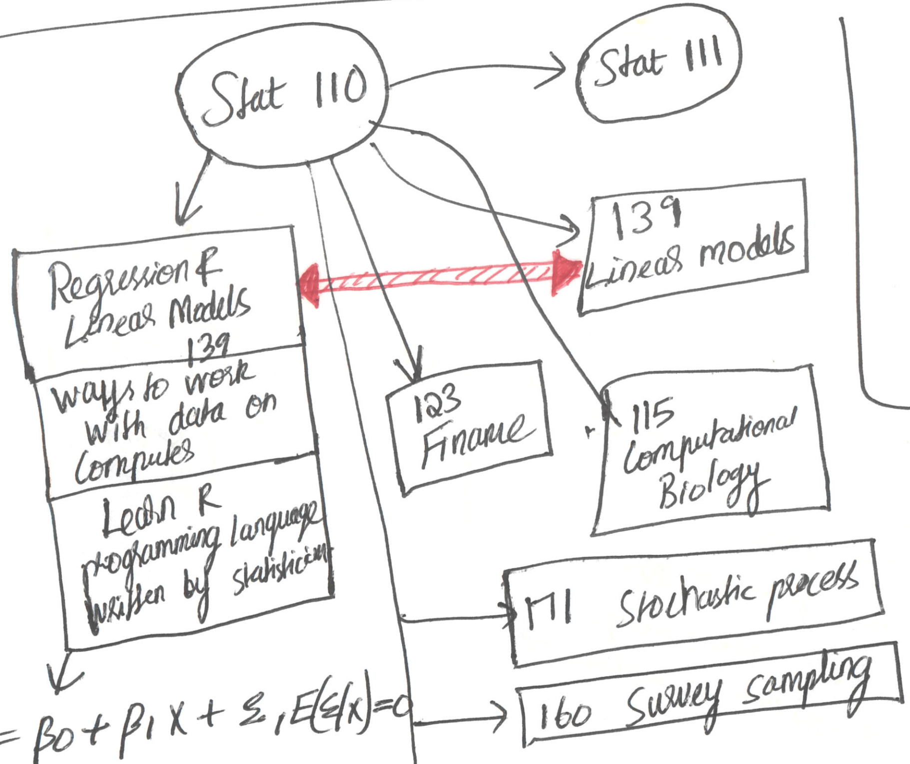
where $\vec{S}$ = stationary distribution

Where $\vec{S}$ = represents page rank.

Top 10 List (according to Rank)

1) Conditioning
2) Symmetry
3) RV's & their distributions
4) Stories proofs & Story for distribution
5) Linearity
6) Indicator Random Variable — Interview problems
7) LOTUS
8) Law of Large Numbers/Central Limit theorem
9) Markov chains
10) Bayes Rule

Tricks, Threads to solve most questions, interviews, new thoughts etc.



Probability

models known → Predict future data

Statistics

Past & Present data → Model the distribution

$y = \beta_0 + \beta_1 X + \varepsilon,\ E(\varepsilon/x) = 0$

predict y given X

β_0, β_1 are unknown parameters

ε = unknown error term

$Cov(y, X) = Cov(\beta_1 X) + Cov(\varepsilon, X)$

$= \beta_1 Var(x) + Cov(\varepsilon, x)$

$\Rightarrow \beta_1 Var(x) = Cov(y, x)$

but $E(\varepsilon|x) = 0$

$E(E(\varepsilon|x)) = 0 \Rightarrow E(\varepsilon) = 0$ Adam's law

$\Rightarrow Cov(\varepsilon x) =$

$E(\varepsilon x) = E(E(\varepsilon x|x))$ Adam's law

$E(x E(\varepsilon|x)) = 0$

$\Rightarrow \beta_1 = \frac{Cov(X,Y)}{Var(X)}$ } Projection in Linear Algebra

Books: Mostly Harmless Econometrics