Holy Shit Problem

m: pigeonholes n: pigeons p: probability

Given m & p find n s.t. ~~that~~ the probability of an overlap is higher than p.

Ans: $n > \frac{1}{2} + \sqrt{2m\log\left(\frac{1}{1-p}\right) + \frac{1}{4}}$

Corr. If $3n+1$ balls are put in n^2 bins, then the probability of overlap is larger than 0.98889

Pf Let there be m bins and n balls.

Then the probability of no ball being together is

$$1-p = \left(\frac{m-1}{m}\right)\left(\frac{m-2}{m}\right)\cdots\left(\frac{m-n+1}{m}\right)$$

$$1-p = \prod_{i=1}^{n-1}\left(1-\frac{i}{m}\right)$$

use $1-x \leq e^{-x}$

$$\left(1-\frac{i}{m}\right) \leq e^{-i/m}$$

$$\prod_{i=1}^{n-1}\left(1-\frac{i}{m}\right) \leq e^{-\left(\frac{n(n-1)}{2m}\right)}$$

$$\Rightarrow 1-p \leq e^{-\left(\frac{n(n-1)}{2m}\right)}$$

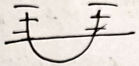
$1-p \le e^{-\left(\frac{n(n-1)}{2m}\right)}$. $1-p > \prod_{i=1}^{n-1}\left(1-\frac{i}{m}\right)$

$\Rightarrow \frac{1}{1-p} < e^{\left(\frac{n(n-1)}{2m}\right)}$

$\Rightarrow \ln\left(\frac{1}{1-p}\right) < \frac{n^2-n}{2m}$

$\Rightarrow n^2-n-2m\ln\left(\frac{1}{1-p}\right) > 0$

$\Rightarrow n = \frac{1 \pm \sqrt{1+8m\ln\left(\frac{1}{1-p}\right)}}{2}$



$= \quad n > \frac{1}{2} + \sqrt{\frac{1}{4}+2m\ln\left(\frac{1}{1-p}\right)}$

or

$n < \frac{1}{2} - \sqrt{\frac{1}{4}+2m\ln\left(\frac{1}{1-p}\right)} < 0$ [Doesn't work]

Small workaround solves sir's problem.

We are to find n

s.t.

the probability of overlap $> p$.

!(prob of overlap) $< 1-p$

$\Rightarrow \underbrace{\prod_{i=1}^{n-1}\left(1-\frac{i}{m}\right)}_{1-\tilde{p}} < 1-p$ is reqd.

So we know now

$1-\tilde{p} < e^{-\frac{n(n-1)}{2m}} \rightarrow e^{-\frac{n(n-1)}{2m}} < 1-p$

Now to get the sols

Heads & Tails

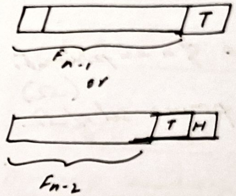
Ex Find out the probability of obtaining no two consecutive heads if a fair coin is tossed 7 times.

$F_1 = |\{T, H\}| = 2$.

F_n: no. of chances s.t. there are no consecutive heads if a fair coin is tossed

$F_2 = |\{TT, TH, HT\}| = 3$

Two cases



$F_n = F_{n-1} + F_{n-2}$

T 2 → H 1 2

$F_{n+1} = F_n + F_{n-1}$

Reqd probability $= \frac{34}{2^7} = \frac{17}{64}$

$$\begin{pmatrix}F_{n+1}\\F_n\end{pmatrix} = \begin{pmatrix}1&1\\1&0\end{pmatrix}\begin{pmatrix}F_n\\F_{n-1}\end{pmatrix}$$

Eigenvalues : $\frac{1+\sqrt{5}}{2}$, $\frac{1-\sqrt{5}}{2}$

$$\begin{pmatrix}F_{n+1}\\F_n\end{pmatrix} = \begin{pmatrix}1&1\\1&0\end{pmatrix}\begin{pmatrix}F_n\\F_{n-1}\end{pmatrix} = \begin{pmatrix}1&1\\1&0\end{pmatrix}^2\begin{pmatrix}F_{n-1}\\F_{n-2}\end{pmatrix}$$

$$= \begin{pmatrix}1&1\\1&0\end{pmatrix}^n\begin{pmatrix}F_1\\F_0\end{pmatrix}$$

$$\begin{pmatrix} F_{n+1} \\ F_n \end{pmatrix} = (PDP^{-1})^n \begin{pmatrix} F_1 \\ F_0 \end{pmatrix}$$

$$= PD^nP^{-1} \begin{pmatrix} F_1 \\ F_0 \end{pmatrix}$$

Axioms of Probability

- Experiment :- An act that can be repeated under similar circumstances
- Sample space : Set of all outcomes of a experiment. (Ω)
- Set of events $(\mathcal{E})$: A subset of the power set of Ω s.t.

1. $\Omega \in \mathcal{E}$
2. If $A \in \mathcal{E}$, then $A^c = \Omega \setminus A \in \mathcal{E}$.
3. $\mathcal{E}$ is closed under countable union.
4. There is a fn $P : \mathcal{E} \to [0,1]$ s.t.

Something Called Vitali Sets.

$$P(\Omega) = 1$$

If $A_1, \dots, \in \mathcal{E}$ s.t. $A_i \cap A_j = \phi$ then

$$P(A_1 \cup A_2 \cup \dots) = \sum_{i=1}^{\infty} P(A_i)$$

Eg: Tossing a coin

$\Omega = \{H, T\}$ $\quad \mathcal{E} = \{\phi, \{H\}, \{T\}, \{H, T\}\}$

$$P(A^c) = 1 - P(A)$$

We know $P(\Omega) = 1$ and $\Omega = A \cup A^c$ } From axioms.

$P(A \cup A^c) = 1$ and $A \cap A^c = \phi$

$\Rightarrow P(A \cup A^c) = P(A) + P(A^c) = 1$

$\Rightarrow \boxed{P(A) = 1 - P(A^c)}$

~~Eg: Rolling a die~~

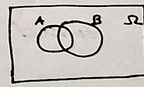
Mutually exclusive events

$A, B \in \mathcal{E}$ are called mutually exclusive if $A \cap B =$

If $A \cap B = \phi$

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

If $A \cap B \neq \phi$



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

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

Also

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

Also $B \cap (A \cap B^c) = \phi$

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

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

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

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

Now use P.I.E. (Incl. Excl).

↳ This can be generalized for arbitrary $n \in \mathbb{N}$.

Exhaustive set of Events

$S \subseteq \mathcal{E}$ is an exhaustive set of events ~~[illegible]~~ if the ~~[illegible]~~ $\bigcup_{A \in S} A = \Omega$

Let $\Omega = \{\omega_n\}_{n=1}^{\infty}$ be a countable sample space.

s.t. $\mathcal{E} = 2^{\Omega}$ $\qquad \omega_n :=$ possible outcomes.

$\Rightarrow \{\omega_n\} \in \mathcal{E} \quad \forall \omega_n \in \Omega$

$1 = P(\Omega) = P\left(\bigcup_{n=1}^{\infty} \{\omega_n\}\right) = \sum_{n=1}^{\infty} P(\{\omega_n\})$

$\Rightarrow \sum_{n=1}^{\infty} P(\omega_n) = 1$

Ex Let $p > 0$ be the probability of obtaining a head if a coin is tossed. Show that if we keep on ~~tap~~ tossing the coin, then the probability of obtaining a head eventually is 1.

Ans: by counting.

probability of all tails after n tosses. $= (1-p)^n$

~~Ans.~~ If we take the limit as $p > 0$ $\quad \underline{1-p < 1}$

$\lim_{n \to \infty} (1-p)^n = 0 \Rightarrow$ Probability of getting ...

$\Omega = \{H, TH, TTH, TTTH, \ldots\} \cup \{T, T, \ldots \times \infty\}$

$P(H) = p \qquad P(TH) = (1-p)p$

$P(T^n H) = (1-p)^n p$

$P(\{H, TH, \ldots\}) = P(H) + P(TH) + \ldots$

$= p + (1-p)p + (1-p)^2 p \ldots$

$\Rightarrow p\left(\sum_{n=0}^{\infty} (1-p)^n\right)$

$\Rightarrow \frac{p}{1-(1-p)} = 1$

~~[illegible]~~ Equally Likely

Let $S \subseteq \mathcal{E}$. We say the events in S are equally likely if $P(A) = P(B)$ for all $A, B \in S$.

More often we say the events A & B are equally likely if $P(A) = P(B)$

Random Experiment

An experiment of which we know the sample space but none of the outcomes occurs with certainty.

~~Non Random Experiments~~

Finite Sample Space

$\Omega = \{\omega_1, \dots, \omega_n\}$

$\mathcal{E} = 2^{\Omega}$

For $A \in \mathcal{E}$ with $A = \{\omega_{n_1}, \dots, \omega_{n_k}\}$

for n_i's $k \in \{1, \dots, n\}$

$P(A) = \sum_{i=1}^{k} P(\omega_{n_i}) = \boxed{\frac{\# A}{n}}$

countably additive since $\{\omega_{n_i}\}$'s are pairwise disjoint

This is only true for equally likely events. (all events in Ω has to be equally likely).

$P(\Omega) = 1$ and $P(\omega_i) = f$ [Equally likely]

$P(\omega_1 \cup \dots \cup \omega_n) = \sum_{i=1}^{n} P(\omega_i) = nf = 1$

$\Rightarrow f = 1/n$

$\Rightarrow P(A) = \sum_{i=1}^{k} P(\omega_{n_i}) = f(\# A)$

$= \boxed{\frac{\# A}{n}}$

This always requires all events in the sample space to be equally likely which frankly is trivial.

CLASSICAL Def^n of PROBABILITY

Suppose a random experiment results in m mutually exclusive, exhaustive and equally likely outcomes. Let there be $n(A)$ outcomes which are favourable to event A. Then the probability of occurrence of event A is

$$P(A) = \frac{n(A)}{m}$$

There is a problem. we use probability to define equally likely. so cyclic def^n.

For infinite sets the events can never be equally likely.

Relative freq def^n. of probability

If you repeat a random experiment n times and if an event A occurs $f_n(A)$ times, then $f_n(A)/n$ is called the rel. freq of A.

$$P(A) = \lim_{n \to \infty} \frac{f_n(A)}{n}$$

[This limit may not exists]

Boole's inequality

Probability space: $(\Omega, \mathcal{E}, P)$

$A_1, \ldots, A_n \in \mathcal{E}$. Then

$$P(A_1 \cup \ldots \cup A_n) \leq P(A_1) + \ldots + P(A_n)$$

Pf

$$B_i = \begin{cases} A_i & \text{if } i = 1 \\ A_i \setminus \bigcup_{j=1}^{i-1} A_j & \text{if } i > 1. \end{cases}$$

$$P\left(\bigcup_{i=1}^{n} B_i\right) = P\left(\bigcup_{i=1}^{n} A_i\right) \quad \forall j \in \{1, \ldots, n\}$$

B_i's are mutuall exclusive

$$= P(B_1) + \ldots + P(B_n)$$

$B_j \subseteq A_j \quad \forall j$

$\Rightarrow P(B_j) \leq P(A_j)$

$$\Rightarrow P(A_1 \cup \ldots \cup A_n) = \sum_{j=1}^{n} P(B_j) \leq \sum_{j=1}^{n} P(A_j).$$

$$\Rightarrow P(A_1 \cup \ldots \cup A_n) \leq P(A_1) + \ldots + P(A_n)$$

(Done)

~~Boole~~ Bon-ferroni's Inequality

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

Pf We proceed by induction.

$n=1$

$P(A_1) \geq P(A_1)$ [True]

$n=2$

$P(A_1 \cap A_2) = P(A_1) + P(A_2) - P(A_1 \cup A_2)$

Now $P(A_1 \cup A_2) \leq 1$.

$\Rightarrow -P(A_1 \cup A_2) \geq -1$

$\Rightarrow P(A_1 \cap A_2) \geq P(A_1) + P(A_2) - 1$. [True]

Induction hypothesis

We assume this holds for k

$$P(A_1 \cap \ldots \cap A_k) \geq P(A_1) + \ldots + P(A_k) - (k-1).$$

Let show it holds for $k+1$

$$P(\underbrace{A_1 \cap \ldots \cap A_k}_{A} \cap \underbrace{A_{k+1}}_{B}) \geq P(A) + P(B) - 1$$

$$\geq P(A_1 \cap \ldots \cap A_k) + P(A_{k+1}) - 1$$

use Induction hypothesis $\geq P(A_1) + \ldots + P(A_{k+1}) - (k-1) - 1$

$= -k$

[Thus proved via Induction]

Conditional Probability

$(\Omega, \mathcal{E}, P) :=$ Probability space.

$B \in \mathcal{E}$ has already occured as result of the experiment.

→ Since our ~~sample sp~~ event B has occured, the sample space reduces to B.

Now let $A \in \mathcal{E}$.

If $A \cap B = \phi$, then A will not occur as sample space is now B.

If $A \cap B \neq \phi$, then the probability of this event is measured realative to B.

Defn)
Let $B \in \mathcal{E}$ be s.t. $P(B) > 0$. Then for $A \in \mathcal{E}$, we define

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

In particular $P(B/B) = 1$ like $P(\Omega) = 1$

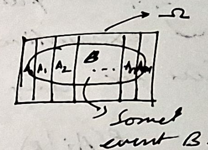
The last statement essentially shows how B is reduced to the sample space.

Lemma (Total prob)

Let $(\Omega, \mathcal{E}, P)$ be a prob space and let $\{A_i\}_{i=1}^{\infty}$ be pairwise mutually exclusive, exhaustive events s.t. $P(A_0) = 0$ ad $P(A_i) > 0 \ \forall i \in \mathbb{N}$

Then for any $B \in \mathcal{E}$, we have

$$P(B) = \sum_{i=1}^{\infty} P(A_i) P(B/A_i)$$



~~Pf~~ Let's do it once for the the finite case.

$B = (B \cap A_0) \cup (B \cap A_1) \cup \dots \cup (B \cap A_n)$

$B \cap A_0 \subseteq A_0$ and $B \cap A_1 \subseteq A_1$

Since A_0 and A_1 are mutually exclusive $B \cap A_0$ and $B \cap A_1$

Now we exclude the A_0 case as there might events (outcomes) which have zero prob.

Now however $P(B/A_i)$ won't be defined.

Thus we have $A_0 = 0$ to allow for such possibilities, we still need to consider A_0 as removing A_0 will not mean exhaustiveness

~~B~~ If $(A_1 \cup \dots \cup A_n) = \Omega$.

$$P(B) = P(\overset{0}{B \cap A_0}) + P(B \cap A_1) + \dots + P(B \cap A_n)$$

$$= \sum_{i=1}^{n} P(A_i) P(B/A_i)$$

We can't just extend to ∞ case.

If $\bigcup_{i=0}^{\infty} A_i = \Omega$.

$$P(B) = \sum_{j=1}^{\infty} A_j \, P(B/A_j). \quad [\text{Proved}]$$

Bayes Theorem

Let $(\Omega, \mathcal{E}, P)$ be a prob space and let $\{A_i\}_{i=1}^{\infty}$ be pairwise mutually exclusive and exhaustive events with $P(A_i) > 0 \ \forall i \in \mathbb{N}$. Then for any $B \in \mathcal{E}$ with $P(B) > 0$ we have

$$P(A_j/B) = \frac{P(A_j)\,P(B/A_j)}{P(B)}$$

Pf: $$P(A_j/B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(A_j)\,P(B/A_j)}{\sum_{i=1}^{\infty} P(A_i)\,P(B/A_i)}$$

Ex) Let's roll a fair die till we get an outcome. Call A_n the event in which we stop at the n^{th} roll. Let B be the event that all the outcomes preceding the last one are odd. Determine $P(A_m/B)$

Ans $P(A_n) = \frac{1}{6}\left(\frac{5}{6}\right)^{n-1}$ $\quad P(B/A_n) = \frac{P(B \cap A_n)}{P(A_n)}$

$$P(A_m/B) = \frac{P(A_m)\,P(B/A_m)}{P(B)} = \frac{(1/6)(3/6)^{n-1}}{1/6\,(5/6)^{n-1}} = \left(\frac{3}{5}\right)^{n-1}$$

$$P(A_m/B) = \frac{P(A_m)\,P(B/A_m)}{\sum_{n=1}^{\infty} P(A_n)\,P(B/A_n)}$$

$$= \frac{1/6\,(5/6)^{m-1}(3/6)^{m-1}}{\sum_{n=1}^{\infty} 1/6\,(5/6)^{n-1}(3/5)^{n-1}} = \frac{1}{2^m}$$

Ex-2 Suppose you find someone interesting and you'd like to ask him/her out for a coffee. Let's assume there are 3 mutually exclusive, exhaustive & equally likely cases.

A: He/She finds you interesting

B: He/She feels indifferent to you.

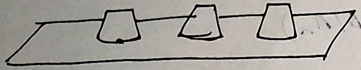
C: He/She is repulsed by you.

$P(Y/A) = 0.9$ $\quad P(Y/B) = 0.5$ $\quad P(Y/C) = 0.1$

i) Find the prob that he/she accepts your invitation ? Ans $= 0.5$

(ii) Given that he/she accepts your invitation find probability that she finds you interesting too.

Ex-3 [Monty hall]



One has a coin. Blah Blah.

Switch or no switch,

Ans A: Your initial guess is right $= 1/3$

A^c: " " " " wrong $= 2/3$

L: Your friend lifts an empty cup $= 1$.

we must compare $P(A|L)$ and $P(A^c/L)$

~~P(A|L)~~ $P(L) = P(L|A) = 1$

. L is independent of A.

$$P(A/L) = \frac{P(A)\ P(L|A)}{P(L)} = P(A) = 1/3$$

$$P(A^c/L) = \frac{P(A^c)\ P(L|A)}{P(L)} = P(A^c) = 2/3$$

So we should switch as given friend lifts an empty, you are wrong is more probable than you choose the correct ~~x~~ cup.

Simple thought expt (not thought but written).

Initial guess	Coin is under cup	Stick to initial guess	Switch
B	A	X	✓
C	A	X	✓
A	A	✓	X

WLOG can be extended 6 more can — $\frac{3}{9}$ — $6/9$

Ex "Random" Monty Hall

The presenter does not know where the coin resides and she lifts a cup which happens to empty. (by sheer luck)

Would you still switch? $P(L) = 2/3$

Independence of Events

Let A, B be 2 events such the occurrence of one event does not influence the other.

In other words, ~~xxx~~ (Assume non zero $P(A)$ & $P(B)$)

* $P(A|B) = P(A)$ and $P(B|A) = P(B)$

so, we can reduce it further;

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

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

↳ This can also be taken as the formal defn of independence of events.

Mutual vs Pairwise independence

If $S \subseteq \mathcal{E}$ s.t. $P(A\cap B) = P(A)P(B)$ $\forall\, A, B \in S$, we say the events are pairwise independence.

where if $\forall\, T \subseteq S$ we have

$$P\left(\bigcap_{A\in T} A\right) = \prod_{A\in T} P(A),$$

all the events in S are called mutually independent.

Statistician's Joke

The statistician computed the probability of having a bomb B_1 on the plane. Let's say $P(B_1) = p$. Apparently, under the assumption of independence, he computed the prob of having two bombs on the plane.

$$\cancel{P(B_1 B_2) = p^2} \quad P(B_1\cap B_2) = p^2$$

Instead, he must have computed

$$P(B_2|B_1) = \frac{P(B_1\cap B_2)}{P(B_1)} = \frac{p^2}{p} = p$$

The fucker carried his own bomb, which is incredibly stupid. However that is not a random event. $P(B_1) = 1$. So the sample space is reduced.

Erdös (1976) [Conjecture]

Given $m \in \mathbb{N}$, $m \geq 3$, any sequence $\{\alpha_n\}$ with $\alpha_n \in \mathbb{N}$ s.t. $\sum_{n=1}^{\infty} \frac{1}{\alpha_n} = \infty$, we have an arithmetic progression of length m in $\{\alpha_n\}$.

Mutually Exclusive vs Independence

~ $A\cap B = \phi$.

$\Rightarrow P(A\cap B) = 0$

• $P(A\cap B) = P(A)\cdot P(B)$

Random Monty Hall

$f'(f(x))f'(x) = 1$ 

A: Your initial guess is correct.

L: Your friend lifts a cup which happens to be empty.

$P(A|L)$ and $P(A^c|L)$

$P(L|A) = 1$

$P(L|A^c) = \frac{1}{2}$

$$P(A|L) = \frac{P(A\cap L)}{P(L)} = \frac{P(A)\,P(L|A)}{P(A)\,P(L|A) + P(A^c)\,P(L|A^c)}$$

$$= \frac{1}{2}$$ → no advantage in switching.

Base-Rate fallacy / false positive paradox.

Consider a very rare disorder which affects 0.1% of the population.

Let there be a test for this disease with 99% sensitivity (true positive).

99% specificity (true negative)

"Sensitivity" of a test is 100% × probability of identification of the disease in a person, when he/she is actually affected by that disease.

"Specificity" of a test is 100 × probability of correctly identifying a negative case

D: Presence of disease N: Negative test result.

P: Positive test result

$P(P|D) = 0.99$ $P(N|D^c) = 0.99$

$P(D) = 0.001$

i) Find $P(D|P)$.

ii) Given that a person has tested positive once, find the prob. that he/she the disease if he test positive again.

(iii) given that a person has tested ~~negative~~ positive twice, find the prob that he/she has the disease if he/she tests positive again.

Ans

i) $$P(D|P) = \frac{P(P|D)\,P(D)}{P(P)}$$

~~P(P) =~~ ~~P(P∩D)~~ $P(P) = P(P\cap D) + P(P\cap D^c)$

$= P(D)\,P(P|D) + P(D^c)\,P(P|D^c)$

$= 0.001 \times 0.99 + 0.999 \times 0.0$

~~0.0208~~

$P(D|P) \simeq 0.09$

ii) $P(D) \approx P(D/P)$

$P(D/P) = 0.09 = P(D)$

Since we do not care about $P(D^c)$ as ~~the~~ the person has already tested positive.

$$P(D/P) = \frac{P(D) \cdot P(P/D)}{P(D)P(P/D) + P(D^c)\,P(P/D^c)}$$

$$= \frac{0.09 \times 0.99}{0.09 \times 0.99 + 0.\cancel{8}91 \times \cancel{0.89}\ 0.01} \approx 0.91$$

---

base rate fallacy

	Prob of getting a seq of heads	Prob of having false positive reports of a test with 50% sensitivity
H, P	1/2	1/2
HH, PP	1/4	1/4
HHH, PPP	1/8	1/8

Yesterday's problem

Consider a test with 99% sensitivity and 99% specificity for a very rare disease which affects only 0.1% of the population.

$P(D/P) = 0.09 \rightarrow P(D/PP) = 0.91 \rightarrow P(D/PPP) = 0.99$

$P(D/(P)_n)$ is a monotonically increasing seq if

sensitivity + specificity > 1

$\Rightarrow P(\cancel{D}/P/D) + P(N/D^c) > 1$

and $P(D/(P)_n) \longrightarrow 1$ as $n \to \infty$.

Why does one test not suffice?

The chance of being affected is 1 in 1000. The specificity and sensitivity is high but since the disease is very rare; the chance of having the disease is only 1 out of 11, given a ⊕ve test.

The Coffee Problem

$P(A/Y) = 0.6 \qquad P(Y/A) = 0.9$

$P(N/A^c) = 1 - P(Y/A^c)$

$$= 1 - \frac{P(Y \cap A^c)}{P(A^c)}$$

$$= 1 - \frac{P(Y \cap A^c)}{1 - P(A)} = 1 - \frac{P(Y)\,P(A^c/Y)}{1 - P(A)}$$

$$P(N|A^c) = 1 - \frac{P(Y)(1-P(A|Y))}{1-P(A)}$$

$$\approx 1 - \frac{\cancel{1}/2\,(1-0.6)}{0.67}$$

$$\approx 0.7$$

$P(Y|A)$

$\underline{P(A) = \frac{1}{3}} = P(A|Y)\frac{P(Y)}{P(A)}$

$\frac{P(Y|A)}{P(A|Y)} P(Y)$

$\frac{0.9}{0.6} \quad 0.5$

$\frac{3}{2} \times \frac{1}{2} = \frac{3}{4}$

$P(A|\underbrace{YY\ldots Y}_{n \text{ times}}) \to 1$ as $n \to \infty$

Random Variables

Roll a die : You'll get an outcome in $\{1, 2, \ldots, 6\}$

Ask a person for his/her ph. no. You'll get a 10 digit no. (assuming you get a numerical response)

Guess the 1st letter of a prayer's name.

We may generally quantify the outcomes by assigning ~~the~~ 26 real nos to the Roman alphabet

Generally a random variable is a measurable function.

Analogy b/w measurable and cts functions.

<u>cts funcs</u>

Let $f: S \to S'$ be cts
↳ topological space.

$\Updownarrow$

For every open set $O' \subseteq S'$, $f^{-1}(O') \subseteq S$ is open.

Let $f: S \to S'$ be a measurable f^n.
↳ (measurable space)

$P: \mathcal{E} \to [0,1] \to$ [Prob measure]

$L: \mathbb{R} \to \mathbb{R} \to$ [length measure]

If $O' \subseteq S'$ is measurable, then $f^{-1}(O') \subseteq S$ is also measurable.

RANDOM VARIABLE

A real valued function on the sample space s.t. the pre-image of every interval is an event.

* For this course by a r.v, we'll always mean a real valued r.v.

Def^n (Probability dist func)

Let $(\Omega, \mathcal{E}, P)$ be a prob space and $X: \Omega \to \mathbb{R}$ be a random variable. Then X translates $(\Omega, \mathcal{E}, P)$ to $(\mathbb{R}, \mathcal{E}_X, P_X)$, where $\mathcal{E}_X = \{A \subseteq \mathbb{R} / X^{-1}(A) \in \mathcal{E}\}$

$P_X(A) = P(X^{-1}(A))$

Enumerate the rationals in $[-1,1]$

as $\{q_1, q_2, \ldots\}$ then consider

$\{q_i + Q'\}$ $\quad q_i + Q' \subseteq [-1, 2]$

$$l(q_i + Q') = l(Q') = c$$

$$c \leq \sum_{i=1}^{\infty} l(q_i + Q') \leq 3.$$

Notation

Henceforth, will use the notations $P_X(A)$, $A \in \mathcal{E}_X$ & $P(X \in A)$ interchangeably.

In particular, will write $P_X((-\infty, a])$ as $P(X \leq a)$.

Random Variables

$X : \Omega \to \mathbb{R}$

s.t. $X^{-1}((-\infty, a]) \in \mathcal{E}$ $\forall a \in \mathbb{R}$.

e.g. $(-\infty, a) = \bigcup_{n=1}^{\infty} (\infty, a - \frac{1}{n}]$

The facts that $\mathcal{E}$ is closed under countable unions and complementation is also carried over to $\mathcal{E}_X$.

e.g. Any interval in $\mathbb{R}$ can be written in terms of $(-\infty, a]$ where $a \in \mathbb{R}$.

(a, b)

$\cup (c, d)$

$b < b$

Lemma 1

$\forall x \in \mathbb{R}$, we have $\{x\} \in \mathcal{E}_X$

Pf

$$\{x\} = [x, x]$$

$$= (\underbrace{(-\infty, x)}_{\mathcal{E}_X} \cup \underbrace{(x, \infty)}_{\mathcal{E}_X})^c$$

As $\mathcal{E}_X$ is closed under countable union and complementation, $\{x\} \in \mathcal{E}_X$.

Cor

$\forall$ r.v. $X : \Omega \to \mathbb{R}$,

$\Downarrow$ $\quad P(X = x) = P_X(X^{-1}(x))$ is well defined.

Probability mass function

Defn [CDF: Cumulative distribution func]

For $a \in \mathbb{R}$, we define $F_X(a) = P(X \leq a)$

* The cdf is an increasing function because for $a \leq b$, $F_X(a) = P(X \leq a) = P(X^{-1}(-\infty, a]) \leq$

$F_X(b) = P(X \leq b) = P(X^{-1}(-\infty, b])$

Cor: The CDF ~~iff~~ ~~is~~ is continuous $\Leftrightarrow$ its left continuous.

$\Updownarrow$

The CDF is uniformly cts

Defn (cts random variable)

A random variable with a continuous CDF.

Thm Let X be a r.v.

Then we be a r.v.

Thus we have

$P(X = x) > 0$ iff F_X is discontinuous at x.

Pf

Since F_X is monotone, it has only jump discontinuous. Since F_X is right continuous, a jump discontinuity can only occur at x iff

$$F(x^-) < F_X(x)$$

$\Leftrightarrow$ $P(X < x) < P(X \leq x)$

$\Leftrightarrow$ $P(X < x) < P(X < x) + P(X = x)$

$\Leftrightarrow$ ~~$P(X \leq x)$~~ $P(X = x) > 0$

Cor The set of values of an r.v. where its PMF is non-zero is countable.

Pf Since F_X is monotone, the set of its discontinuities is countable.

Thm The sum $\sum_{x : P(X = x) > 0} P(X = x)$ is well defined i.e. convergent for every random variable.

Pf $D_X :=$ The set of discontinuities of F_X

$$0 \leq P(X \in D_X) = \sum_{x \mid P(X = x) > 0} P(X = x) \leq P(X \in \mathbb{R}) = 1$$

~~If X is cts, then $D_X = \phi$~~

$\sum_{D_X} P(X = x)$ is a non decreasing function,

Since D_X is countable, $\sum_{D_X} P(X = x)$ is well defined as $\sum_{D_X} P(X = x)$ is bdd and monotonic

---

Defn (Classification of r.v.)

$$\sum_{x \mid P(X = x) > 0} P(X = x) = \begin{cases} 0 & \text{if } X \text{ is a cont. r.v.} \\ \mu & \text{for some } \mu \in (0,1), \text{ Then } X \text{ is mixed r.v.} \\ 1 & X \text{ is a discrete r.v.} \end{cases}$$

Properties of CDF

$F : \mathbb{R} \to \mathbb{R}$

i) F is non-~~increasing~~ decreasing

ii) $F(x) \in [0,1]$

iii) $\lim_{x \to \infty} F(x) = 1$

iv) $\lim_{x \to -\infty} F(x) = 0$

(v) F is right continuous at every pt.

Probability Density Function (PDF)

Let $X : \Omega \to \mathbb{R}$ be a r.v.

If $\exists$ a fx $f_X : \mathbb{R} \to \mathbb{R}_{\geq 0}$ s.t.

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt \quad \forall x \in \mathbb{R}$$

we call f_X a PDF of X.

(*) F must be continuous. (take $\int_{-\infty}^{x+\varepsilon} f_X(t) = \int_{-\infty}^{x+\varepsilon} f_X(t)\,dt$

Since $\lim_{x \to \infty} F(x) = 1$

$$\Rightarrow \int_{-\infty}^{\infty} f_X(t)\,dt = 1.$$

$$P(a < X \leq b) = F_X(b) - F_X(a)$$

$$= \int_a^b f_X(t)\,dt$$

f_X need not be continuous. However, if it is a continuous func, then the fundamental thm of calculus $\Rightarrow$ $f_X(t) = F_X'(t)$.

Note

So if F_X is differentiable, its natural to choose $f_X = F_X'$

Mixed Random Variables

Thm

If X is a random variable s.t. ~~E~~ F_X has a discontinuity and F_X is continuous and strictly increasing in some interval $I \subseteq \mathbb{R}$, $I \neq \phi$.

then X is a mixed random r.v.

Pf

Let $a, b \in I$ s.t. $a < b$

$\Rightarrow F_X(b) > F_X(a)$

Let D_X be the set of discontinuities of F_X

$D_X \cap (a, b] = \phi$

$\Rightarrow P(X \in D_X \cup (a,b]) \leq P(X \in \mathbb{R}) = 1$

$\Rightarrow P(X \in D_X) + P(X \in (a,b]) \leq 1.$

$$\Rightarrow \sum_{x \cdot P(X=x) > 0} P(X=x) + \underbrace{F_X(b) - F_X(a)}_{> 0} \leq 1$$

$$\Rightarrow \sum_{x | P(X=x) > 0} P(X=x) \leq 1 - (F_X(b) - F_X(a))$$

$$\Rightarrow \sum_{x | P(X=x) > 0} P(X=x) < 1 \Rightarrow X \text{ is a mixed random r.v.}$$

Hal HaHa!!!

Discrete Random Variable

Thm

If a random variable X takes only countably many values, then X is a discrete r.v. $\{\frac{1}{n}\}$

Pf

Let C be the range of values assumed by X and

let $C_+ = \{x \in C \mid P(X=x) > 0\}$ &

$C_0 = \{x = C \mid P(X=x) = 0\}$

$C = C_+ \cup C_0$

$C_+ \cap C_0 = \phi$

$$\sum_{x \in C} P(X=x) = \sum_{x \in C_+} P(X=x) + \sum_{x \in C_0} P(X=x)$$

Since X takes only in C, we have

$P(X \in \mathbb{R} \setminus C) = 0$

$\Rightarrow P(X = x) = 0 \quad \forall x \in \mathbb{R} \setminus C$ as $P(x) \geq 0 \ \forall x$

$1 = P(X \in \mathbb{R})$

$= P(X \in C) + \overbrace{P(X \in \mathbb{R} \setminus C)}^{0}$

$\Rightarrow P(X \in C_+) + P(X \in C_0) = 1$

$\Rightarrow \sum_{x \in C_+} P(X=x) = 1 \Rightarrow X$ is discrete

[Proved]

Cor

If a random variable X takes values in a discrete set, it is a discrete r.v.

$\Downarrow$

A set of isolated pts in $\mathbb{R}$.

For some $\varepsilon > 0$, and $x \in A$,

$V_\varepsilon(x) \cap A = \{x\}$.

Defn

A Bernoulli / Binary r.v. assumes exactly two distinct values.

Notation

If is a Bernoulli r.v. taking values in $\{0, 1\}$ (output)

s.t. $P(X=1) = p$ then we write

$X \sim \text{Bernoulli}(p)$ → parameter

Eg (Discrete Uniform r.v.)

If X is a rv assumes n distinct values with equal probabilities, we write

$X \sim \text{uniform}(n)$ ⇒ $\sim$ read as "follows"

$X \sim \text{uniform}(2)$

$\Rightarrow X \sim \text{bernoulli}(\frac{1}{2})$.

Notation, (...)

Let x & Y be two random variables on the sample space, Then for $A \in \mathcal{E}_x$ and $B \in \mathcal{E}_Y$.

$$P(x \in A, y \in B) := P(\{x \in A\} \cap \{y \in B\})$$

$A \in \mathcal{E}_x$, $x^{-1}(A) \in \mathcal{E}$

$B \in \mathcal{E}_Y$, $x^{-1}(B) \in \mathcal{E}$

$$P(\{x \in A) \cap (x \in B\}) = P(x \in A, y \in B)$$

If $P(Y \in B) > 0$, we define the cond^n prob

$$P(x \in A | y \in B) = \frac{P(x \in A, y \in B)}{P(y \in B)}.$$

Def^n [Independence of r.v]

Two r.v. x and y are said to be independent if $P(x \in A | y \in B) = P(x \in A)$ $\forall A \in \mathcal{E}_x$, $B \in \mathcal{E}_Y$ s.t. $P(y \in B) > 0$.

$P(y \in B | x \in A) = P(y \in B)$ $\forall A \in \mathcal{E}_x, B \in \mathcal{E}_Y$ s.t. $P(x \in A) > 0$

$\Leftrightarrow P(x \in A, y \in B) = P(x \in A) P(y \in B)$

[Better def^n].

Def^n [Pairwise independence]

Let $\{x_i\}_{i \in S}$ be a set of r.v. s.t

$$P(x_i \in A_i ; x_j \in A_j) = P(x_i \in A_i) P(x_j \in A_j)$$

$\forall i, j \in S$ and $\forall A \in \mathcal{E}_{x_i}$ and $A_j \in \mathcal{E}_{x_j}$

Then the set of r.v. are pairwise independent

Def^n [Mutual Independence]

Let $\{x_i\}_{i \in S}$ be a set of r.v. s.t.

$$P\left(\bigcap_{i \in T} \{x_i \in A_j\}\right) = \prod_{i \in T} P(x_i \in A_i) \quad \forall A_i \in \mathcal{E}_{x_i}$$

$\bigcap_{i \in T} \{x \in A_j\}$

$\forall$ countable subsets T of S.

Def^n (iid) random variables.

Independent and Identically distributed r.v. $\{x_i\}_{i \in S}$.

- $\{x_i\}_{i \in S}$ are mutually independent.
- $F_{x_i}(x) = F_{x_j}(x)$ $\forall i, j \in S$ & $\forall x \in \mathbb{R}$.

Mutual independence $\Rightarrow$ Pairwise Independence.

Exercise

Let $\{a_n\}_{n \in \mathbb{N}}$ be a seq. of non-negative numbers s.t. $\sum_n a_n = 1$. Let S be a countable subset of $\mathbb{R}$.

Let's give the elements of S some enumeration.

Define a fn, $F : \mathbb{R} \to [0,1]$ by

$$F(x) = \sum_{\substack{s_n \in S \\ s_n \leq x}} a_n$$

Show that F is the CDF of a discrete r.v.

Expectation of a func g of a random variable.

$g : \mathbb{R} \to \mathbb{R}$ and let X be a random variable.

If the sum $\sum_{x \mid P(X=x)>0} g(x)\, P(X=x)$ converges absolutely,

then we call this as the expectation of g

and we write it as : $E(g(X))$

If X is a cts. r.v. with a PDF, then

the $E(g) := \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$ iff

the integral converges absolutely.

Moments (n^{th} moment) $\Rightarrow g := x^n$

Variance : $g(x) = (x-\mu)^2$

$\mu = E(X)$

Defn [Random Vector]

Let $X_1, \ldots, X_n$ be random variables defined on the same sample space. Then the ordered tuple

$\tilde{X} := (X_1, \ldots, X_n)$

is called a random vector.

Defn [CDF of a random vector] [Joint CDF]

$F_{\tilde{X}}(x_1, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$

Gen. from $\mathbb{R}^1 \longrightarrow \mathbb{R}^n$

Defn [PMF of a random vector] [Joint PMF]

$P(X=x_1, \ldots, X=x_2) = P(\{X_1 = x_1\} \cap \{X_2 = x_1\} \cap \ldots \cap \{X_n = x$

Ex: Toss a fair coin thrice and let

X: No. of heads obtained Y: No. of tails obtai

$X=i$ \ $Y=j$	0	1	2	3	$P(X=i)$
0	0	0	0	1/8	1/8
1	0	0	3/8	0	3/8
2	0	3/8	0	0	3/8
3	1/8	0	0	0	1/8
$P(Y=j)$	1/8	3/8	3/8	1/8	

$\longrightarrow$ Marginal PMF.

Summing over all except the ith value in the tuple

Marginal PMF [Defn]

Let $X_1, \ldots, X_n$ be jointly distributed ~~F~~ discrete r.v.s (i.e. $\tilde{X} = (X_1, \ldots, X_n)$).

Then the map $x_i \longmapsto P(X = x_i)$ is called the 'i^{th}' marginal PMF of $\tilde{X}$.

Lemma

Let $X_1, \ldots, X_n$ be jointly distributed discrete r.v.s. Then the marginal PMF of X_1 is given by

$$P(X_1 = x_1) = \sum_{x_2, \ldots, x_n \,/\, P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) > 0} P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

Pf: Follows from Total prob. lemma

Defn [Conditional PMF and Conditional CDF]

Let X, Y and Z be jointly distributed ~~F~~ discrete r.v.'s. Then we define the conditional PMF of X given $Y = y$ by

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

Given $P(Y = y) > 0$

If X, Y are independent

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

The conditional CDF of X is given ~~by~~ $Y = y$ is

$$P(X \leq x \mid Y = y) = \frac{P(X \leq x, Y = y)}{P(Y = y)}.$$

$$= \sum_{\substack{r \leq x \\ P(X = r, Y = y) > 0}} \frac{P(X = r, Y = y)}{P(Y = y)}$$

Defn [Joint PDF of a cts. random vector]

Let $X_1, \ldots, X_n$ be cts r.v. If $\exists\, f_{\tilde{X}} : \mathbb{R}^n \longrightarrow \mathbb{R}^n_{\geq 0}$ s.t.

$$F_{\tilde{X}}(x_1, \ldots, x_n) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\int_{-\infty}^{x_3} \cdots \int_{-\infty}^{x_n} f_{\tilde{X}}(t_1, \ldots, t_n)\, dt_1 \ldots dt_n$$

then $f_{\tilde{X}}$ is called the joint PDF of $X_1, \ldots, X_n$.

If $F_{\tilde{X}}$ is differentiable to n^{th} order, then

$$\boxed{f_{\tilde{X}}(t_1, \ldots, t_n) = \frac{\partial^n F_{\tilde{X}}(t_1, \ldots, t_n)}{\partial t_1 \partial t_2 \ldots \partial t_n}}.$$

Marginal PDF

The i^{th} marginal PDF is the function which sends $x_i \longmapsto f_X(x_i)$ ~~the PDF of the~~ ~~in the fn which sends~~ is called the i^{th} marginal PDF.

The conditional PDF of X given $Y=y$ is

$$f_{X|Y}(x/y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

Conditional CDF

$$F_{X|Y}(x/y) = \frac{\int_{-\infty}^{x} f_{X,Y}(t,y)\,dt}{f_Y(y)}.$$

Lemma Let f_X and f_Y be cts as well.
Let X and Y be two independent cts. r.v. defined over the same sample space, and let

$f_{X,Y}: \mathbb{R}^2 \to \mathbb{R}^2_{\geq 0}$ be their joint PDF.

Then $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$. if

Sketch of the PDF. $f_{X,Y}$ is cts, (Then marginal pdf is cts).

Since f_X & f_Y are cts, we have

$$P(x-\varepsilon \leq X \leq x+\varepsilon,\ y-\varepsilon \leq Y \leq y+\varepsilon)$$

$$= P(x-\varepsilon \leq X \leq x+\varepsilon)\, P(y-\varepsilon \leq y \leq y+\varepsilon)$$

$$= \left(\int_{x-\varepsilon}^{x+\varepsilon} f_X(t)\,dt\right)\left(\int_{y-\varepsilon}^{y+\varepsilon} f_Y(t)\,dt\right) \cong f_X(x) f_Y(y)\, \varepsilon^2$$

Dividing both sides by ε^2 and taking the limit and taking the limit as $\varepsilon \to 0$ we get

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

If X and Y are independent, then the conditional PDF of X given Y is

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

$\Rightarrow$ conditional CDF of X is also given by F_X.

Dirichlet Thm # Imp for a lotta theorems we prove later.

Let $\sum_{n=1}^{\infty} \alpha_n$ be an absolutely cgt. series. Then for any bijection $f: \mathbb{N} \to \mathbb{N}$, the series $\sum_{n=1}^{\infty} \alpha_{f(n)}$ converges absolutely and we have

$$\sum_{n=1}^{\infty} \alpha_{f(n)} = \sum_{n=1}^{\infty} \alpha_n.$$

Pf:- Proving only the last equality suffices. Since we may replace α_n by $|\alpha_n|$ and α_{n+1} by $|\alpha_{n+1}|$ in it to obtain absolute convergence.

Let S_n and S_n' denote the partial sums of the series $\sum_k \alpha_n$ and $\sum_k \alpha_{f(n)}$.

Since $\{S_n\}$ converges by defn, we have.

$\exists N \in \mathbb{N}$ s.t. $\forall m, m' \in \mathbb{N}$ s.t.

$m' \geq m \geq N$, we have

$$\left|\sum_{n=m}^{m'} \alpha_n\right| = |S_{m'} - S_{m-1}| < \varepsilon.$$

Since and

$$\left|\sum_{n=m}^{m'} \alpha_n\right| < \varepsilon$$

For $k \geq N \geq M$

$$|S_{k'} - S_k| \leq \sum_{n=N+1}^{k'} |\alpha_n| < \varepsilon$$

where $k' = \max\{k, f(1), \ldots, f(M)\}$

$\Rightarrow$ $S_{k'}$ and S_k converge to the same pt.

$\Rightarrow$ Every rearrangement of the same series converges to the same pt as the original series.

Quiz on 4. 3. 24 instead of 7. 3. 24.

Conditional CDF of jointly distributed cts random variables. Let X, Y be two jointly distributed cts r.v. and let $f_{X,Y}$ be their joint PDF while f_X and f_Y are the marginal PDFs.

$$F_{X|Y}(x,y) = P(X \leq x \mid Y = y)$$

$$= \int_{-\infty}^{x} f_{X|Y}(t|y)\, dt$$

$$= \frac{1}{f_Y(y)} \int_{-\infty}^{x} f_{X,Y}(t,y)\, dt$$

In particular if X and Y are independent,

$$f_{X,Y}(t,y) = f_X(t)\, f_Y(y)$$

$$F_{X|Y}(x,y) = \int_{-\infty}^{x} f_X(t)\, \cancel{f_Y(y)}\, dt \left(\frac{1}{\cancel{f_Y(y)}}\right)$$

$$= \int_{-\infty}^{x} f_X(t)\, dt$$

$$= F_X(x)$$

Thus we can see conditional CDF equals marginal CDF when r.v.s are independent.

If $X_1, \dots, X_n$ are mutually independent cts. r.v., defined over the same space and if

$\{i_1, \dots, i_m\} \subseteq \{1, 2, \dots, n\}$ s.t.

$f_{i_1 i_2 \dots i_m}$ is the joint PDF of $X_{i_1}, \dots, X_{i_m}$, then we have

$$f_{X_{i_1} \dots X_{i_m}} = f_{X_{i_1}}(x_{i_1}) \dots f_{X_{i_m}}(x_{i_m})$$

$$\forall x_{i_1}, \dots, x_{i_m} \in \mathbb{R}$$

Thm

If X, Y are r.v.s over the same sample space having finite expectations, then $\forall\, a, b \in \mathbb{R}$, we have $E(aX+bY) = aE(X) + bE(Y)$

Pf (Discrete case).

X, Y are discrete r.v.s

$$aE(X) + bE(Y) = a \sum_{x \mid P(X=x)>0} x P(X=x) + b \sum_{y \mid P(Y=y)>0} y P(Y=y)$$

$$= a \sum_{x \mid P(X=x)>0} x \sum_{y \mid P(X=x, Y=y)>0} P(X=x, Y=y)$$

$$+ b \sum_{y \mid P(Y=y)>0} y \sum_{x \mid P(X=x, Y=y)>0} P(X=x, Y=y)$$

By dirichlet Theorem.

$$aE(X) + bE(Y) = \sum_{x, y : P(X=x, Y=y)>0} (ax+by) P(X=x, Y=y)$$

$$= E(aX+bY)$$

(continuous case).

Let X, Y be jointly distributed cts. r.v. with joint PDF $f_{X,Y}$. Then.

$$aE(X) + bE(Y) = a \int_{-\infty}^{\infty} x f_X(x)\,dx + b \int_{-\infty}^{\infty} y f_Y(y)\,dy$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \text{ and}$$

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

$$aE(X) + bE(Y) = a \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy$$

$$+ b \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy$$

[Fubini's theorem]

$$= a \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)\,dx\,dy$$

$$+ b \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x,y)\,dx\,dy$$

$$= E(aX+bY)$$

~~Defn [Expectation of a r.v. vector]~~

Cor. If $X_1, \dots, X_n$ are jointly distributed r.v. with finite expectation, then $\forall a_1, \dots, a_n \in \mathbb{R}$.

$$E(a_1X_1 + \dots + a_nX_n) = a_1E(X_1) + \dots + a_nE(X_n)$$

Defn (Expectation of a fn of n r.v. defined over the same ~~spa~~ sample space):

$X_1, \dots, X_n$ are r.v. defined over the same sample space and let $g: \mathbb{R}^n \to \mathbb{R}$.

discrete r.v.

$$E(g) = \sum_{x_1, \dots, x_n \mid P(X_1 = x_1, \dots, X_n = x_n) > 0} g(x_1, \dots, x_n) P(X_1 = x, \dots, X_2 = x).$$

This should absolutely cgt.

cts. r.v.

$$E(g) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) f_{X_1 \dots X_n}(x_1, \dots, x_n) \, dx_1 \dots dx_n$$

iff the RHS converges absolutely.

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

$$\sigma_X = \sqrt{Var(X)}$$

↳ standard deviation of X.

Thm

Let X and Y be independent r.v. defined defined over the same sample space s.t. $E(X)$ and $E(Y)$ exists, then

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

Pf Discrete case

$$E(X)E(Y) = \sum_{x \mid P(X=x) > 0} x P(X=x) E(Y)$$

$$= \sum_{x \mid P(X=x) > 0} x P(X=x) \sum_{y \mid P(Y=y) > 0} y P(Y=y)$$

$$= \sum_{x, y \mid P(X=x, Y=y) > 0} xy \, P(X=x) P(Y=y)$$

Now if X and Y are independent,

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

$$\Rightarrow E(X)E(Y) = \sum_{x, y \mid P(X=x, Y=y) > 0} xy \, P(X=x, Y=y)$$

$$= E(XY)$$

Continuous case follows by Fubini's Theorem.

$\underline{\text{Def}^n}$ [Covariance]

A crude measure of linearly dependency between 2 r.v.s.

Let X & Y be r.v. (jointly distributed) s.t. $E(X)$ and $E(Y)$ exists. Thus we define

$$\operatorname{cov}(X,Y) = E(XY) - E(X)E(Y)$$

$\underline{\text{Def}^n}$ [Correlation coefficient of X and Y]

$$\rho_{xy} = \frac{\operatorname{cov}(x,y)}{\sigma_x \sigma_y}$$

For any two r.v.s we have $-1 \le \rho_{x,y} \le +1$

$\underline{\text{Proof}}$ $\operatorname{Var}\left(\frac{x}{\sigma_x} \pm \frac{y}{\sigma_y}\right)$

Rewriting the notes bcoz I slept in class.

$$\operatorname{Var}(X) = E\left(\left(X - E(X)\right)^2\right)$$

Note: $\operatorname{Var}(X)$ is a non-negative quantity.

We define standard deviation of X

$$\sigma_x = \sqrt{\operatorname{Var}(X)}$$

$\underline{\text{Thm}}$

Let X and Y are independent r.v.s defined over the ~~samp~~ same sample space, s.t. $E(X)$ and $E(Y)$ exists.

Then

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

$\underline{\text{Pf}}$ (Discrete case)

We know $E(X)$ and $E(Y)$ exists

$$* \quad E(X) = \sum_{x \mid P(X=x)>0} x\,P(X=x)$$

$$E(Y) = \sum_{y \mid P(\cancel{X=}Y=y)>0} y\,P(\cancel{X=}Y=y)$$

Let $D_x = \{x \mid P(X=x)>0\}$

$D_y = \{y \mid P(Y=y)>0\}$

$$E(X)E(Y) = \left(\sum_{x \mid P(X=x)>0} x\,P(X=x)\right)\left(\sum_{y \mid P(Y=y)>0} y\,P(Y=y)\right)$$

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

Since they converge absolutely * by def^n, we can rearrange them.

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

$$= \sum_{D_x, D_y} xy\,P(X=x, Y=y) = E(XY)$$

Cont case

$$E(X).E(Y) = \int_{-\infty}^{\infty} x f_X(x)\,dx \int_{-\infty}^{\infty} y f_Y(y)\,dy$$

By Fubini's theorem,

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

Since X and Y are independent

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$$

$$= E(XY)$$

Defn. [Covariance]

A crude measure between linear dependency between two random variables

Let X & Y be two r.v.s (jointly distributed) s.t. E(X) and E(Y) exists.

Then we define

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

Eg-1 Let $Y = aX + b$. Then $cov(X,Y)$?

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

$$= E(X(aX+b)) - E(X)E(aX+b)$$

$$= aE(X^2) + b\cancel{E(X)} - a(E(X))^2 - bE(X)$$

$$= a\left(E(X^2) - (E(X))^2\right)$$

$$= a\,(Var(X)).$$

Eg-2 $Y = X^2$

Let X be a r.v. s.t. $E(X) = E(X^3) = 0$

Then

$$cov(X,Y) = E(X^3) - E(X)E(X^2) \quad P(X=x) = \begin{cases} 1 & x=0 \\ 0 & o/w \end{cases}$$

$$= 0$$

But X and Y are dependent.

So $cov(X,Y) = 0 \not\Rightarrow X, Y$ are independent

X, Y are indep $\Rightarrow cov(X,Y) = 0$

Def[n] [correlation coefficient]

$$\rho_{x,y} = \frac{\operatorname{cov}(x,y)}{\sigma_x \sigma_y}$$

$\sigma_x, \sigma_y \neq 0$

Thm Let x, y be r.v.s. we have

$$-1 \le \rho_{x,y} \le 1$$

Pf

we use the fact that

$$\operatorname{Var}\left(\frac{x}{\sigma_x} \pm \frac{y}{\sigma_y}\right) \geq 0$$

$$E\left(\left(\frac{x}{\sigma_x} \pm \frac{y}{\sigma_y}\right)^2\right) - \left(E\left(\frac{x}{\sigma_x} \pm \frac{y}{\sigma_y}\right)\right)^2 \geq 0$$

$$\frac{E(x^2)}{\sigma_x^2} + \frac{E(y^2)}{\sigma_y^2} \pm \frac{2E(xy)}{\sigma_x \sigma_y}$$

$$- \left(\frac{E(x)}{\sigma_x}\right)^2 - \left(\frac{E(y)}{\sigma_y}\right)^2 \mp \frac{2E(x)E(y)}{\sigma_x \sigma_y} \geq 0$$

$$\frac{\operatorname{Var}(x)}{\sigma_x^2} + \frac{\operatorname{Var}(y)}{\sigma_y^2} \pm \frac{2\operatorname{cov}(x,y)}{\sigma_x \sigma_y} \geq 0$$

$$\Rightarrow 2 \pm 2\frac{\operatorname{cov}(x,y)}{\sigma_x \sigma_y} \geq 0$$

$$\Rightarrow 1 \pm \rho_{x,y} \geq 0$$

$$\Rightarrow -1 \le \rho_{x,y} \le 1$$

Eg $\operatorname{cov}(x,y) = a \operatorname{Var}(x)$ $\quad Y = aX + b$

$$\rho_{x,y} = \frac{\operatorname{cov}(x,y)}{\sigma_x \sigma_y} \qquad \sigma_y = |a|\sigma_x$$

$$= \frac{a \operatorname{Var}(x)}{|a| \sigma_x^2}$$

$$= \frac{a}{|a|} = \operatorname{sgn}(a)$$

$$\rho_{x,y} \in \{-1, 1\}$$

Covariance matrix

Let $X_1, \ldots, X_n$ be jointly distributed r.v.s.

Then their covariance mx is defined by

$$\left(\mathrm{cov}(X_i, X_j)\right)_{n\times n} \quad \text{Now } \mathrm{cov}(X_i, X_j) = \mathrm{cov}(X_j, X_i) \ \forall i, j$$

$\Rightarrow$ covariance matrix is symmetric $\Rightarrow$ Hermitian.

$\Rightarrow$ Eigenvalues are real.

$\mathrm{cov}(X, X) = \mathrm{Var}(X)$

Conditional expection & conditional variance

Discrete case

$$P(X=x \mid Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)}$$

Continuous case

$$f_{X,Y}(X=x \mid Y=y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

$g: \mathbb{R} \to \mathbb{R}$

$$E(g(X) \mid Y=y) = \sum_{x \mid P(X=x, Y=y) > 0} g(x)\, P(X=x \mid Y=y)$$

iff this sum converges absolutely.

$E(X \mid X=y)$ is the conditional expection of X, given $Y=y$

Conditional variance

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

$$\mathrm{Var}(X \mid Y=y) = E(X^2 \mid Y=y) - \left(E(X \mid Y=y)\right)^2$$

HW

$$\mathrm{Var}(X \mid Y=y) = E\left(\left(X - E(X \mid Y=y)\right)^2 \mid Y=y\right)$$

$E(X \mid Y=y)$ is a fxn of y. $\quad E(X \mid Y) = E(X \mid Y=y)$

$D_y = \{y \mid P(Y=y) > 0\}$

$D_{x,y} = \{x \mid P(X=x, Y=y) > 0\}$

Thm

$$E(E(X \mid Y)) = E(X)$$

Pf (Discrete)

$$E(E(X \mid Y)) = \sum_{D_y} E(X \mid Y=y)\, P(Y=y)$$

$$= \sum_{D_y} \sum_{D_{xy}} x\, P(X=x \mid Y=y)\, P(Y=y)$$

$$= \sum_{D_y} \sum_{D_{xy}} x\, P(X=x, Y=y)$$

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

$$= \sum_{D_x} x\, P(X=x) = E(X).$$

Cts case

$$E(E(X|Y)) = \int_{-\infty}^{\infty} E(X|Y=y) f_{Y}(y)\, dy$$

$$\left[\{x \mid P(X=x \mid Y=y) \geq 0\} = \{x \mid P(X=x, Y=y) > 0\} \right] \quad P(Y=y)$$

$$\boxed{P(Y=y) > 0}$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\, dx\, f_Y(y)\, dy$$

$$= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} \frac{f_{X,Y}(x,y)}{\cancel{f_Y(y)}} \cancel{f_Y(y)}\, dy\, dx$$

$$= \int_{-\infty}^{\infty} \left(x \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy \right) dx$$

$$= \int_{-\infty}^{\infty} x\, f_X(x)\, dx$$

$$= E(X).$$

Thus for both cases

$$E(E(X|Y)) = E(X)$$

Thm

$$Var(X) = Var(E(X|Y)) + E(Var(X|Y))$$

Pf

$$Var(E(X|Y)) \cancel{+ E(Var(X|Y))}$$

$$= E\left((E(X|Y))^2\right) - \left(E(E(X|Y))\right)^2$$

$$E(Var(X|Y)) = E\left(E(X^2|Y) - (E(X|Y))^2\right)$$

$$= E(E(X^2|Y)) - E(E(X|Y)^2)$$

$$Var(E(X|Y)) + E(Var(X|Y)) = E(E(X^2|Y)) - (E(X))^2$$

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

$$\Rightarrow Var(E(X|Y)) + E(Var(X|Y)) = Var(X)$$

Thm

$$\rho_{X,Y} = \pm 1 \Rightarrow Y = aX + b$$

$$-1 \leq \rho_{X,Y} \leq 1 \Rightarrow Var\left(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}\right) \geq 0$$

$$\Rightarrow -\sigma_X \sigma_Y \leq Cov(X,Y) \leq \sigma_X \sigma_Y$$

$$\Rightarrow Var\left(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}\right) \geq 0$$

$y = x^2$; X^k

$E(X^4) = E(X) = E(X^3) = E(X^2)$

$E(X^3) - E(X)E(X^2)$

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

$= Var(X)$

$\sigma_X = \sqrt{Var(X)}$

$\sigma_Y = \sqrt{Var(X^2)}$

$= \sqrt{E(X^4) - (E(X^2))^2}$

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

Cor-1 ~~cov(X,Y) ≥ -σx σy ≤ 0~~

$\mathrm{cov}(X,Y) + \sigma_x \sigma_y \geq 0$.

$\Leftrightarrow \quad \mathrm{Var}\left(\frac{X}{\sigma_x} + \frac{Y}{\sigma_y}\right) \geq 0$

$\mu_x = E(X) \quad , \mu_y = E(Y)$.

$\Rightarrow \quad \rho_{x,y} = -1$

$\Rightarrow \quad \mathrm{Var}\left(\frac{X}{\sigma_x} + \frac{Y}{\sigma_y}\right) = 0$

$\Rightarrow \quad \frac{X}{\sigma_x} + \frac{Y}{\sigma_y} = 0$

$\Rightarrow \quad \frac{X}{\sigma_x} + \frac{Y}{\sigma_y} = \frac{E(X)}{\sigma_x} + \frac{E(Y)}{\sigma_y}$] ⟹ (Read more)

$\Rightarrow \quad \frac{Y}{\cancel{\sigma_y}} = \sigma_y\left(-\frac{X}{\sigma_x} + \frac{E(X)}{\sigma_x} + \frac{E(Y)}{\sigma_y}\right)$

$\Rightarrow \quad Y = AX + B$.

$A = -\frac{X\sigma_y}{\sigma_x} \quad , B = \frac{E(X)}{\sigma_x} + \frac{E(Y)}{\sigma_y}$

⟹ Y is indeed a linear func with prob-1.

$\mathrm{Var}(X) = E((X - E(X))^2) = \sum_{D_x} \underbrace{(x-\mu_x)^2}_{\geq 0} \underbrace{P(X=x)}_{\cancel{\geq} 0} = 0$

$(x-\mu_x)^2 \geq 0$ ~~almost~~ everywhere.

for discrete ~~dis~~ case. if $P(X=x) > 0$

$\mathrm{Var}(X) = \int (x-\mu_x)^2 f_X(x)\,dx = 0$

$\Rightarrow \quad (x-\mu_x)^2 \geq 0$

Thus $f_X(x)$ ~~dx~~ must be a zero function with measure 0 ⟹ ~~P(X=x) ≥ 0 almost~~

$\Rightarrow \quad (x-\mu_x)^2 = 0$ almost everywhere.

Def^n (Binomial r.v.).

A r.v. X is a binomial r.v. that is the sum of n iid Bernoulli(p) r.v.

$X \sim \mathrm{Binomial}(n,p) \quad , X = X_1 + X_2 + \dots + X_n$

Eg Toss a coin 20 times. $E(X) = \sum_i E(X_i) = \sum_i p = np$

X : The no of heads obtained.

$$P(X = r) = \begin{cases} 0 & r \in \{0, \dots, n\} \\ \sum_{\substack{S \subseteq \{1,\dots,n\} \\ |S| = r}} \prod_{i \in S} P(X_i = 1) \prod_{j \notin S} P(X_j = 0) \end{cases}$$

$P(X=r) = {}^nC_r \, p^r (1-p)^{n-r}$.

\# Suppose ~~on~~ avg you receive 3 msgs/hr. Find the probability that you receive that you receive at least 1 msg in the next 15 mins.

Ans: $N \sim Bin(60, p)$

$E(N) = 60p = 3$

$p = \frac{1}{20}$

Define $X \sim Bin(15, \frac{1}{20})$.

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

$P(X=0) = \binom{n}{0} (1-p)^n$

$= (1-p)^n = \left(1-\frac{1}{20}\right) = \left(\frac{19}{20}\right)^{15}$

$P(X>0) = 1 - \left(\frac{19}{20}\right)^{15}$

$= \frac{(20)^{15} - (19)^{15}}{(20)^{15}}$

Poisson r.v.

Def^n: The limiting case of a binomial dist where $n \to \infty$ but $E(X) = np = \lambda$ = remains constant,

$$P(X=r) = \begin{cases} 0 & \text{if } r \notin \{0\} \cup \mathbb{N} \\ \lim_{n\to\infty} \binom{n}{r} (p_n)^r (1-p_n)^{n-r} \end{cases} \quad X \sim \text{poisson}(\lambda)$$

$E(X) = \lambda = np_n$.

$$P(X=r) = \lim_{n\to\infty} \frac{n!}{(n-r)!\,r!} \left(\frac{\lambda}{n}\right)^r \left(1-\frac{\lambda}{n}\right)^{n-r}$$

$$= \lim_{n\to\infty} \frac{n!}{(n-r)!\,r!} \left(\lambda^r (n-\lambda)^{n-r}\right)$$

$$= \frac{\lambda^r}{r!} \lim_{n\to\infty} \frac{n!}{(n-r)!} (n-\lambda)^{n-r}$$

$$= \frac{\lambda^n}{r!} \lim_{n\to\infty} \overbrace{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{r-1}{n}\right)}^{\lim \to 1} \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^r}$$

$$= \frac{\lambda^n}{r!} \lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda} \frac{\lambda^n}{r!}$$

For a Poisson Dist, $\boxed{E(X) = Var(X) = \lambda}$

Geometric distribution.

Consider an r.v. which counts the no. of tosses required for obtaining the first head if you keep on tossing a coin.

Defn [Geometric]

$\{X_i\}_{i=1}^{\infty}$ i.i.d Bernoulli r.v.

We call X a geometric(p) if the following hold.

$\forall n \in \mathbb{N}$, $X = n$ if $X_n = 1$ & $X_i = 0$ $\forall i < n$

& $P(X=r) = 0 \;\forall r \notin \mathbb{N}$

$$P(X=r) = P(X_1 = 0, \ldots, \& X_r = 1)$$
$$= P(X_1=0) \ldots P(X_r = 1)$$
$$= (1-p)^{r-1} p$$

$$P(X=r) = \begin{cases} (1-p)^{r-1} p & r \in \mathbb{N} \\ 0 & r \notin \mathbb{N} \end{cases}$$

We now have the generalised version of Geometric.

Defn [Pascal dist]

$X = X_1 + \ldots + X_n$ where X_i's are i.i.d Geometric(p)

$\Rightarrow X \sim \text{Pascal}(n, p)$

PMF

$P(X=0) = 0 \quad \forall r \notin \{n, n+1, n+2, \ldots\}$

$$P(X=r) = \sum_{\substack{r_1+r_2+\ldots+r_n = r \\ r_i \in \mathbb{N}}} P(X_1 = r_1, \ldots, \& X_n = r_n)$$

$$= \sum_{\substack{r_1+r_2+\ldots r_n = r \\ r_i \in \mathbb{N}}} P(X_1 = r_1) \ldots P(X_n = r_n)$$

$$= \sum_{\substack{r_1+r_2+\ldots+r_n = r \\ r_i \in \mathbb{N}}} (1-p)^{r_1-1} (1-p)^{r_2-1} \ldots (1-p)^{r_n-1} p^n$$

$$= \sum_{\substack{r_1+r_2+\ldots r_n = r \\ r_i \in \mathbb{N}}} (1-p)^{r-n} p^n$$

$$\Rightarrow \binom{r-1}{n-1} (1-p)^{r-n} p^n$$

Better argument.

for ~~...~~ $(r-1)^{th}$ toss it is binomial distribution, so $P(X=r-1) = \binom{r-1}{n-1} (1-p)^{r-n} p^{n-1}$

$$P(X=r) = \binom{r-1}{n-1}(1-p)^{r-n}\, p^{n}.$$

Hypergeometric dist

N balls, m of them are blue,

N − m are white.

X: the no. of red balls obtained in a draw of n balls.

$$P(X=r) = 0 \quad \forall\, r \notin \{0, \dots, \min\{n, m\}\}$$

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

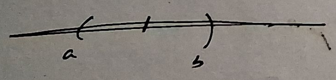
↓

Hypergeometric r.v. PMF

---

cts. uniform distr.

R.E.
choose a pt
arbitrarily
randomly from
an interval (a, b)



X: R.V. taking the value you've chosen

$$X \sim U(a, b)$$

If the PDF of a r.v. X is a non-zero constant over a ~~finite~~ bdd. interval and zero everywhere, then X is defined to be a cts uniform r.v.

$$\int_a^b f_X \, dx = 1 \qquad f_X(c) = c$$

$$\Rightarrow f_X \int_a^b dx = 1$$

$$\Rightarrow f_X(x) = \frac{1}{(b-a)}$$

→ PDF takes a constant value in

Normal random variable

CLT

Let X_i's be iid, the distribution.

$\overline{X_n} = \frac{x_1 + \cdots + x_n}{n}$ converges to a normal distribution

Def^n (Normal r.v.)

Let X be a r.v. whose PDF is given by

$$f_X(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma}}$$

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

We assume now $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

Lemma-1 If $X \sim N(\mu, \sigma^2)$, $Y = aX + b$, $a \neq 0$

then $Y \sim N(a\mu + b, a^2\sigma^2)$

Pf

$$F_Y(t) = P(Y \leq t)$$

$$= P(aX + b \leq t)$$

$$= \begin{cases} P\left(X \leq \frac{t-b}{a}\right) & a > 0 \\ P\left(X \geq \frac{t-b}{a}\right) & a < 0 \end{cases}$$

Since X is continuous r.v.,

$$P(X \geq x) = P(X > x)$$

$$\Rightarrow \begin{cases} F_X\left(\frac{t-b}{a}\right), & a > 0 \\ 1 - F_X\left(\frac{t-b}{a}\right), & a < 0 \end{cases}$$

Since $f_X(x)$ is continuous, ~~then~~ we have

$$F_X'(x) = f_X(x)$$

Since $F_X'(x)$ exists, $F_Y'(y)$ exists, we may take $F_Y'(y)$ to be a pdf of Y.

$$F_Y'(t) = \frac{1}{|a|} F_X'\left(\frac{t-b}{a}\right)$$

$$= \frac{1}{|a|\sigma\sqrt{2\pi}} e^{-\frac{(t-b-a\mu)^2}{2\sigma^2 a^2}}$$

$$= \frac{1}{|a|\sigma\sqrt{2\pi}} e^{-\frac{(t-(a\mu+b))^2}{2\sigma^2 a^2}}$$

$$\Rightarrow Y \sim N(a\mu + b, a^2\sigma^2)$$

Corollary $Z \sim N(\mu, \sigma^2)$

$$Z \mapsto \frac{Z-\mu}{\sigma}$$

$$\Rightarrow Z \sim N(\mu, \sigma^2) \longmapsto \frac{Z-\mu}{\sigma} \sim N(0, 1)$$

Defn [Standard Normal Dist]

$X \sim N(0,1)$

$$f_X(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

Lemma

Let $\phi(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ then the CDF of the std normal r.v.

(i) $\lim_{x \to \infty} \phi(x) = 1$ (ii) $\phi(-x) = 1 - \phi(x)$

Pf $\phi(-x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-x} e^{-\frac{t^2}{2}} dt$ $t' = -t$

$dt' = -dt$

$$= \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-\frac{t^2}{2}} dt$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

$$= 1 - \phi(x)$$

Lemma 03 Let $X \sim N(0,1)$

Then $E(X^n) = \begin{cases} 0 & \text{if } n \text{ is odd} \\ (n-1)!! & \text{if } n \text{ is even} \end{cases}$

$(n-1)!! = (n-1)(n-3) \cdots 5 \cdot 3 \cdot 1$

Pf Let n be odd,

we have $E(X^n) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^n e^{-t^2/2} dt$

$= 0$

Since $t^n e^{-t^2/2}$ is an odd function.

For n is even.

$$E(X^n) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^n e^{-t^2/2} dt$$

$$= \frac{1}{\sqrt{2\pi}} \cdot t^{n-1} e^{-t^2/2} \Big|_{-\infty}^{\infty}$$

$$+ \frac{1}{\sqrt{2\pi}} (n-1) \int_{-\infty}^{\infty} t^{n-2} e^{-t^2/2}$$

$$= + \cancel{\frac{1}{\sqrt{2\pi}}} (n-1) E(X^{n-2})$$

Continuing this recurrence relⁿ, we have,

$E(X^n) = (n-1) \ldots 5.3.\, E(X^2) = (n-1)!!$

<u>Cor.</u> $X \sim N(0,1)$

$E(X) = 0 \qquad Var[X] = 1.$

<u>Cor-2</u>

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

$$E\left(\left(\frac{X-\mu}{\sigma}\right)^n\right) = \begin{cases} 0 & n \text{ is odd} \\ (n-1)!! & n \text{ is even.} \end{cases}$$

$$\Rightarrow \underbrace{E\left((X-\mu)^n\right)}_{n^{th} \text{ central moment.}} = \begin{cases} 0 & n \text{ is odd} \\ (n-1)!!\,\sigma^2 & n \text{ is even} \end{cases}$$

$\Rightarrow E(X) = \mu \;;\; Var(X) = \sigma^2$

(2) <u>Claim</u> $\lim_{x\to\infty} \phi(x) = \lim_{x\to\infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$

exists

Observe $e^{-t^2/2} \in (0,1] \quad \forall t \in \mathbb{R}$

If we have $t \geq 2$ then $e^{-t^2/2} \leq e^{-t}$

$t \geq 2$

$t^2 \geq 2t \Rightarrow -t^2 \leq -2t. \Rightarrow -\frac{t^2}{2} \leq -t.$

$\Rightarrow e^{-t^2/2} \leq e^{-t}.$

Let us consider

$a_n = \int_{-n}^{n} e^{-t^2/2}\, dt$

(a_n) is an increasing sequence.

Cor-2

$$E(X^2) = \mu^2 + \sigma^2$$

Cor-3

$$E(X^3) = \mu^3 + 3\mu\sigma^2$$

$$E((X-\mu)^3) = 0$$

$$\Leftrightarrow E(X^3) = \mu^3 + 3\mu\sigma^2$$

So in gen if you know

$E(X^j)$ & $\forall j \in \{1, \dots, n-1\}$

then expanding $E((X-\mu)^n)$, you may obtain $E(X^n)$

Recall

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

$P(|X-\mu| \leq \sigma) \simeq 0.68$

$P(|X-\mu| \leq 2\sigma) \simeq 0.95$

$P(|X-\mu| \leq 3\sigma) \simeq 0.997$

Markov's Inequality

Let X be a r.v. and $a \in \mathbb{R}_{>0}$

$$P(|X| \geq a) \leq \frac{E(|X|)}{a}$$

Pf

$p = P(|X| \geq a)$ and define $\hat{X} = \begin{cases} 1 & |X| \geq a \\ 0 & \text{otherwise} \end{cases}$

$\hat{X} \sim \text{Bernoulli}(p)$

$$\int_a^b f + \int_{f(a)}^{f(b)} f^{-1} = bf(b) - af(a)$$

$$\int_{f(a)}^{f(b)} f^{-1}(x)\,dx$$

Let

$f(u) = x$

$f'(u)\,du = dx$

$x f'(u)$

$x f(u)\Big|_a^b - \int_a^b f(u)\,du$

$E(\hat{x}) = p$

$\hat{x} \leq \frac{|x|}{a}$ $\quad |x| \geq a$

$|x| \geq a, \ \hat{x} = 1$

$E(\hat{x}) \leq E\left(\frac{|x|}{a}\right)$ $\quad |x| \not\geq a$

$\& \ \hat{x} = 0$

$\Rightarrow P(|x| \geq a) \leq E\left(\frac{|x|}{a}\right)$ $\quad \Rightarrow |x| \geq 0$

Chebyshev's Inequality

If Y is a r.v. with $E(Y) = \mu$ and $Var(Y) = \sigma^2$ and if $b > 0$, then

$$P(|Y - \mu| \geq b) \leq \frac{\sigma^2}{b^2}$$

Pf Let $x := (Y-\mu)^2$ and $a = b^2$

$P(|x| \geq b^2) = P(|Y - \mu| \geq b)$

$\leq \frac{E((Y-\mu)^2)}{b^2} = \frac{\sigma^2}{b^2}$

$\Rightarrow P(|Y - \mu| \geq b) \leq \frac{\sigma^2}{b^2}$

general $t\sigma$ inequality

If X is a r.v. with $E(X) = \mu$ & Var $\sigma^2 > 0$ $\quad \forall t > 0$

$$P(|X - \mu| < t\sigma) \geq 1 - \frac{1}{t^2}$$

Pf Use chebyshev.

$Y = X$.

$b^2 = t\sigma$

$$P(|X - \mu| \geq t\sigma) \leq \frac{1}{t^2}$$

$\Rightarrow -P(|X - \mu| \geq t\sigma) \geq -\frac{1}{t^2}$

$\Rightarrow P(|X - \mu| < t\sigma) \geq 1 - \frac{1}{t^2}$

Weak Law of large numbers

Population: All the values taken by a r.v.

Sample: A finite set of values assumed by a r.v.

random sample: A set of n iid r.v. $X_1, \ldots, X_n$ with population mean μ and var σ^2.

Sample mean $\bar{X}_n = \frac{X_1 + \ldots + X_n}{n}$

$E(\bar{X}_n) = \frac{n\mu}{n} = \mu$ $\quad Var(\bar{X}_n) = \frac{\sigma^2}{n}$

Thm

Let $X_1, \ldots, X_n$ be a r. sample from a population with mean μ and var σ^2

then $\forall \varepsilon > 0$, we have.

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon}$$

Pf

Put $Y = \bar{X}_n$, $b = \varepsilon$ $\quad Var = \frac{\sigma^2}{n}$

Markov Chains

Discrete time finite state space: Markov chain with stationary transition probabilities

Markov Property and Markov Chains

Defⁿ 1 Let $\{X_n\}_{n=1}^{\infty}$ be a seqⁿ of r.v. taking values in $\{0, \ldots, N\} \rightarrow$ states.

Markov property

$$P(X_{n+1} = s_{n+1} / X_n = s_n, X_{n-1} = s_{n-1}, \ldots, X_0 = s_0)$$

$$= P(X_{n+1} = s_{n+1} / X_n = s_n)$$

If $\{X_n\}_{n=0}^{\infty}$ satisfy the Markov Chain Property, we call $\{X_n\}_{n=0}^{\infty}$ a M.C. and if $X_n = s$ for some n, then we say that the system is <u>at state s</u> <u>at time n</u>.

Transitional probabilities

$$P(X_{n+1} = j / X_n = i)$$

If this probabilities are independent of n, then these are called stationary transitional probabilities

$$P_{ij} = P(X_{n+1} = j / X_n = i)$$

$$P_{ij} = P(X_1 = j / X_0 = i)$$

Transitional prob matrix / Stochastic matrix

$$\mathbb{P} = (P_{ij})_{i,j=0}^{N}$$

Lemma-1 If $\mathbb{P}$ is a stochastic mx, then each row of $\mathbb{P}$ sums to 1.

Pf $\sum_{j=0}^{N} P_{ij} = \sum_{j=0}^{N} P(X_1 = j / X_0 = i)$

Joint prob distribution.

$P(X_m = s_m, X_{m-1} = s_{m-1}, \dots, X_0 = s_0)$

$= P_{s_{m-1}}$

$P(X_1 = j \mid X_1 = 0)$

$P(X_1 = 0 \mid X_0 = i)$

$P_{ij}^{(n)}$: n^{th} state transition probability matrix.

$P_{ij}^{(n)} = P(X_n = j \mid X_0 = i)$

Thm If $\mathbb{P}^{(n)}$ is the n^{th} state Tr pr mx. of a M.C., then

$$P(X_n = j) = \sum_{i=0}^{N} P_{ij}^{(n)} P(X_0 = i)$$

Pf

$$P(X_n = j) = \sum_{i=0}^{N} P(X_n = j / X_0 = i)$$

$$= \sum_{i=0}^{N} P(X_n = j / X_0 = i) P(X_0 = i)$$

$$= \sum_{i=0}^{N} P_{ij}^{(n)} P(X_0 = i)$$

$\blacksquare$

Chapman Kolmogorov Eqn.

$$\mathbb{P}^{(n)} = \mathbb{P}^n$$

Pf $\mathbb{P}^{(1)} = \mathbb{P}$ Let's assume the claim holds for $m \in \mathbb{N}$,

$$P_{ij}^{(m+1)} = P(X_{m+1} = j / X_0 = i)$$

$$= \sum_{k=0}^{N} P(X_{m+1} = j, X_m = k / X_0 = i)$$

$$= \sum_{k=0}^{N} P(X_{m+1} = j / X_m = k, X_0 = i) \; P(X_m = k / X_0 = i)$$

$$= \sum_{k=0}^{N} P(X_{m+1} = j / X_m = k) \; P_{ik}(m)$$

$$= \sum_{k=0}^{N} P_{ik}^{(m)} P_{kj} = P_{ij}^{(m+1)}$$

$$\Rightarrow \mathbb{P}^{(m+1)} = \mathbb{P}^{(m)} \mathbb{P}$$

We know by induction hypothesis

$$\mathbb{P}^{(m)} = \mathbb{P}^{m}$$

$$\Rightarrow \mathbb{P}^{(m+1)} = \mathbb{P}^{m+1}$$

Corrollary

$$\mathbb{P}^{n} = \mathbb{P}^{n-r} \mathbb{P}^{r} \quad \forall r \in \{0, 1, \ldots, n\}.$$

$$\mathbb{P}^{(0)} = \mathbb{I}$$

$$P_{ij}^{(0)} = P(X_0 = j / X_0 = i) = \delta_{ij}$$

$$\mathbb{P}^2 v = \mathbb{P}(\mathbb{P} v) = \mathbb{P} v = v \quad \Rightarrow \mathbb{P}^2 v = v$$

Accessible states

$P(X_n = j / X_0 = i) > 0$ for some $n \in \mathbb{N} \cup \{0\}$

We say that the state "j" is accessible from state "i" and we denote it by "$i \to j$"

Communication

Two states i, j are communicating if $i \to j$ and $j \to i$. We denote it by $i \leftrightarrow j$.

Now communication is an equivalence reln, we can partition it into equivalence classes which are communicating.

If a M.C. has only one communicating class, it is an irreducible MC.

Probability of revisiting a state

$f_s = P(X_n = s$ for some $n \in \mathbb{N} / X_0 = s)$.

So the probability of visiting a state again is $1 - f_s$.

$$f_s = P\left(\bigcup_{n=1}^{\infty} \{X_n = s\} / X_0 = s\right)$$

Recurrent state: A state s s.t. $f_s = 1$

Transient state: A " s s.t. $f_s < 1$.

Absorbing state : A state s s.t. $P(X_n = s \mid X_0 = s) = 1 \; \forall n \in \mathbb{N}$

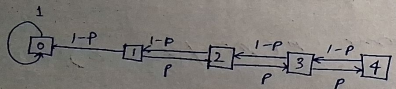
Absorbing $\Rightarrow$ recurring
recurring $\not\Rightarrow$ absorbing.

} If we have multiple communicating classes, then some states would become irrelevant later.
Something to read about.

Gambler's ruin

Gambler starts with gambling with an amt. ₹M and at each round of round of gamble, he wins ₹1 with prob p and loses ₹1 with prob ~~q~~ $1-p$

Unless $p < 1/2$, the casino may go bankrupt with a positive prob.



So the markov chain will ultimately reach 0 if a target amount T is not set. So the MC will reach 0 faster that it reaches infinity.
The choice of "T" could be an interesting thing!

X_n = Gambler's balance at time n.

Then n^{th} gamble $Y_n = \{$

$P(X_n \mid X_m)$

$P(X_{n+1} = +1 \mid X_n = -1)$

$= P($~~X_{n+1}~~ $=$

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n} \qquad E(\bar{X}_n) = \mu = \text{population mean}$$

$$Var(\bar{X}_n) = \frac{\sigma^2}{n}$$

CLT $\Rightarrow$ weak law of large numbers

Moment Generating function

The MGF of a r.v. X, if $E(e^{|tx|})$ exists in a neighbourhood of 0, then we define the MGF of X as

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

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

The n^{th} derivative gives us the n^{th} moment,

$$M_X^{(n)}\big|_{t=0} = E(X^n)$$

Prop 1 If X & Y are ~~iid~~ independent r.v, then

$$M_{X+Y} = M_X M_Y$$

Pf

$$M_{X+Y}(t) = E\left(e^{t(X+Y)}\right) = E(e^{tX})\, E(e^{tY})$$

$$= M_X M_Y$$

Prop 2

If X and Y are r.v, taking values in $\{0, \dots, n\}$ s.t. $M_X = M_Y$ in some neighbourhood of 0, then the PMFs of X and Y are the same.

Pf $M_X = M_Y$

$$\Rightarrow E(e^{tX}) = E(e^{tY})$$

$$\Rightarrow \sum_{x=0}^{n} e^{tx} P(X=x) = \sum_{y=0}^{n} e^{ty} P(Y=y)$$

$$\Rightarrow \sum_{r=0}^{n} e^{tr} \left(P(X=r) - P(Y=r)\right) = 0$$

Let $z = e^t$

$$\Rightarrow \sum_{r=0}^{n} \left(P(X=r) - P(Y=r)\right) z^r = 0$$

Now by defn it has uncountably many 0s in some neighbourhood of 0. However a polynomial can only attain 0 n times unless ~~all are not~~ they are identically 0 ...

Prop:

X, Y r.v. assuming values in $\{s_0, \dots, s_n\}$ s.t. $M_X = M_Y$ in some neighbourhood of 0. Then the PMFs of X and Y are the same.

Pf

$$M_X(t) = M_Y(t)$$

$$\Leftrightarrow \sum_{r=0} \left(P(X=s_r) - P(Y=s_r)\right) e^{ts_r} = 0$$

Now e^{ts_r} is are linearly independent, as they are in distinct eigenspaces of $\frac{d}{dx}$. Thus they are independent.

$$\Rightarrow P(X=s_r) = P(Y=s_r)$$

[Prove]

Thm (Uniqueness of MGF). If the MGF of a r.v. exists, then it determines the distribution of X uniquely.

Idea of Pf

cts case. MGF of a cts r.v. is $\int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx$ is the two sided Laplace transform of the PDF of f_X.

$\Rightarrow$ Taking the inverse Laplace transform, we will get back PDF

[W Feller, KL Chung]

$M_X(t) = \sum_{n=0}^{\infty} \frac{E(X^n)t^n}{n!}$ defines a complex analytic fn in a nbhd. So if two such functions match in some nbhd of 0, then, we obtain an analytic function, which contains uncountably many points. which is a contradiction as analytic functions have countably many zeroes.

MGF of standard normal distribution $= e^{t^2/2}$

MGF of Bernoulli r.v.

$X \sim$ Bernoulli (p)

$M_X(t) = E(e^{tX}) = e^{tx} P(X=1) + e^{t0} P(X=0)$

$= pe^t + 1 - p$

MGF of Bin(n,p)

$X \sim$ Bin(n,p) $\quad X = X_1 + \dots + X_n$

$\Rightarrow X_1, \dots, X_n$ are iid r.v.s and $X_i \sim Ber(p)$ $\forall i$

$M_X(t) = (pe^t + 1 - p)^n$.

MGF of Poisson

$X \sim$ Poisson(λ)

$M_X(t) = E(e^{tX}) = \sum_{j=0}^{\infty} e^{tj} e^{-\lambda} \frac{\lambda^j}{j!} = e^{\lambda(e^t-1)}$

Suppose $X = X_1 + \dots + X_n$ are independent poisson r.v. with parameters $\lambda_1, \dots, \lambda_n$.

$M_{X_1+\dots+X_n}(t) = e^{(e^t-1)(\lambda_1+\dots+\lambda_2)}$

If $X = X_1 + \dots + X_n$ where X_is $\sim$ Poisson(λ_i), we have by uniqueness of MGF,

$X \sim$ Poisson$(\lambda_1 + \dots + \lambda_n)$

MGF of exponential dist

$X \sim Exp(\lambda)$

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

$= \lambda \int_0^{\infty} e^{-(\lambda - t)x} dx$

Only converges iff $\lambda > t$

$= \frac{\lambda}{\lambda - t} \ \forall t < \lambda$

Chernoff bounds.

X is an MGF with M_X. Then for $a \in \mathbb{R}$, we have

$P(X \geq a) \leq \inf_{t \geq 0} \frac{M_X(t)}{e^{at}}$

$P(X \leq a) \leq \inf_{t \leq 0} \frac{M_X(t)}{e^{at}}$

Pf (case : $t \geq 0$)

Since e^x is increasing f^n, we have

$$P(X \geq a) = P(e^{tX} \geq e^{ta}) \leq \frac{E(e^{tX})}{e^{ta}} \quad \text{[By Markov]}$$

$$\Rightarrow P(X \geq a) \leq \frac{M_X(t)}{e^{ta}}$$

$$\Rightarrow P(X \geq a) \leq \inf_{t \geq 0} \frac{M_X(t)}{e^{ta}}$$

Similarly let us take

$t \leq 0$

$$P(X \leq a) = P(e^{tX} \geq e^{ta}) \leq \frac{E(e^{tX})}{e^{ta}}$$

$$\Rightarrow P(X \leq a) \leq \inf_{t \leq 0} \frac{E(e^{tX})}{e^{ta}}$$

__Thm__ (Levy - Lindeberg CLT)

Let $X_1, \dots, X_n$ be iid r.v. with $E(X_i) = \mu$ and variance $\sigma^2 > 0$

Define $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$.

Then the distribution of $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ converges to the std norm dist as $n \to \infty$.

$$\lim_{n \to \infty} \left(P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq a \right) \right) = \Phi(a)$$

__Lemma__

Converges in MGF $\Rightarrow$ Convergence in CDF at all points where the CDF is continuous.

Pf. $Y_i = \frac{X_i - \mu}{\sigma/\sqrt{n}}$

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\frac{1}{n}\sum X_i - \mu}{\sigma/\sqrt{n}}$$

$$= \frac{\sum X_i - n\mu}{\sqrt{n}\,\sigma} = \frac{\sum Y_i}{\sqrt{n}}$$

Let $M(t)$ defined as $M_{Y_i}(t)$

$\Rightarrow$ The MGF of $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ is $M\left(\frac{t}{\sqrt{n}}\right)^n$

We define $L(t) = \log(M(t))$

$L(t) = \log M(t)$

To show

$$\lim_{n \to \infty} nL\left(\frac{t}{\sqrt{n}}\right) = \frac{t^2}{2}$$

$M(t) = M_{Y_i}(t) = E(e^{t^2 \frac{y_i}{2}})$

$M(0) = E(e^0) = 1$

$M'(0) = 0$

$M''(0) = 1$

$L(0) = \log E(M(0))$, $L'(0) = 0$, $L''(0) = 1$

Let $x = \frac{1}{n}$

$$\lim_{x \to 0} \frac{L(t\sqrt{x})}{\sqrt{x}} \quad [\text{L'Hospital}]$$

$$= \lim_{x \to 0} \frac{L'(t\sqrt{x})t^{2}}{\sqrt{x}} = \lim_{x \to 0} \frac{L''(t\sqrt{x})\, t^2}{2} = \frac{t^2}{2}$$

Order statistics

Let $X_1, \dots, X_n$ be iid r.v.

$X_{(1)} = \min\{X_1, \dots, X_n\}$

$X_{(j)} =$ j^{th} smallest minima

$X_{(n)} = \max\{X_1, \dots, X_n\}$

Prop^n Let F be the common CDF of $X_1, \dots, X_n$ (iid rv)

Then the CDF of $X_{(j)}$

$$F_{X_{(j)}} = \sum_{r=j}^{n} \binom{n}{r} F(x)^r (1 - F(x))^{n-r}$$

Pf S, S' are disjoint subsets of $\{1, \dots, n\}$ of cardinality of m and m', then

$P(X_s \leq x \; \forall s \in S, \; X_{s'} > x \; \forall s' \in m')$

$\Rightarrow \left(\prod_{s \in S} P(X_s \leq x)\right)\left(\prod_{s \in S'} P(X_{s'} > x)\right)$

$= [F(x)]^m [1 - F(x)]^{m'}$

$F_{X_{(j)}}(x) = P(X_{(j)} \leq x)$

$= P($ Atleast j of the r.v.s have value less than $x)$

$= \sum_{r=j}^{n} P($Exactly r of the r.v. are less than or equal to $x)$

$$= \sum_{r=j}^{n} \binom{n}{r} F(x)^r [1 - F(x)]^{n-r}$$

Cor $F_{X_{(1)}} = 1 - (1 - F(x))^n$

$F_{X_{(n)}}(x) = (F(x))^n$

Cor If $X_1, \dots, X_n$ are iid discrete r.v.s, with CDF $= F$ then the PMF of their j^{th} order statistic given by $P(X_{(j)} = t) = F_{X_{(j)}}(t) - \lim_{s \to t^-} F_{X_{(j)}}(s)$

$F(x) = \int$

Coro

$$P(X_{(1)} = t) = 1 - (1 - F(t))^n - \lim_{s \to t^-} (1 - (1 - F(s)^n))$$

$$= \underbrace{\lim_{s \to t^-} (1 - F(s))^n - (1 - F(t))^n}$$

$$P(X_{(n)} = t) = F(t)^n - \lim_{s \to t^-} (F(s))^n$$

- - - - - - - - - - - - - - -

Prop Let $x_1, \ldots, x_n$ be iid cts r.v. with a common differentiable CDF F and a common PDF $f = F'$. Then the PDF of $X_{(j)}$ is given by

$$f_{(j)}(t) = n \binom{n-1}{j-1} F(t)^{j-1} (1 - F(t)^{n-j}) f(t)$$

Proof

$$f_{x_j}(t) = F'_{x_j}(t)$$

we can do this since $F(t)$ is cont. differentiable $\Rightarrow$

$$F_{X_{(j)}}(t) = \sum_{r=j}^{n} \binom{n}{r} F(t)^r (1 - F(t))^{n-r}.$$

is also cts differentiable.

Let us differentiate $F_{x_j}(t)$

$$= \sum_{r=j}^{n} \binom{n}{r} r (F(t))^{r-1} (1 - F(t)) f(t)$$

$$- \sum_{r=j}^{n} \binom{n}{r} (n-r) F(t)^r (1 - F(t))^{n-r-1} f(t)$$

Now we have

$$f_{X_{(j)}}(t) = n f(t) \Big(\sum_{r=j}^{n} \binom{n-1}{r-1} F(t)^{r-1} (1 - F(t))^{n-r}$$

$$- \sum_{r=j}^{n} \binom{n-1}{r} F(t)^r (1 - F(t))^{n-r-1} \Big)$$

$$= n f(t) \Big(\sum_{r=j}^{n} \binom{n-1}{r-1} F(t)^{r-1} (1 - F(t))^{n-r}$$

$$- \sum_{r=j+1}^{n} \binom{n-1}{r-1} F(t)^{r-1} (1 - F(t))^{n-r} \Big)$$

$$= n f(t) \binom{n-1}{j-1} F(t)^{j-1} (1 - F(t))^{n-j}$$

$$\Rightarrow f_{X_{(j)}}(t) = n f(t) \binom{n-1}{j-1} F(t)^{j-1} (1 - F(t))^{n-j}$$

1st Order. stat

$$\cancel{f_{X_{(1)}}} \; f_{X_{(1)}}(t) = n f(t) (1 - F(t))^{n-1}$$

$$f_{X_{(n)}}(t) = n f(t) F(t)^{n-1}$$

Joint PMF of order statistics

Let $X_1, \ldots, X_n$ be iid r.v. with a common PMF f. Then the joint PMF is given by

$$P(X_{(1)} = x_1, \ldots, X_{(n)} = x_n)$$

$$= \begin{cases} 0 & \text{unless } x_1 \le \ldots \le x_n \\ \dfrac{n!}{r_1! \ldots r_n!} f(t_1)^{r_1} \ldots f(t_m)^{r_m} & \text{where} \end{cases}$$

x_i's assume the values t_i and r_i's are the no. of times t_i's appear. Here t_i's are for distinct r.v.s.

Pf

Let $\tilde{x} := (x_1, \ldots, x_n)$ and let $G_{\tilde{x}}$ denote the group of permutations of $\tilde{x}$.

Let t_i's be the values assumed by the x_i's and let the values be repeated r_i times.

Now we have

$$X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$$

unless $x_1 \le x_2 \le \ldots \le x_n$, we have

$$P(X_{(1)} = x_1, X_{(2)} = x_2, \ldots, X_{(n)} = x_n) = 0$$

Since $X_{(i)}$'s are the ith order statistics, thus the $X_{(i)}$'s are ordered, thus the values they assume also must be ordered. Suppose if we have 2 elements, then max $\le$ min is a false rel^n. Thus $P(\max \le \min) = 0$

Note

$\tilde{x} = (x_1, \ldots, x_n)$.

Let g be a permutation map.

$$(x_1, \ldots, x_n) \mapsto (g(x_1), \ldots, g(x_n))$$

Here g swaps two elements of the tuple.

i.e. $g(x_j) = x_i$ and $g(x_i) = x_j$

Now let us assume

$\tilde{x} = (x_1, \ldots, x_n)$ s.t.

$$x_1 \le x_2 \le \ldots \le x_n$$

$|G_{\tilde{x}}|$ = no. of permutations of the tuple

$$= \frac{n!}{r_1! \ldots r_m!}$$

$P(X_{(1)} = x_1, X_{(2)} = x_2, \ldots, X_{(n)} = x_n)$

= the probability that the values $x_1, \ldots, x_n$ are assumed by the initial rv's $X_1, \ldots, X_n$ in some order.

$$P(X_{(1)}=x_1, \ldots, X_{(n)}=x_n)$$

$$= \sum_{g\in G_x} P(X_1=g(x_1), \ldots, X_n=g(x_n))$$

$$= \sum_{g\in G_x} f(x_1) f(x_2) \cdots f(x_n)$$

$$= \frac{n!}{r_1!\, r_2! \cdots r_m!} [f(x_1)] \cdots [f(x_n)]$$

$$= \frac{n!}{r_1!\, r_2! \cdots r_m!} (f(t_1))^{r_1} \cdots (f(t_m))^{r_m}$$

[Joint PDF]

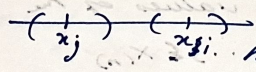
Thm

Let $X_1, \ldots, X_n$ be iid cts r.v. with a common cts PDF f. Then the joint PDF of $X_{(1)}, \ldots, X_{(n)}$ is given by

$$\begin{cases} n!\, f(x_1) \cdots f(x_n) & \text{if } x_1 \le x_2 \le \ldots \le x_n \\ 0 & \text{otherwise.} \end{cases}$$

Pf Let $\underline{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$

if $i, j \in \{1, 2, \ldots, n\}$ are s.t. $i<j$ & $x_i > x_j$ for sufficiently small δ, we have



we have disjoint neighbourhoods.

$$\delta < \frac{|x_j - x_i|}{2}.$$

$$(x_i-\delta, x_i+\delta) \cap (x_j-\delta, x_j+\delta) = \phi$$

$$P(x_i-\delta \le X_{(i)} \le x_i+\delta,\ x_j-\delta \le X_{(j)} \le x_j+\delta) = 0$$

as $X_{(j)} \ge X_{(i)}$

But this is not possible as $X_{(i)}$ takes values strictly greater than $X_{(j)}$.

Now

$P(x_1-\delta \le X_{(1)} \le x_1+\delta, \ldots, x_n-\delta \le X_{(n)} \le x_n+\delta)$

is a subevent of

$P(x_i-\delta \le X_{(i)} \le x_i+\delta,\ x_j-\delta \le X_{(j)} \le x_j+\delta)$

Since

P(~~happens~~ Event E) = 0

⇒ P(subevent S) = 0

⇒ If S happens ⇒ E happens

Thus we define the values of the joint PDF of $X_{(1)}, \dots, X_{(n)}$

at a pt $\tilde{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$

where for some $i < j$

$x_i > x_j$

$\Rightarrow P(E) = 0$

Now lets assume $\tilde{x} = (x_1, \dots, x_n)$ be s.t.

$x_1 \le x_2 \le \dots \le x_n$

Let $f_{\tilde{x}}$ be the joint PDF of $x_1, \dots, x_n$

Therefor any permutation $(i_1, \dots, i_n)$ of $(1, 2, \dots, n)$, we have

$$f_{\tilde{x}}(x_{i_1}, \dots, x_{i_n}) = f(x_1) \cdots f(x_n)$$

For sufficiently small $\varepsilon > 0$,

$$P\left(x_{i_1} - \frac{\varepsilon}{2} \le X_1 \le x_{i_1} + \frac{\varepsilon}{2}, \dots, x_{i_n} - \frac{\varepsilon}{2} \le X_n \le x_{i_n} + \frac{\varepsilon}{2}\right)$$

$$= \prod_{j=1}^{n} P\left(x_{i_j} - \frac{\varepsilon}{2} \le X_j \le x_{i_j} + \frac{\varepsilon}{2}\right)$$

$$\approx \prod_{j=1}^{n} \varepsilon f(x_{i_j}) = \varepsilon^n f(x_1) \cdots f(x_n)$$

Now if

$x_1 < x_2 < \dots < x_n$, then

for sufficiently small $\varepsilon > 0$, we have

$$P\left(x_1 - \frac{\varepsilon}{2} \le X_{(1)} \le x_1 + \frac{\varepsilon}{2}, \dots, x_n - \frac{\varepsilon}{2} \le X_{(n)} \le x_n + \frac{\varepsilon}{2}\right)$$

$$= n!\, \varepsilon^n f(x_1) \cdots f(x_n)$$

Since we sum over all permutations of $(1, 2, \dots, n)$ and there are $n!$ such permutations. Since they is in strict inequality.

Hence we define the value of the joint PDF of $X_{(1)}, \dots, X_{(n)}$ at $\tilde{x} = (x_1, \dots, x_n)$ with $x_1 < \dots < x_n$ by dividing both sides ε^n and taking the limit $\varepsilon \to 0$

$$g(x_{(1)}, \dots, x_{(n)}) = n! f(x_1) \cdots f(x_n)$$

Now for $\tilde{x} = (x_1, \dots, x_n)$ with $x_1 \le x_2 \le x_3 \dots$ for $x_1 \le \dots \le x_n$ ~~by the limit~~ we get joint PDF by the limit

$$\lim_{(r_1, \dots, r_n) \to \tilde{x}} n! f(r_1) \cdots f(r_n) = n! f(x_1) \cdots f(x_n)$$

$i_j < i_j$, $x_{i_j} \Leftrightarrow x_i$ $v_i + v_j = v+1$

So the joint PDF r.v.s is given by

$$\begin{cases} n!\, f(x_1) \cdots f(x_n) & \text{if } x_1 \le x_2 \cdots \le x_n \\ 0 & \text{o/w} \end{cases}$$

!! continuity reqd.
