PROBABILITY THEOREM

GEORGE SALMAN

- Some Chosen problems
- The Probability Module Axioms
- Conditional Probabilty
- Bernulli-Trials
- Discrete-Random-Variables
- Absolutely-Continuous-Random-Variable
- Tail-Formula
- Borel-Sets
- Random Variables & The Connection with Borel Sets
- Variance
- Random vectors
- Common Dirstribute Funtions
- Function Of Random Vector
- Sum of independent Random Variables
- Convolution
- Conditional Probaility Funcitons
- Covariance
- Gaussian Vectors
- Chebyshev's and markov inequality
- Sequence of infinite events
- Law of Large numbers
- Central Limit Theorem
- Challenging Questions

Some Chosen problems

Problem. giving permutations $\{1, 2, ..., n\}$,$n \in \mathbb{N}$.

what is the probabilty that in random permutation we will get fixed point ?

Solution. Denote $A = \{$set of permutation in which 1 is a fixed point $\}$
we are looking for $\mathbb{P}(A)$.

$$\mathbb{P}(A) = \frac{|A|}{|\{S_n : [n] \to [n]\}|} = \frac{(n-1)!}{n!} = \frac{1}{n}$$

Exercise. what is the probability that in a random permutation obtain (1,2) as transposition? i.e (1 mapped to 2 vice versa).

Solution. denote B the event in which we are choose all the permutation s.t $1 \to 2, 2 \to 1$ so we have $(n-2)!$

Since for 1 we have 1 option and for 2 1 option and for the 3 we have $n-3$ and 4 $n-4$ So :

$$1 \cdot 1 \cdot (n-2) \cdot (n-3) \cdot .... \cdot 1 = (n-2)!$$

hence ,

$$\mathbb{P}(B) = \frac{|B|}{|\{S_n : [n] \to [n]\}|} = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}.$$

Exercise. In a jug there is 5 balls- two are black and 3 are white,

we take out 2 balls with return, what is the probability that the second ball is black ?

Solution. denote $\Omega = \{(1,2),(1,3),(2,4), ...(5,4)\}$, then $|\Omega| = 20$

$$\mathbb{P}(C) = \frac{|C|}{|\Omega|} = \frac{8}{20} = \frac{2}{5}$$

.

Exercise. In a set of n people what is the probabilty the exsist two people with the same birthday?

if $n \geq$366 then its 1. using peigenhole principle.

Denote $A =${event in which exsist at least two people with the same birthday}.

It's more comfortable to find $\mathbb{P}(A^c) = P$ (all have different birthday date),

we have 365^noptions to choose birtday date for n people.we count the options of choosing n different birthday date we have

$$(365) \cdot ... \cdot (366-n) \text{ So}, \mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \frac{(365)....(366-n)}{365^n}.$$

(*) its true for $\forall n$ birthday has the same probability

probability	number of people (n)
~25%	14
~51%	23
~71%	30
~90%	40

Probability Module

Definition. The Axiomatic Probability Module (Ω, F, P)

$\Omega-$ Sample space (all possible results).

$F-$Collection of Subsets of Ω, if $A \in F$, A is called even.

$P : F \to [0,1]$ is a probability function $\mathbb{P}(A)$ =probability of A, $A \in F$.

<u>Axioms of F:</u>

$F1$ **::** $\emptyset \in F$

$F2$ **::** if $A \in F$ then $A^c \in F$

$F3$ **::** if $A_k \in F$, $\forall k \in \mathbb{N}$ then $\bigcup\limits_{k=1}^{\infty} A_k \in F$

Remark. F3 Axioms yields also for finite sequence of events using construction of infinte sequence which obtain returning of the finite sequence.

<u>Axioms of P:</u>

$A.$**:** $\mathbb{P}(\emptyset) = 0$

B.: if $A \subset B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$
C.: if $A \subset B$ then $\mathbb{P}(A \backslash B) = \mathbb{P}(A) - \mathbb{P}(B)$
D.: if $A_1, ..., A_n$ events then $\mathbb{P}(\overset{n}{\underset{k=1}{\cup}} A_k) \leq (\sum\limits_{k=1}^{n} \mathbb{P}(A_k)$.

Proof. In induction
Base : for $n = 1$ (trivial)
Step : $n \geq 1$ assume it's true for n and we will show for $n + 1$.

$$\mathbb{P}(\overset{n+1}{\underset{k=1}{\cup}} A_k) = \mathbb{P}(\overset{n}{\underset{k=1}{\cup}} A_k \cup A_{n+1}) \leq_* \mathbb{P}(\overset{n}{\underset{k=1}{\cup}} A_k) + \mathbb{P}(A_{n+1}) = \sum_{k=1}^{n+1} \mathbb{P}(A_k)$$

. □

when in $*$ we used the fact that $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$.

Claim. .

$$\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$$

.

Proof. .

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

□

Claim. If A_k a sequence of events S.T $\mathbb{P}(A_k) = 1, \forall k \in \mathbb{N}$ then $\mathbb{P}(\underset{k}{\cap} A_k) = 1$

Proof. .

$$\mathbb{P}((\underset{k}{\cap} A_k)^c) = \mathbb{P}((\cup A_k^c) \leq \sum \mathbb{P}(A_k^c) = 0 \Rightarrow \mathbb{P}(\underset{k}{\cap} A_k) = 1.$$

□

Claim. If $\{A_k\}$ is a countable sequence *not-decreasing th*en $\mathbb{P}(\overset{\infty}{\underset{k=1}{\cup}} A_k) = \underset{k \to \infty}{lim} \mathbb{P}(A_k)$.

Proof. Define $A_0 = \emptyset$, then :

$$\mathbb{P}(\overset{\infty}{\cup} A_k) = \mathbb{P}(\overset{\infty}{\cup}(A_k - A_{k-1})) =_{*B_K = A_k - A_{k-1}} = \sum^{\infty} [\mathbb{P}(A_k) - \mathbb{P}(A_{k-1})] = \underset{k \to \infty}{lim} \mathbb{P}(A_k) - \mathbb{P}(A_0) = \underset{k \to \infty}{lim} \mathbb{P}(A_k).$$

□

Theorem. *Inclusion–Exclusion Principle.*
$\forall k \in \mathbb{N}$, A_k are events then the following satisfied :

$$\mathbb{P}(\overset{n}{\cup} A_k) = \mathbb{P}(A_1) + .. + \mathbb{P}(A_n) - (\mathbb{P}(A_1 \cap A_2) + ... + \mathbb{P}(A_{n-1} \cap A_n) + .. + (-1)^{n+1} \mathbb{P}(A_1 \cap .... \cap A_n)$$

.

Remark. We Generalized The Theorem to $\mathbb{N}$.
According to definition of the probability space (Ω, F, P) We can notice that :

- F is closed under countable unions

- P is defined for all elements of F.
- P is additive for union of disjoint elements of F.

The colloraly of the definition and the De Morgan Laws is:

- F is closed under countable intersections.

Let us consider a series A_i where $A_i \in F$,$i = 1, 2..$ The probability of the union can be constructed recursively as follows: First step is clear :

$$\mathbb{P}(\bigcup_{i=1}^{1} A_i) = \mathbb{P}(A_1)$$

And we know for it that $A_1 \in F$.

The recursive step is

$$\mathbb{P}(\bigcup_{i=1}^{n} A_i) = \mathbb{P}(\bigcup_{i=1}^{n-1} A_i) + \mathbb{P}(A_n) - \mathbb{P}([(\bigcup_{i=1}^{n-1} A_i \cap A_n])$$

The challenge is whether this $\bigcup_{i=1}^{n} A_i \in F$. We can resolve it as follows: from the recursion we know that $\bigcup_{i=1}^{n-1} A_i \in F$ and from problem formulation that $A_n \in F$.

Since F is closed under intersections, we know that $(\bigcup_{i=1}^{n-1} A_i \cap A_n) \in F$. Thus, the recursive step is well defined.

Hence, repeating it till infinity, We obtain the calculation for any countable set.

Claim. If $\{A_k\}$ is a sequence of countable *not-incresing events then* $\mathbb{P}(\bigcap_{k=1}^{\infty} A_k) = \lim_{k\to\infty} \mathbb{P}(A_k)$

Proof. we will calculate the probability of the complement event :

$$\mathbb{P}(\bigcap_{1}^{\infty} A_k)^c) = \mathbb{P}(\bigcup_{1}^{\infty} A_k^c) = \lim_{k\to\infty} \mathbb{P}(A_k^c) = 1 - \lim_{k\to\infty} \mathbb{P}(A_k)$$

ant the claim stem by Axiom P2. □

DISCUSSION ABOUT INFINITE PROBABILITY MODULES

- if we choose a random natural number in a uniform way we get contradiction since

$$1 = \mathbb{P}(\mathbb{N}) = \mathbb{P}(\bigcup_{n\in\mathbb{N}}\{n\}) = \sum_{n=1}^{\infty} \mathbb{P}(\{n))$$

denote α=$\mathbb{P}(\{n)\}$ if $\alpha > 0$ then the sum diverges, and if $\alpha = 0$ then the sum .

Corollary. *there is no way to define uniform probability on* $\Omega = \mathbb{N}$.

- The module $\Omega = [0, 1]$ we require uniform-probability i.e $\mathbb{P}(A) = \mathbb{P}(A^r)$ (cyclic right shifting by r).

we will try to do identical trick :

$$1 = \mathbb{P}([0,1]) = \mathbb{P}(\underset{[0,1]}{\cup}\{x\}) = \underset{[0,1]}{\cup}\mathbb{P}(\{x\}) = 0$$

Maybe indeed we require that $\mathbb{P}(\{x\}) = 0, \forall x \in$[0,1] in other hand $\exists E \subseteq [0,1]$ S.T:

(1) $E^r \cap E^s = \emptyset, \forall s, r \in [0,1], r \neq s$
(2) $\cup E^r = [0,1]$

Hence we got :

$$1 = \mathbb{P}([0,1]) = \mathbb{P}(\underset{\mathbb{Q}\cap[[0,1]}{\cup}E^r) = \sum_{r\in\mathbb{Q}\cap[[0,1]}\mathbb{P}(E)^r = \sum_{r\in\mathbb{Q}\cap[[0,1]}\mathbb{P}(E)$$

from here we get contradiction again as before, So, there is no uniform probability module on $[0,1]$.

- possible to find $F \underset{\neq}{\subseteq} 2^{[0,1]}$ S.T every interval $[a,b] \in F$ and there is p uniform on F

- F is called Sigma-Algebra if it satisfies Axioms $F1, F2, F3$.

CONDITIONAL PROBABILITY

Let (Ω, P, F) sample space and B event not trivial ($\mathbb{P}(B) > 0$)

Define new probability function on F :$\mathbb{P}(A|B)$,the conditional probability given by :

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

Lemma. *$A \to \mathbb{P}(A|B)$ is a probability function which satisfy $P1, P2, P3$.*

Proof. . □

$P1$ :: $\mathbb{P}(\emptyset|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\emptyset)}{\mathbb{P}(B)} = \frac{0}{\mathbb{P}(B)} = 0$
$P2$ :: $\mathbb{P}(A^c|B) = \frac{\mathbb{P}(A^c\cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B\setminus(A\cap B))}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} - \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = 1 - \mathbb{P}(A|B)$
$P3$ :: $\mathbb{P}(\cup A_k|B) = \frac{\mathbb{P}([\cup A_k]\cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\cup(A_k\cap B])}{\mathbb{P}(B)} = \frac{\sum\mathbb{P}(A_k\cap B)}{\mathbb{P}(B)} = \sum\mathbb{P}(A_k|B)$

Remark. $\mathbb{P}(A\cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$ if $\mathbb{P}(B) > 0$.

Lemma 1. *(THE FULL PROBABILITY FORMULA)*

Assume that there is collection of $\{B_k\}_{k=1}^{\infty}$ of events (not trivial) imply partitionn of Ω i.e :

$$\Omega = \underset{k}{\cup}B_k, \mathbb{B}_k \cap B_d = \emptyset, k \neq d$$

Then :

$$\mathbb{P}(A) = \sum_k \mathbb{P}(A|B_k)\mathbb{P}(B_k)$$

Example. Given m balls white and n black in jug, we take two balls with no return
$A = \{\text{second ball black}\}$
$B_1 = \{\text{First ball white}\}$
$B_2 = \{\text{First ball black}\}$

SECOND-2th
FIRST-1th
BLACK-B
WHITE-W

By the full probability formula we get :

$$\mathbb{P}(2th - B) = \mathbb{P}(2th - B|1th - B) \cdot \mathbb{P}(1th - B) + \mathbb{P}(2th - B|1th - W) \cdot \mathbb{P}(1th - W)$$
$$=(\tfrac{n-1}{n+m-1}) \cdot (\tfrac{n}{n+m}) + (\tfrac{n}{n+m-1}) \cdot (\tfrac{m}{n+m}) = (\tfrac{n(n+m-1}{(n+m-1)(n+m)}) = \tfrac{n}{n+m}.$$

Proof. .

$$\mathbb{P}(A) = \mathbb{P}(\cup(A \cap B_k)) = \sum_k \mathbb{P}(A \cap B_k) = \sum_k \mathbb{P}(A \backslash B_k)\mathbb{P}(B_k)$$

□

BAYES-FORMULA

Assuming $\mathbb{P}(A) > 0, \mathbb{P}(B) > 0$ then :

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

Proof. we need to show that $\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)} \iff \mathbb{P}(B|A)\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B)$
i.e $\mathbb{P}(B \cap A) = \mathbb{P}(A \cap B)$ which is trivial. □

Example. a medecine is effective in many levels with 3 type of disease:
$A-$affective in 0.8, disease recover $0,1$
$B-$affective in 0.6, disease recover $0,5$
$C-$affective in 0.4, disease recover $0,4$

If person show any of those syptoms and the medicine help, what the probability it was disease A?

Solution. we are looking for:

$$\mathbb{P}(A|recovered) = \frac{\mathbb{P}(recovered|A)\mathbb{P}(A)}{\mathbb{P}(recovered)} = \frac{0.8 \cdot 0.1}{\mathbb{P}(recovered)}$$

We will calculate $\mathbb{P}(recovered)$ by the full probability formula

$$\mathbb{P}(recovered) = \mathbb{P}(recovered|A)\mathbb{P}(A) + \mathbb{P}(recovered|B)\mathbb{P}(B) + \mathbb{P}(recovered|C)\mathbb{P}(C)$$
$$=0.8 \cdot 0.1 + 0.6 \cdot 0.5 + 0.4 \cdot 0.4 = 0.54$$

So in total :

$$\mathbb{P}(A|recovered) = \frac{0.8}{0.54} \sim 0.15$$

Definition. We says that two event A, B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

Remark. Notice that:
If $\mathbb{P}(B) > 0$ stem that $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A)$.
If $\mathbb{P}(A) > 0$ stem that $\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \mathbb{P}(B)$.

Definition. (Generelaized)
$n \in \mathbb{N}$, $A_1, ..., A_n$ events independent if :

$$\forall i \neq j, \mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \cdot \mathbb{P}(A_j)$$

$$\forall i \neq j \neq k, \mathbb{P}(A_i \cap A_j \cap A_k) = \mathbb{P}(A_i) \cdot \mathbb{P}(A_j) \cdot \mathbb{P}(A_k)$$

$$\vdots$$

$$\mathbb{P}(A_1 \cap A_2 ... \cap A_n) = \prod_{k=1}^{n} \mathbb{P}(A_i)$$

Example. We will give example of three events A, B, C independent in pairs but not independent

$$\Omega = \{1, 2, 3, 4\},$$

$$\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \mathbb{P}(\{3\}) = \mathbb{P}(\{4\}) = \frac{1}{4}$$

Define the following :

$$A = \{1, 2\}, B = \{1, 3\}, C = \{2, 3\}$$

Then :

$$\mathbb{P}(B) = \mathbb{P}(A) = \frac{1}{2}, \mathbb{P}(A \cap B) = \mathbb{P}(\{1\}) = \frac{1}{4}$$

Then :

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Hence, A, B are independent and in the same way, B, C and A, C.
But $A \cap B \cap C = \emptyset$ Which mean :

$$0 = \mathbb{P}(A \cap B \cap C) \neq \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C) = \frac{1}{8}$$

So A, B, C, are not independent.

Claim. If A, B independent then A, B^c independent.

Proof. .

$$\mathbb{P}(A \cap B^c) = \mathbb{P}(A \backslash A \cap B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$$

$$= \mathbb{P}(A) - \mathbb{P}(A) \cdot \mathbb{P}(B) = \mathbb{P}(A)(1 - \mathbb{P}(B)) = \mathbb{P}(A) \cdot \mathbb{P}(B^c)$$

□

Claim. (Generalized)

If $A_1, ..., A_n$ independent then every n events from type $A^{r_1}, ..A^{r_n}$ when $r \in \{0,1\}^n$ and $A^1 = A^c.A^0 = A$ Independent.

Proof:: We will show that :

$$\mathbb{P}(A_1^c \cap ....A \cap A_{l+1} \cap ...A_n) = \prod_{i=1}^{l} \mathbb{P}(A_i^c) \cdot \prod_{i=1}^{n} \mathbb{P}(A_i)$$

Using induction on l .

$l = 0$ is the given since $A_1, ..., A_n$ are independent.

assume that the lema hold for l and we wills how for $l + 1$.

$$\mathbb{P}(A_1^c \cap .... \cap A_{l+1}^c \cap A_{l+2} \cap ... \cap A_n) = \mathbb{P}(A_1^c \cap .. \cap A_l^c \cap (\Omega \backslash A_{l+1}) \cap A_{l+2} \cap ... \cap A_n)$$

$$= \mathbb{P}(A_1^c \cap ... \cap A_l^c \cap \Omega \cap A_{l+2} \cap .. \cap A_n) \backslash [\bigcap_{k=1}^{l} A_k^c \bigcap_{k=l+1}^{n} A_k])$$

$$= \mathbb{P}(A_1^c \cap ...A_l^c \cap \Omega \cap ... \cap A_n) - \mathbb{P}(A_1^c \cap .. \cap A_l^c \cap A_{l+1} \cap .. \cap A_n)$$

$$= \mathbb{P}(A_1^c) \cdots \mathbb{P}(A_l^c) \cdot \mathbb{P}(A_{l+3}) \cdots \mathbb{P}(A_n)[1 - \mathbb{P}(A_{l+1})] =_{1-\mathbb{P}(A_{l+1}) = \mathbb{P}(A_{l+1}^c)} \prod_{k=1}^{l+1} \mathbb{P}(A_i^c) \cdot \prod_{i=1}^{n} \mathbb{P}(A_i)$$

Q.E.D.

Collection of non-finite events

Definition. we will say that $\{A_\alpha\}, \alpha \in A$ when A is a set of indexes not finite.(not countable necessarily)

Is a collcetion of independent events $A_{\alpha_1}, ..., A_{\alpha_r}$ independent $\forall r \in \mathbb{N}$

Moreover, $\forall \alpha_1 < .. < \alpha_r \in A$.

Bernulli-Trials

A finite sequence or counable trials in which they could get $0 \vee 1$

(success or failure, head or tale) which satisfies :

(1) The probability of every trial to get success is equal for all trials and define by p when $0 \leq p \leq 1$.
(2) the event $H_k = \{\text{success in trial } k\}$, $k = 1, 2, .., n$ are independent .

Example. two trials of rolling a coin twice

$\mathbb{P}(Success) = 0.5$

$\mathbb{P}(\text{identical to first roll}) = 0.8$

what is the probability to success in the second roll?

Solution. we will use the full probability Formula :

$$\mathbb{P}(succes{-}2) = \mathbb{P}(success{-}2|success{-}1)\cdot\mathbb{P}(success{-}1)+\mathbb{P}(success{-}2|success{-}2)\cdot\mathbb{P}(success{-}2) =$$

$$0.2\cdot 0.5+0.8\cdot 0.5 \quad = \quad 0.5$$

this are bernulli trials, although the probility preserve between the steps.

Exercise. In n bernulli trials with success parameter p
what is the probability for $k-success$? $k = 0, 1, 2....$

Example. what the probabilty for 3 success in 8 trials?

Solution. we will describe the trials as a squence of $0, 1$ then :
{3 success out of 8}={(11100000), {11010000}, ..(00000111)}
when

$$\mathbb{P}(11100000) = p^3\cdot q^5$$

$$\mathbb{P}(11010000) = p^2qpq^4 = p^3q^5$$

And identical for the rest, some number of options to success 3 out of 8 is $\binom{8}{3}$
hence we get

$$\binom{8}{3}p^3q^5$$

In generalizd case we get

$$\mathbb{P}(k-succcess-from-n) = \binom{n}{k}p^kq^{n-k}$$

Remark. The sample-space Ω is uniform for all events $\{k\text{-success}\}$ and indeed

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(\bigcup_{k=0}^{n}\{k-success\}) = \sum_{k=0}^{n}p^kq^{n-k} = (p+q)^n$$

Exercise. In infinte aequence of bernulli trials with paramater succeess p
what is the probabilty that the first time we get success in trial n?

Solution. first notice that :

$$\mathbb{P}(\bigcap_{k=1}^{n-1}\{fail-in-trial-k\}\cap\{\{success-in-trialn\}) = \prod_{k=1}^{n-1}q\cdot p = q^{n-1}\cdot p$$

Check::
$\Omega = \bigcup_{n=1}^{\infty}$**:** {first success in n trial} indeed,

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(\bigcup_{n=1}^{\infty}\{first-success-in-n-trial\} = \sum_{n=1}^{\infty}pq^{n-1} = \frac{p}{1-q} = 1$$

Notes::

- All that true if $p > 0$
- unfortunately, there was non accuracy, Since we didn't take into account the event $A_\infty = \{no-success\}$ but $\forall m, \mathbb{P}(A_\infty) < q^m$ (i.e the prbability of m failure on the beggining) hence $\mathbb{P}(A_\infty) = 0$

Claim. If $\mathbb{P}(A\backslash \overset{\infty}{\cup} A_k) = 0$ for $\{A_i\}$ dijoint events then, $\mathbb{P}(A) = \sum\limits_{k=1}^{\infty} \mathbb{P}(A_k)$

Proof. Notice that :

$$A = \bigcup_{k=1}^{\infty} A_k \cup (A\backslash \bigcup_{k=1}^{\infty} A_k) \Rightarrow_{\mathbb{P}(A_\infty)=0, A\backslash \bigcup\limits_{k=1}^{\infty} A_k = A_\infty} \mathbb{P}(A) = \sum_{k=1}^{\infty} \mathbb{P}(A_k) + \mathbb{P}(A_\infty)$$

□

RANDOM VARIABLES

First we will start with probability space in which $F = 2^\Omega$ (every possible subset of Ω)

Definition. Let (Ω, F, P) be probability space in which $F = 2^\Omega$, function $X : \Omega \to \mathbb{R}$ is a Random-Variable.

Example. $\Omega = \{0,1\}^n, \mathbb{P}(\{\epsilon_1, ..., \epsilon_n\}) = p^{\sum_1^n \epsilon_k} \cdot q^{n - \sum_1^n \epsilon_k}$ when $p \in (0,1), q = 1 - p, (\epsilon_1, ...\epsilon_n) \in \{0,1\}^n$this is bernolli module of n trials.

Define $\underset{w \in \Omega}{X((\epsilon_1, ...\epsilon_n))} = \sum\limits_{k=1}^{n} \epsilon_k$ i.e $X =$ number of success
Define $Y((\epsilon_1, ...\epsilon_n)) = \sum\limits_{k=1}^{n} \epsilon_k \epsilon_{k-1}$ i.e $Y=$ number of consecutive pairs of success.

Definition. Let X be random variable which define on Ω and assume that X get a finite number or countable values then X called *Discrete-Random-Variable*.

Remark. previous examples were describing *Discrete-Random-Variable*.

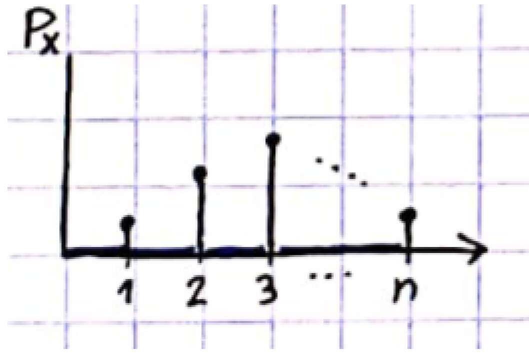
Definition. let X be random variable on $(\Omega, 2^\Omega, P)$ then the function $\mathbb{P}_x : \mathbb{R} \to [0,1]$ which defined by $\mathbb{P}_x(x) = \mathbb{P}(\{\omega | X(\omega\} = x\})$ is called probability function of x.

Example. if exist n bernulli trials with parameter p and $X =$ number of success then

$$\mathbb{P}_X(3.5) = 0$$

$$k \in \mathbb{N} : \mathbb{P}_X(k) = \mathbb{P}(k - success) = \binom{n}{k} p^k q^{n-k}$$

$$\sum_{x \in \mathbb{R}} \mathbb{P}_X(x) = 1 = \mathbb{P}(\Omega)$$



Remark. notice that this sum is legal since since there is only countable valuse which interfer in the sum.

GENERAL-CASE::
(Ω, F, P)**:** is a sample space when F not necessiraly $F = 2^{\Omega}$

Definition. (Generalized)
$X : \Omega \to \mathbb{R}$ is a random variable if $\forall x \in \mathbb{R}$, $B_x = \{\omega | X(\omega\} \leq x\} \in F$

Remark. notie that :

- Instead of $B_x = \{\omega | X(\omega) \leq x\}$ we can write $\mathbb{P}(B_x) = \mathbb{P}(X \leq x)$.
- Indeed $\{X = x\} \in F\ \forall x$ if we assume $\{X \leq x\} \in F, \forall x$ since if for every x. $\{X \leq x\} \in F$ then the complement $\{X > x\} \in F$,$\forall x$. in particular $\forall n$,$\{X > x - \frac{1}{n}\} \in F$ Hence, $\{X \leq x\} \cap \bigcap_{n \in \mathbb{N}} \{X > x - \frac{1}{n}\} = \{X = x\} \in F$

Corollary. *Also in case in which* $F \neq 2^{\Omega}$, $\mathbb{P}_X(x) = \mathbb{P}(X = x)$ *defined well for discrete-Random-Variable* X.

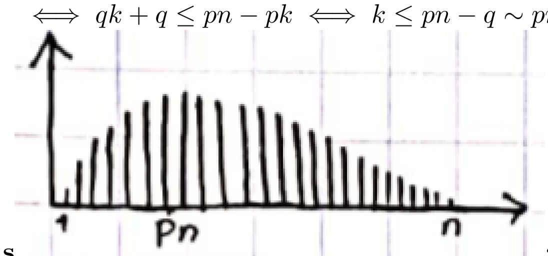
REMIND-THE-EXAMPLE:
$X =$**:** number of success in n bernulli trials
$$\mathbb{P}_X(x) = \left\{ \begin{array}{cc} \binom{n}{x} p^x q^{(n-x)}, & x = 0, .., n \\ 0 & otherwise \end{array} \right\}\text{:}$$

The probability function called binomial with parameters $n \in \mathbb{N} \wedge p \in [0, 1]$
we will say X is a binomial random variable, $X \sim Bin(n, p)$

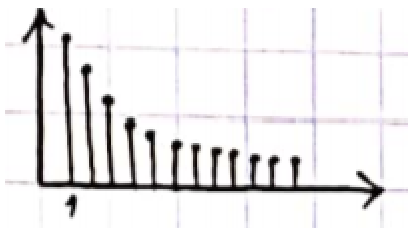
Remark. The binomial probability function have 1 max (or 2 adjacent).
it's a increasig funtion till max and decrease after that (as the following diagram)

check-max::
$$\mathbf{P}_X(k) \leq \mathbb{P}_X(k+1) \iff \binom{n}{k} p^k q^{n-k} \leq \binom{n}{k+1} p^{k+1} q^{n-k-1} \iff \frac{1}{n-k} \cdot q \leq \frac{1}{k+1} p\text{:}$$

$$\iff qk + q \leq pn - pk \iff k \leq pn - q \sim pn\text{:}$$

s  **:**

Example. (Geometric Random-Variable)
we will look at the infinite sequence of bernulli trials.
Y = number of trials till we get first time success
$\mathbb{P}_{\gamma}(k) = p \cdot q^{k-1}$ $(k \in \mathbb{N})$
function like that called Geometric function, $Y \sim Geom(p)$



Example. (Poisson Random-variable)
A random variable discrete is poisson with parameter $\lambda > 0$
if it has probability function :

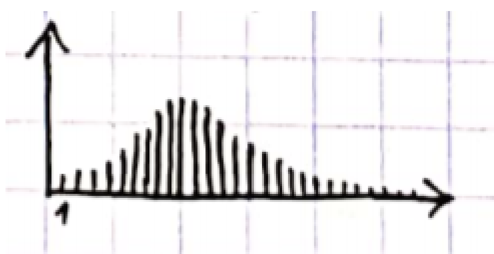
$$\mathbb{P}_X(k) = C\frac{\lambda^k}{k!}, C \in \mathbb{R}$$

We will find C:

$$\mathbb{P}(\Omega) = 1 \Rightarrow \sum_{k=0}^{\infty} \mathbb{P}_X(k) = 1 \Rightarrow C \cdot \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = 1 \Rightarrow C \cdot e^{-\lambda} = 1 \Rightarrow C = e^{-\lambda}$$

i.e

$$\mathbb{P}_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}$$



Exercise. A company produce valves, the valve could be defective in probability $\beta = 0.008$ and independent to others.

The company check a 1000 valves, if 8 or more are defect then it close it's maintenance, what is the probability that it will close?

Solution. We are talking about 1000 bernulli trials with probability of 0.008 to success in every trial .

Denote X =number of success, we need to calculate $\mathbb{P}(X \geq 9)$

$$\mathbb{P}(X \geq 8) = \sum_{8}^{1000} \binom{1000}{8} (0.008)^k (0.992)^{1000-k} = 1 - \sum_{0}^{7} \binom{1000}{8} (0.008)^k (0.992)^{1000-k}$$

Theorem. *Let $\lambda > 0$ and let X_n Random variables $S.T$ $X_n \sim Bin(n, \frac{\lambda}{n}), \forall n$ $S.T$, $\frac{\lambda}{n} < 1$ then :*

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

Remark. $\mathbb{P}(X_\infty = k)$ when $X_\infty \sim poiss(\lambda)$

Corollary. *when n is big enough for example in the previous example we can say that*

$$\mathbb{P}(X_{1000} \leq 7) \sim e^{-8}(1 + \frac{8}{1} + \frac{64}{2} + ... + \frac{8^7}{7!}) = 0.54$$

when 8 stem form $\lambda = n \cdot p = 1000 \cdot 0.008 = 8$

Proof. Notice that :

$$\mathbb{P}(X_n = k) = \binom{n}{k} p_n^k q_n^k = \binom{n}{k} (\frac{n}{\lambda})^k (1 - \frac{\lambda}{n})^{n-k}$$

$$= \frac{\lambda^k}{k!} \cdot \frac{n \cdot (n-1) \cdots (n-k+1)}{n \cdot n \cdot n} \cdot (1 - \frac{\lambda}{n})^n (1 - \frac{\lambda}{n})^{-k} \underset{n\to\infty}{\to} \frac{\lambda^k}{k!} \cdot 1 \cdot 1 \cdots 1 \cdot e^{-\lambda} \cdot 1 = e^{-\lambda}\frac{\lambda^k}{k!}$$

□

Definition. A random variable X on a probability space (Ω, F, P) is a absolutely continious random variable

If exist non-negative funtion and integrable on $\mathbb{R}$, $f_X(x)$ s.t for every $I = (a, b) \subseteq \mathbb{R}$ (boundded ot not bounded)

$$\mathbb{P}(a < X < b) = \int_a^b f_X(x)dx$$

f is called the density function of X.

Remark. If Y is a discrete randokm variable, then :

$$\mathbb{P}(a < Y < b) = \sum_{a<y<b} \mathbb{P}_\gamma(y)$$

Remark. $\int_{-\infty}^{\infty} f_X(x)dx = 1$ for every random variable X Since, the integral is $\mathbb{P}(-\infty < X < \infty) = \mathbb{P}(\Omega)$

Remark. if X absolutely continious with $f_X(x)$ and $\widetilde{f}(x)$ another function S.T $\{x \in \mathbb{R} | \widetilde{f}(x) \neq f_X(x)\}$ finite or countable with no limit points, then $\widetilde{f}(x)$ is also density function of X.

Remark. When we assume that f_X is integrable, we mean in the improper inegral, and it could be for f_X a finite or countable number of singular points (not limits one), i.e f_X not necessarily bounded, but the integral on $\mathbb{R}$ is 1.

Claim. if X random variable then $\forall x_0 \in \mathbb{R}, \mathbb{P}(X = x_0) = 0$

Proof. $\forall \epsilon > 0$,

$$\mathbb{P}(X = x_0) \leq \mathbb{P}(x_0 - \epsilon < X < x_0 + \epsilon) = \int_{x_0-\epsilon}^{x_0+\epsilon} f_X(x)dx \leq M \cdot 2\epsilon$$

assumig that $|f_X(x)| < M$ in neighberhood of x_0

$$\Downarrow$$

$$\mathbb{P}(X = x_0) = 0$$

If x_0 a singular point i.e, $f_X(x) \underset{\substack{x\to x_0^+ \\ x\to x_0^-}}{=} \infty$, $\int\limits_{x_0-\epsilon}^{x_0+\epsilon} f_X(x)dx \underset{\epsilon\to 0}{\to} 0$

So anyway we get that $\mathbb{P}(X = x_0) = 0$ $\square$

Example. (Uniform-Random-Variable)

$$f_X(x) = \left\{ \begin{array}{cc} C > 0 & x \in (\alpha, \beta) \\ 0 & x > \alpha \\ 0 & x < \beta \end{array} \right\}$$

A rendom variable X with density called uniforom, $X \sim U[\alpha, \beta]$, adn the value of f_X in a finite points is not important, we can define it also in finite number points.

The value of C should be $\frac{1}{\beta-\alpha}$ in order the Volume obtain 1.

Example. (Exponential-Random-Variable)

$$f_X(x) = \left\{ \begin{array}{cc} \lambda e^{-\lambda x} & , x > 0 \\ 0 & , otherwise \end{array} \right\}$$

Check:

$\int\limits_{-\infty}^{\infty} \lambda e^{-\lambda x} dx = \lambda \frac{e^{-\lambda x}}{\lambda}|_{\infty}^{0} = 1$**:**

X Random variable Gamma with parameter $\Gamma \in \mathbb{N}, \lambda > 0$

$$f_X(x) = \left\{ \begin{array}{cc} \frac{\lambda^r}{(r-1)!} x^{r-1} e^{-\lambda x}, & x > 0 \\ 0 & , x \leq 0 \end{array} \right\}$$

It's a family with two-parameters (Γ, λ) of a Randon-Variable which obtain exponential density as a private case when $\Gamma = 1$.

Example. (Normal.Gaussian-Random-Variable)

if $f_X(x) = C \cdot e^{-\frac{x^2}{2}} dx, C \in \mathbb{R}$.

We will calculate the Value of C, Denote $I = \int\limits_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx$

$$I^2 = \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \int\int_{\mathbb{R}^2} e^{\frac{x^2+y^2}{2}} dxdy = \int_0^{2\pi} (\int_0^{\infty} e^{-\frac{r^2}{2}} rdr)d\theta = 2\pi e^{\frac{-r^2}{2}}|_0^{\infty} = 2\pi$$

$$\Downarrow$$

$$I^2 = 2\pi \Rightarrow I = \sqrt{2\pi} \Rightarrow C = \frac{1}{\sqrt{2\pi}}$$

Remark. We got r as a result of determinant of Jacobian Matrix.

X Random-Variable (Normal) General if :

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} . x \in \mathbb{R}$$

when $\mu \in \mathbb{R}, \sigma > 0$ are two parameters.

Definition. Let X Random-Variable which defined on (Ω, F, P) [X is descrete or absolutely continious]

The distirbution function $F_X(x)$ which defined :

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\{\omega | X(\omega) \leq x\}) = \mathbb{P}(X^{-1}((-\infty, x]))$$

Theorem. *every distirbution function satisfy :*

- $F_X(x)$ not decreasing
- $\lim_{x\to\infty} F_X(x) = 1, \lim_{x\to-\infty} F_X(x) = 0$
- $F_X(x)$ is continious in every right neighberhood $\forall x \in \mathbb{R}$

Proof. We will show the it's true

(1). Assuming $x_1 < x_2$ then :

$$F_X(x_1) = \mathbb{P}(X \leq x_1) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \leq x_1\}) \leq \mathbb{P}(\{\omega \in \Omega | X(\omega) \leq x_2\}) = F_X(x_2)$$

(2). Notice that ;

$$\lim_{x\to\infty} F_X(x) =_{Heine-Theorem} \lim_{n\to\infty} F_X(n) = \lim_{n\to\infty} \mathbb{P}(\{\omega \in \Omega | X(\omega) \leq n\}) = \mathbb{P}(\bigcup_{n=1}^{\infty} \{\omega | X(\omega) \leq n\}) = 1$$

$$\lim_{n\to\infty} F_X(-n) = \lim_{n\to\infty} \mathbb{P}(X \leq -n) = \mathbb{P}(\bigcap_{n=1}^{\infty} \{\omega | X(\omega) \leq -n\}) = 0$$

(3). Notice that :

$$\lim_{n\to\infty} F_X(x+\frac{1}{n}) = \lim_{n\to\infty} \mathbb{P}(X \le x+\frac{1}{n}) = \mathbb{P}(\bigcap_{n=1}^{\infty}\{X \le x+\frac{1}{n}\}) = \mathbb{P}(X(\omega) \le x\}) = F_X(x)$$

□

Remark. If X Discrete Random-Variable, $F_X(x) = \mathbb{P}(X \le x)$;

(1) $F_X(x) = \mathbb{P}(X \le x) = \sum_{y \le x} \mathbb{P}_X(y)$ (It's a discrete-sum)

(2) $\mathbb{P}_X(x) = F_X(x) - \lim_{x\to x^-} F(x)$

Proof. (2)

$$F_X(x^-) = \lim_{n\to\infty} F_X(X - \frac{1}{n}) = \mathbb{P}(\bigcup_{n=1}^{\infty}\{X \le x - \frac{1}{n}\}) = \mathbb{P}(X < x)$$

$$F_X(x) - F_X(x^-) = \mathbb{P}(X \le x) - \mathbb{P}(X < x) = \mathbb{P}(X = x) = \mathbb{P}_X(x)$$

□

Reminder-(Normal-Random-Variable):

If X Normal with μ, σ^2 ($X \sim N(\mu, \sigma^2)$) then :

$$\mathbb{P}(X \in (a,b) = \phi(\frac{b-\mu}{\sigma}) - \phi(\frac{a-\mu}{\sigma})$$

Proof. Notice that :

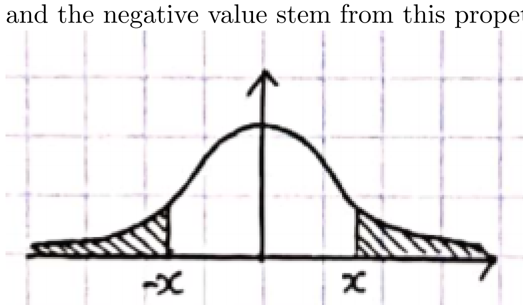
$$\mathbb{P}(y_1 < Y < y_2) = \frac{1}{\sqrt{2\pi}\sigma}\int_{y_1}^{y_2} e^{-\frac{(y-\mu)^2}{2\sigma^2}}dy =_{\substack{t=\frac{y-\mu}{\sigma}\\ dt=\frac{dy}{\sigma}}} \frac{1}{\sqrt{2\pi}}\int_{\frac{y_1-\mu}{\sigma}}^{\frac{y_2-\mu}{\sigma}} e^{\frac{-t^2}{2}}dt = \mathbb{P}(X \in (a,b) = \phi(\frac{y_2-\mu}{\sigma}) - \phi(\frac{y_1-\mu}{\sigma})$$

□

Remark. If $x > 0$ then :

$$\phi(-x) = \mathbb{P}(X \le -x) = \mathbb{P}(X > x) = 1 - \phi(x)$$

So, it's enough to hold a positive values of ϕ in a table and the negative value stem from this propety.

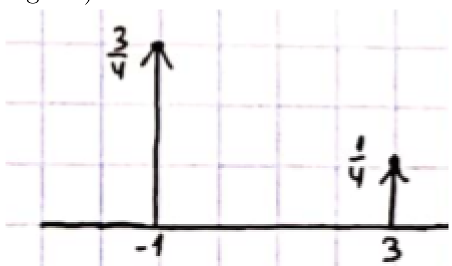


EXPECTATION

Example. Assuming ;

$$X = \left\{ \begin{array}{cc} -1, & \frac{3}{4} \\ 3, & \frac{1}{4} \end{array} \right\}$$

(it's like rolling a not fair coint when n probability $\frac{3}{4}$ it get -1 and probability $\frac{1}{4}$ it get 3)



We want to know "Center of mass" In this case we will say :

$$\mu_X = \frac{3}{4}(-1) + (\frac{1}{4}) \cdot 3 = 0$$

Definition. .

(1) if X discrete Random-Variable then $E[X] = \mu_X$ defined as

$$E[X] = \sum_{x \in \mathbb{R}} x\mathbb{P}_X(x)$$

(2) if X absolutely continious, $E[X] = \int\limits_{-\infty}^{\infty} xf_X(x)dx$ when f_X density function.

Then Sum in (1), and integral in (2) :

(1) if $\sum\limits_{x>0} x\mathbb{P}_X(x) < \infty$ and $\sum\limits_{x<0} x\mathbb{P}_X(x) > -\infty$ then there is no problem.
(2) if $\sum\limits_{x>0} x\mathbb{P}_X(x) = \infty$ and $\sum\limits_{x<0} x\mathbb{P}_X(x) > -\infty$ then $E[X] = \infty$.
(3) if $\sum\limits_{x>0} x\mathbb{P}_X(x) < \infty$ and $\sum\limits_{x<0} x\mathbb{P}_X(x) = -\infty$ then $E[X] = -\infty$.
(4) if $\sum\limits_{x>0} x\mathbb{P}_X(x) = \infty$ and $\sum\limits_{x<0} x\mathbb{P}_X(x) = -\infty$ then $E[X]$ not defined.

Example. .

- $X \sim Geom(P)$ then ;

$$E[X] = \sum_{k=1}^{\infty} kpq^{k-1} = p\sum_{k=1}^{\infty} kq^{k-1} =_{integrartion-element-element} p(\sum_{k=1}^{\infty} q^k)' = p(\frac{1}{1-q})'$$

$$= p(\frac{1}{(1-q)^2}) = \frac{p}{p^2} = \frac{1}{p}$$

- $X \sim Poiss(\lambda)$ then :

$$E[X] = \sum_{k=1}^{\infty} ke^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda}\lambda\sum_{k=0}^{\infty}\frac{\lambda^k}{(k)!} = e^{-\lambda}\cdot\lambda\cdot e^{\lambda} = \lambda$$

- $X \sim Bin(n,p)$ then :

$$E[X] = \sum_{k=0}^{n} k\frac{n!}{k!(n-k)!}p^k q^{n-k} = n!q^n x\sum_{k=1}^{n}\frac{x^{k-1}}{(k-1)!(n-k)!} = \frac{n!}{(n-1)!}q^n x\sum_{k=1}^{n}\frac{x^{k-1}(n-1)!}{(k-1)!(n-k)!}$$

$$= nq^n x\sum_{k=1}^{\infty}\binom{n-1}{k-1}x^{k-1} = nq^n x\sum_{k=0}^{n-1}\binom{n-1}{k}x^k 1^{(n-1)-k} = nq^n x(1+x)^{n-1}$$

$$= nq^n\frac{p}{q}(1+\frac{p}{q})^{n-1} = nq^n\frac{p}{q}(\frac{1}{q})^{n-1} = np$$

- $X \sim U[a,b]$ then :

$$E[X] = \frac{1}{b-a}\int_a^b xdx = \frac{x^2}{2(b-a)}|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$$

- $X \sim Exp(\lambda)$ then :

$$E[X] = \lambda\int_0^{\infty} xe^{-\lambda x}dx = \underbrace{\lambda x\frac{e^{-\lambda x}}{\lambda}|_0^{\infty}}_{=0} + \lambda\int_0^{\infty} 1\frac{e^{-\lambda x}}{\lambda}dx = \frac{e^{-\lambda x}}{\lambda}|_{\infty}^{0} = \frac{1}{\lambda}$$

- $X \sim N(\mu,\sigma^2)$,by semmetry, the intiuition is that $E[X] = \mu$ we will use the following Theorem in Next page.

Theorem. *If X absolutely continious random variable with symmetric density in x_0 i.e $\forall t, f_X(x_0+t) = f_X(x_0-t)$ then $E[X]$ exist and $E[X] = x_0$*

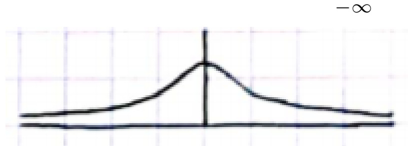
Example. Look at :

$$f_X(x) = C\cdot\frac{1}{1+x^2}$$

This is a function with symmetric density in x_0 **But** $E[X]$ doesn't exist.
Since :

$$C\cdot\int_0^{\infty} x\cdot\frac{1}{1+x^2}dx = \frac{ln(1+x^2)}{2}|_0^{\infty} = \infty$$

$$C\cdot\int_{-\infty}^{0} x\cdot\frac{1}{1+x^2}dx = -\infty$$



THE-Tail-Formula:

Let X be a random variable.

($\boldsymbol{A}$). If X is discrete then :

(1) $\sum\limits_{x>0} x\mathbb{P}_X(x) = \sum\limits_{x>0}(1 - F_X(x))$
(2) $\sum\limits_{x<0} x\mathbb{P}_X(x) = \sum\limits_{x<0}(-F_X(x))$

($\boldsymbol{B}$). If X is absolutely Contintious then :

(1) $\int\limits_0^\infty x f_X(x)dx = \int\limits_0^\infty (1 - F_X(x))dx$
(2) $\int\limits_{-\infty}^0 x f_X(x)dx = \int\limits_{-\infty}^0 (F_X(x))dx$

Corollary. *if $E[X]$ exist then :*

$$E[X] = \int_{-\infty}^{0} -F_X(x)dx + \int_{0}^{-\infty} (1 - F_X(x))dx$$

Proof. We will show B

($\boldsymbol{B}$.1).

$$\int_0^\infty (1 - F_X(x))dx = \int_0^\infty (\int_x^\infty f_X(t)dt)dx) = \int_0^\infty f_X(t) \int_0^t dxdt = \int_0^\infty t f_X(t)dt$$

($\boldsymbol{B}$.2).

$$-\int_{-\infty}^{0} (F_X(x))dx = -\int_{-\infty}^{0} (\int_{-\infty}^{x} f_X(t)dt)dx = -\int_{-\infty}^{0} f_X(t)(\int_t^0 dx)dt = \int_{-\infty}^{0} t f_X(t)dt$$

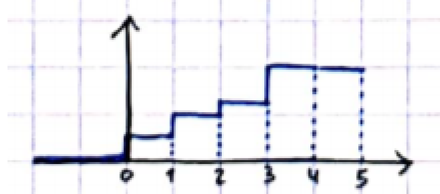
∞

$\square$

Corollary. *If X **Discrete-Random-Variable** which get $0, 1, 2....$ then :*

$$E[X] =_* \sum_{k=0}^{\infty} (1 - F_X(k))$$

*In * we used the Tail-Formula*



Example. A distance competetion, in which a jump is exponential $Exp(\lambda)$, what is the expactation that jumper N win jumper 0.

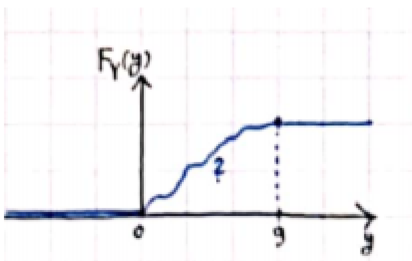
Solution. Notice that :

$$EN = \sum_{n=0}^{\infty}(1 - F_N(n)) = \sum_{n=0}^{\infty}\mathbb{P}(N > n) = \sum_{n=0}^{\infty}\frac{1}{n+1} = \infty$$

TRANSFORMATION OF RANDOM VARIABLES

Example. Let $X \sim U[0,3]$ and define $Y = X^2$ i.e $Y(\omega) = (X(\omega))^2$.

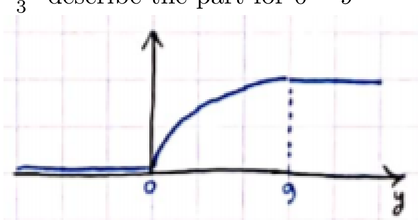
Question: : is Y also absolutely continious? If yes, what is $f_\gamma(y)$? Is $Y \sim U[0,3]$?



We will caclculate $F_\gamma(y)$?
let $0 < y < 9$ then

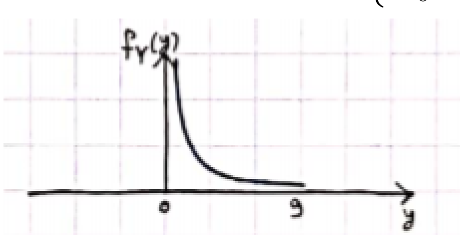
$$F_\gamma(y) = \mathbb{P}(Y \le y) = \mathbb{P}(-\sqrt{y} \le X \le \sqrt{y}) = \mathbb{P}(0 \le X \le \sqrt{y}) = \frac{\sqrt{y}}{3}$$

i.e, If we will drow $F_\gamma(y)$:
$\frac{\sqrt{y}}{3}$ describe the part for $0 - 9$



We will calculate $f_\gamma(y)$:

$$f_\gamma(y) = \left\{ \begin{array}{cc} \frac{1}{6\sqrt{y}} & ,0 < y < 9 \\ 0 & otherwise \end{array} \right\}$$



Definition. Let X **Random-Variable and** $h : \mathbb{R} \to \mathbb{R}$ **"Good Function"**, then $Y = h(X)$ called "Transformation of X" and it's itself **Random-Variable".**

Remark. In the next Lesson we will describe what is a "**Good Function**" and why we need this condition, and if $Eh(x) = h(EY)$

A MIXED DISTIRBUTION

Let X_1, X_2 **Random-Variables** and $0 \leq \alpha \leq 1$ we define :

$$F(x) = \alpha F_{X_1}(x) + (1-\alpha)F_{X_2}(x)$$

$F(x)$ is a distirbution dunction which satify :

- F(x) not decreasing
- $\underset{x\to\infty}{lim} F(x) = 1, \underset{x\to-\infty}{lim} F(x) = 0$ exist.
- $F(x)$ is continious in every right neighberhood.

We are looking for X **Random-Variable** S.T $F(x) = F_X(x)$

Proposition. $X = \alpha X_1(x) + (1-\alpha)X_2$.

Proposition. *Roliing a Coint S.T* $\alpha = success$

$$X = \left\{ \begin{array}{ll} X_1, & Success \\ X_2, & Failure \end{array} \right\}$$

Check::

$\mathbf{F}_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \leq x|Success)\alpha + \mathbb{P}(X \leq x|Failure) \cdot (1-\alpha)$**:**

=$\mathbf{P}(\mathbf{X}_1 \leq x)\alpha + \mathbb{P}(X_2 \leq x)(1-\alpha) = F_{X_1}(x) \cdot \alpha + F_{X_2}(x) \cdot (1-\alpha)$**:**

SO IT WORKS!!

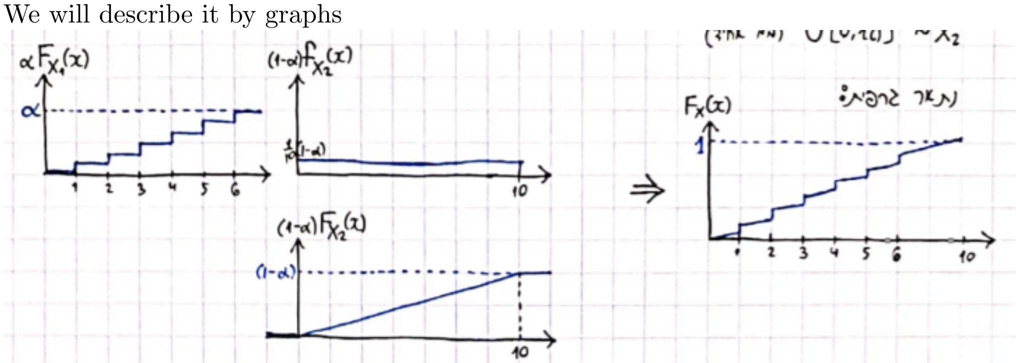
Remark. X called the mix of X_1, X_2.

Remark. X_1 can be discrete and X_2 can be continious- vice versa.

Example. X_1 = result of rolling a fair cube

X_2 ~ $U[0,10]$ (Uniform-Random-Variable)

We will describe it by graphs



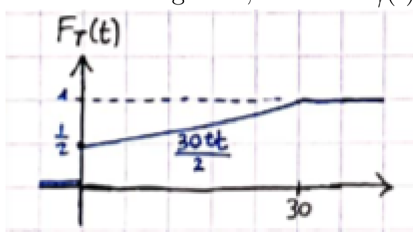
Theorem. *For every Random-Variable X, we can write*

$$F_X(x) = \alpha F^{(d)}(x) + (1-\alpha)F^{(c)}(x)$$

when $0 \leq a \leq 1$.

$F^{(d)}(x)$ is a discrete distirbution function.
$F^{(c)}(x)$ is a continious distirbution function.
if $\alpha = 0$ then X absouletly continious.
if $\alpha = 1$ then X discrete
otherwise, X mixed.

Example. A traffic light change light every $30sec$ green, $30sec$ red,
A car arrive in time X in a cycle of $1min$, $X \sim [0, 60]$
let $T =$ waiting time, what is $F_\gamma(t)$?



$\Rightarrow F_\gamma(t) = \frac{1}{2}F_{T_1}(t) + \frac{1}{2}F_{T_2}(t)$ when $T_1 = 0, T_2 \sim U[0, 30]$

BOREL SETS

Reminder:: F is $\sigma - Algebra$ on $\mathbb{R}$ if :

(1) $\emptyset \in F$.
(2) $B \in F \Rightarrow B^c \in F$.
(3) if $\{B_k\}_{k\in\mathbb{N}} \subseteq F \Rightarrow \underset{k\in\mathbb{N}}{\cup} B_k \in F$.

Theorem. *exist $\sigma - Algebra$, F which obtain F_0 and F is the smallest $\sigma - Algebra$ which obtain F_0 .*
<u>F</u> is called $\sigma - Algebra$ of borel on $\mathbb{R}, F = \beta(\mathbb{R})$, and in general

$$F = \sigma(\{(a, b)\}|\forall a < b)$$

Notice that if the collection are open sets then it's $\sigma - Algebra - Borel$.

Remark. $\sigma - borel$ $\beta(\mathbb{R}^1)$ obain all the open sets, closed, and lot more.

<u>Cantor set</u>

$C_n = \cup\{\text{open sets}\} \in \beta(\mathbb{R})$.

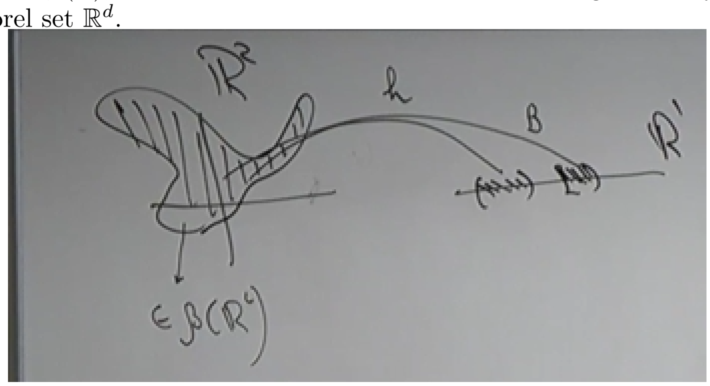
$C = \overset{\infty}{\underset{n=1}{\cap}} C_n \in \beta(\mathbb{R})$
Which mean everything we want we can get a borel set.

Remark. possible to define borel sets on $\mathbb{R}^d$, in the same way $\sigma - Algebra$ which produced by **<u>open balls</u>**

$$B_r(x) = \{y \in \mathbb{R}^d, |y - x| < r\}$$

This is the sigma-borel d -dimensional.

Definition. function $h : \mathbb{R}^d \to \mathbb{R}$ is called borel function if $h^{-1}(B) \in \beta(\mathbb{R}^d)$, $\forall B \in \beta(\mathbb{R})$, i.e h is borel function if the inverse image of every borel set in $\mathbb{R}^1$ is borel set $\mathbb{R}^d$.



A Random Variables & the connection with Borel Sets

REMIDNER.

$X : \mathbb{R} \to \mathbb{R}$ is a random variable if $X^{-1}((a,b)) = \{\omega : X(\omega) \in (a,b)\} \in F$ for every open interval (a,b) .

Theorem. *if X random-variable then $X^{-1}(\beta) \in F$ for every borel-set β (not only for $\beta = (a,b)$) .*

hint to proof :

$$\{\omega : X(\omega) \in (1,2) \cup (9,10)\} = X^{-1}((1,2) \cup (9,10)) = X^{-1}\overbrace{(1,2)}^{\in F} \cup X^{-1}\overbrace{(9,10)}^{\in F} \in_* F$$

explain "*" F i closed for union of sets in F.

Theorem. *if X random-variable and $h : \mathbb{R} \to \mathbb{R}$ then $Y = h(x)$ also is random-variable.*

Remark. before we mentions a **"good function"** we meant to say borel-set.

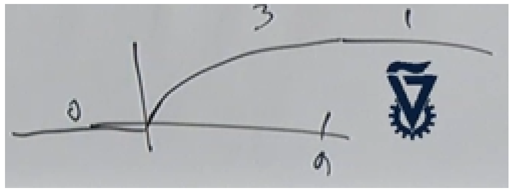
Proof. we need to show that $\forall B \in \beta(\mathbb{R})$, $Y^{-1}(B) \in F$. Notice :

$$Y(\omega) = (h \circ X)(\omega) \Rightarrow Y^{-1}(B) = X^{-1}(\underbrace{\ (h^{-1}(B)\ }_{borel-set}) \in F$$

Since the inverse of random-variable in borel set is borel set. □

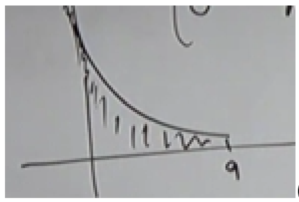
Back to $h(x)$ **:**
$X \sim U[0,3], Y = X^2$ what is $f_Y(y)$?

$$0 \le y \le 9, F_Y(y) = \mathbb{P}(X^2 \le y) = \mathbb{P}(X \in [0, \sqrt{y}]) = \frac{\sqrt{y}}{3}$$



The density of Y is :

$$f_Y(y) = \left\{ \begin{array}{cc} \frac{1}{6\sqrt{y}} & 0 < y < 9 \\ 0 & otherwise \end{array} \right\}$$



(Volume is equal to 1)

What is $E[Y]$?

$$E[Y] = \int_{-\infty}^{\infty} f_Y(y)ydy = \frac{1}{6}\int_0^9 \frac{y}{\sqrt{y}}dy = \frac{1}{6}\int_0^9 \sqrt{y} = \frac{1}{6}\cdot\frac{2}{3}\cdot y^{\frac{3}{2}}|_0^9 = \frac{27}{9} = 3$$

Notice that

$$3 = EY = EX^2 \neq (EX)^2 = \frac{9}{4} = 2.25$$

we will show that

$$EX^2 \vee E[X^2] \geq (E[X])^2$$

Theorem. *let X a absloute continious random-variable and $h : \mathbb{R} \to \mathbb{R}$ borel-function :*

$$\underbrace{Eh(X)}_{Y} = \int_{\mathbb{R}} h(x)f_X(x)dx$$

IN THE EXAMPLE :

$X \sim U[0,3], Y = X^2$ we could find $E[Y]$ like this :

$$E[Y] = \int_0^3 \underbrace{X^2}_{h(x)} \underbrace{\frac{1}{3}dx}_{f_X(x)(density-uniform\sim[a,b]-distibution(\frac{1}{b-a}))} = \frac{x^3}{3\cdot 3}|_0^3 = 3$$

Remark. we didn't need to calculate the density of $h(x)$ by using the Theorem.

Theorem. *if X a <u>discrete</u> random-variable and $h : \mathbb{R} \to \mathbb{R}$ BOREL-FUNCTION then :*

$$E[X] = \sum_{x\in\mathbb{R}} h(x)\mathbb{P}_X(x)$$

Example. $X \sim Geom(p)$ then :

$$EX^2 = \sum_{k=1}^{\infty} k^2 pq^{k-1}$$

Notice that we didn't got the second derivative of series **but** we can do something else ;

$$EX^2 = \sum_{k=1}^{\infty} k(k-1)pq^{k-1} + \underbrace{\sum_{k=1}^{\infty} kpq^{k-1}}_{\frac{1}{p}(expectation - Geometric - Random - Variabel} =_* pq \sum_{k=2}^{\infty} k(k-1)q^{k-2} + \frac{1}{p}$$

$$=_{**} pq(\underbrace{\sum_{k=0}^{\infty} q^k}_{\frac{1}{1-q}})'' + \frac{1}{p} = pq \cdot \frac{2}{(1-q)^3} + \frac{1}{p} = \frac{2pq}{p^3} + \frac{1}{p} = \frac{2q+p}{p^2} = \frac{1+q}{p^2}$$

hence,

$$EX^2 = \frac{1+q}{p^2}$$

Where in * we the series is from 2 since we $k = 1$ add 0 to the sum.
Where in $**$ we added $1 + q$ **but** when we derivative twice to it we get 0.

Check that :

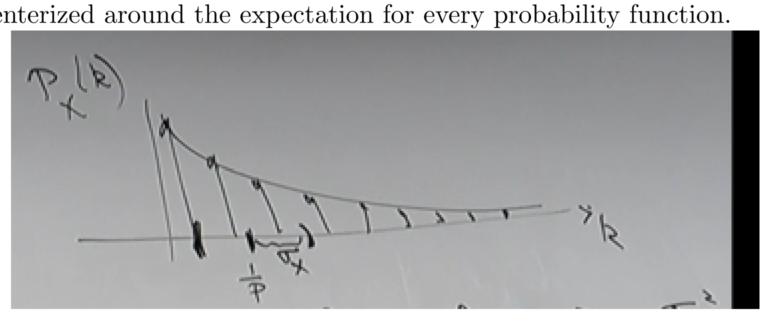
$$EX^2 > (EX)^2$$

$$EX^2 - (EX)^2 = \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2} > 0$$

Remark. Notice that :

The terminology of $EX^2 - (EX)^2$ is called the **Variance** .

Moreover, $\sigma_X = \sqrt{EX^2 - (EX)^2}$ is called **standard deviation** give us and it give us indication about which level of X is compressed far, or stretched far or centerized around the expectation for every probability function.



Theorem. *if X random variable and $h : \mathbb{R} \to \mathbb{R}$is borel function then :*

(1) if X a discrete random variable then $E[X] = \sum\limits_{x\in\mathbb{R}} h(x)\mathbb{P}_X(x)$
(2) id X is absoulote continious random variable then $E\underbrace{h(X)}_{Y} = \int\limits_{\mathbb{R}} h(x)f_X(x)dx$

Remark. We will show the Theorem only for absoulote continious random variable .

Proof. Denote $Y = h(x)$ and assume $Y \geq 0$ $(h : \mathbb{R} \to [0,\infty])$

$$EY = \int_0^\infty \underbrace{(1 - F_Y(y))}_{\mathbb{P}(Y>y)}dy = \int_0^\infty \mathbb{P}(\underbrace{h(x)}_{*} > y)dy = \int_0^\infty \Big(\int\limits_{\{x:h(x)>y\}} f_X(x)dx\Big)dy$$

$*(\mathbb{P}(h(x) > y) = \mathbb{P}(X \in \underbrace{\{x : h(x) > y\}}_{B})$

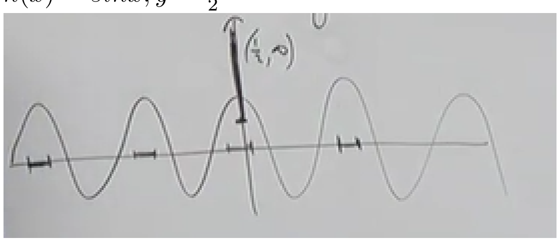
$\mathbb{P}(X \in B) = \int\limits_B f_X(x)dx$, h : is borel function

$h^{-1}((y,\infty))$ = borel set .

We will assume that $h^{-1}((y,\infty)$) not only borel sets **but** a union of countable intervals.

Example.

$h(x) = sinx, y = \frac{1}{2}$



Notice that we do integral on B which obtaine all this emphasized intervals.

$$\int_0^\infty \Big(\int\limits_{\{x:h(x)>y\}} f(x)dx\Big)dy = \int_{-\infty}^\infty \Big(\int\limits_{\{y:h(x)>y\}} f(x)dy\Big)dx = \int_{-\infty}^\infty \Big(f_X(x)\Big(\int\limits_{(0,h(x))} dy\Big)\Big)dx = \int_{-\infty}^\infty f_X(x)h(x)dx$$

in order to remove the assumption that $h(x) \geq 0$ we need to show that :

$$\int_{-\infty}^\infty F_X(y)dy = \int_{-\infty}^\infty h(x)f_X(x)dx$$

Identical way. □

Corollary. *let X random variable with expectation :*

- $E[X + C] = E[X] + C, \forall C \in \mathbb{R}$.
- $E[aX] = aE[X], \forall a \in \mathbb{R}$.

Proof.

$$E[X+C] = \int_{-\infty}^{\infty} (x+C)f_X(x)dx = \int_{-\infty}^{\infty} xf_X(x)dx + C\int_{-\infty}^{\infty} f_X(x)dx = EX + C\cdot 1 = EX + C$$

$$EaX = \sum_x ax\mathbb{P}_X(x) = a\sum_x x\mathbb{P}_X(x) = a \cdot EX$$

Variacne

Definition. let X random variable with finite expectation, the **Variance** of X defined as :

$$\sigma^2(x) = E(\underbrace{(X-\mu)^2}_{\geq 0}) \in [0, \infty]$$

Theorem. *let X random variable and h_1, h_2 borel functions then $h_1 + h_2$ is borel function:*

$$E[h_1(X) + h_2(X)] = E[(h_1 + h_2)(x) = E[h_1(X)] + E[h_2(X)]$$

in term that if right side defined then left side. the proof stem from previous Theorem.

Corollary. *$Var[X] = EX^2 - (EX)^2$ stem that $EX^2 - (EX)^2 \geq 0$ Since ($Var[X] \geq$* 0)

Proof. Notice that :

$$Var[X] = E(X-\mu_x)^2 = E(X^2 - 2\mu_x X + \underbrace{\mu_X^2}_{C}) = E(X^2 - 2\mu_x X) + \mu_X^2 = EX^2 + E(\underbrace{-\mu_X}_{C} X) + \mu_X^2$$

$$= EX^2 - 2\mu_X \underbrace{EX}_{\mu_X} + \mu_X^2 = EX^2 - 2\mu_X^2 + \mu_X^2 = EX^2 - (Ex)^2$$

□

<u>**Another formula for Variance:**</u>

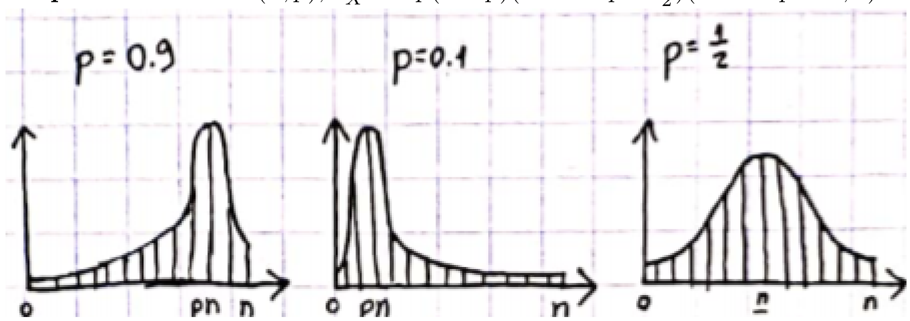
$$\sigma_X^2 = EX(X-1) - \mu_X(\mu_X - 1)$$

Proof. Notice that :

$$EX(X-1) - \mu_X(\mu_X - 1) = EX^2 - EX - \mu_X^2 + \mu_X = EX^2 - \mu_X^2 = \sigma_X^2$$

□

Example. let $X \sim Bin(n,p), \sigma_X^2 = np(1-p)$(max in $p = \frac{1}{2}$)(min in $p = 0, 1$)



Variance Calculation :

$$\sigma_X^2 = EX(X-1) - \underbrace{\mu_X(\mu_X - 1)}_{n^2p^2 - np} = \sum_{k=2}^{n} k(k-1)\frac{n!}{k!(n-k)!}p^k q^{n-k}$$

$$=_{x=\frac{p}{q}} n(n-1)q^n(\frac{p}{q})^n \sum_{k=2}^{n} \frac{(n-2)!x^{k-2}}{(k-2)!((n-2)-(k-2))!} = n(n-1)q^n \frac{p^2}{q^2} \sum_{k=0}^{n-2} \binom{n-2}{2} x^k$$

$$= n(n-1)q^{n-2}p^2(1+x)^{n-2} = ... = n^2p^2 - np^2 \Rightarrow \sigma_X^2 = npq$$

Variance properties:

(1) if $X = a$ (is trivial). then $\sigma_X^2 = 0$

$$EX^2 = a^2, (EX)^2 = a^2 \Rightarrow EX^2 - (EX)^2 = 0$$

(2) $Var(CX) = C^2 Var(X)$

$$Var(CX) = E(CX)^2 - (CEX)^2 = C^2(EX^2 - (EX)^2) = C^2 Var(X)$$

Exercise. let $X \sim N(\mu, \sigma^2)$ what is μ_X? what is σ_X^2?

Solution. 'The answer is μ and σ^2.in the tutorial we saw that if $\sigma = 1, \mu = 0$ then $Var[x] = 1$

lema. if $X \sim N(\mu, \sigma^2)$ then exists random variable $Z \sim N(0,1)$ S.T $X = \mu + \sigma Z$

Proof. $X = \mu + \sigma(\frac{X-\mu}{\sigma})$ so take $Z = (\frac{X-\mu}{\sigma})$ we will check that $Z \sim Norm(0,1)$

$$F_Z(z) = \mathbb{P}(Z \le z) = \mathbb{P}((\frac{X-\mu}{\sigma}) \le z) = \mathbb{P}(x \le \mu + \sigma z) = \phi(\frac{(\mu + \sigma z) - \mu}{\sigma}) = \phi(z)$$

$$\Downarrow$$

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

□

Corollary. *Notice that :*

$$VarX = Var(\mu + \sigma Z) = Var(\sigma Z) = \sigma^2$$

proberty of variance. (monotonic)

let X, Y two random variables on (Ω, F, P) S.T $X(\omega) \leq Y(\omega)$. then if both variance exist then :

$$EX \leq EY$$

Using-Theorem:

if $X(\omega) \leq Y(\omega)$ then $F_X(x) \geq F_Y(x)$

Proof. Notice that :

$$F_X(x) = \mathbb{P}(X \leq x) \geq \mathbb{P}(Y \leq x) = F_Y(x)$$

□

proof proberty of expectation. (monitonic)

we will show that $EX \leq EY$ using the tail formula :

$$EX = -\int_{-\infty}^{\infty} F_X(x)dx + \int_{0}^{\infty} (1 - F_X(x))dx \leq \int_{-\infty}^{0} F_Y(x)dx + \int_{0}^{\infty} (1 - F_Y(x))dx = EY$$

JENSEN - INEQUALITY

look at $I \subset \mathbb{R}$(bounded or not, open or closed)

let $h : I \to \mathbb{R}$ be convex function and X random variable which got it's values from I.

if μ_X exist and $Eh(x)$ exist then :

$$Eh(x) \geq h(EX)$$

Reminder :

h convex on I if $h(p_1x_1 + p_2x_2) \leq p_1h(x_1) + p_2h(x_2)$, $\forall p_1, p_2 \geq 0, p_1 + p_2 = 1$ and $x_1, x_2 \in I$.

Example. LOOK AT :

(1) $h(x) = x^2$ is convex on $\mathbb{R}$, and bu jensen inequality we have that $EX^2 \geq (EX)^2$.
(2) $h(x) = x^4$ is convex on $\mathbb{R}$ and by jensen inequality $EX^4 \geq (EX)^4$.

Proof. $\forall x_0 \in I$, $\exists c \in \mathbb{R}$ S.T $h(x) \geq h(x_0) + c(x - x_0)$. $\forall x \in I$.

In particular,

$$h(x) \geq h(\mu_X) + c(x - \mu_X), \forall x \in I$$

i.e we get :

$$h(X(\omega)) \geq h(\mu_X) + c(X(\omega) - \mu_X), \forall \omega \in \Omega$$

.

Using the monotonic proberty we have that ;

$$Eh(X) \geq h(\mu_X) + c \underbrace{E(X - \mu_X)}_{E[X] - \mu_X = 0} \Rightarrow E[h(X)] \geq h(\mu_X) = h(E[X])$$

$\square$

Random Vectors

Definition. let (Ω, F, P) a probability space and $x_1, x_2, .., x_n$ random variables on it.

$X = \vec{X} = (x_1, .., x_n)$ is called **RANDOM VECTOR** with dimention n.

Remark. sometimes we are required to write it $\vec{X} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.

Exercise. let $X \sim Bin(2, \frac{1}{2})$, $Y \sim Bin(2, \frac{1}{2})$ what is $\mathbb{P}(X = Y)$?

Solution. no enough information.

Exercise. $X = \left\{ \begin{matrix} 0, & \frac{1}{2} \\ 1, & \frac{1}{2} \end{matrix} \right\}, Y = \left\{ \begin{matrix} 0, & \frac{1}{2} \\ 1, & \frac{1}{2} \end{matrix} \right\}$

We will fill the probability table..\mathbb{P}_{Y}

$X \backslash Y$	0	1	
0	α	$\frac{1}{2} - \alpha$	$\frac{1}{2}$
1	$\frac{1}{2} - \alpha$	α	$\frac{1}{2}$
	$\frac{1}{2}$	$\frac{1}{2}$	

If yes then $\mathbb{P}(X = Y) = 2\alpha$ but α not given.

Definition. radom vector $(X_1, ..X_n)$ is **discrete if all the** $X_k, k = 1, 2.., n$ are discrete random variables.

it's equivelant to X get only finite number of vectors (finite or counatble).

Definition. let $(X_1, .., X_n)$ random vector then define

$$\mathbb{P}_{(X_1,..X_n)}(X_1, ..., X_n) = \mathbb{P}(X_1 = x_1, ..., X_n = x_n)$$

is called the probability function.

Theorem. *let (X, Y) be a discrete random variable with dimention 2 then,*

- $\mathbb{P}_X(x) = \sum_{y \in \mathbb{R}} \mathbb{P}_{X,Y}(x, y) \backslash$
- $\mathbb{P}_Y(y) = \sum_{x \in \mathbb{R}} \mathbb{P}_{X,Y}(x, y)$

Independent random variables :

we will describe the random varaibles X, Y and the following U, V :

$X\backslash Y$	1	2	3	4	5	6	
1	$\frac{1}{36}$	$\frac{1}{36}$	$\cdots$	$\cdots$	$\cdots$	$\frac{1}{36}$	$\frac{1}{6}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\cdots$	$\cdots$	$\cdots$	$\frac{1}{36}$	$\frac{1}{6}$
3	$\vdots$	$\vdots$	$\ddots$			$\vdots$	$\vdots$
4	$\vdots$	$\vdots$		$\ddots$			$\vdots$
5	$\vdots$	$\vdots$			$\ddots$		$\vdots$
6	$\frac{1}{36}$	$\frac{1}{36}$				$\frac{1}{36}$	$\frac{1}{6}$
	$\frac{1}{6}$	$\frac{1}{6}$	$\cdots$	$\cdots$	$\cdots$	$\frac{1}{6}$	

Notice that

$$\mathbb{P}(X=1,Y=2)=\frac{1}{36}=\mathbb{P}(X=1)\cdot\mathbb{P}(Y=2)=\frac{1}{6}\cdot\frac{1}{6}=\frac{1}{36}$$

So indeed X,Y are independent.

$U\backslash V$	1	2	3	4	5	6	
1	$\frac{1}{12}$	$\frac{1}{60}$	$\cdots$	$\cdots$	$\cdots$	$\frac{1}{60}$	$\frac{1}{6}$
2	$\frac{1}{60}$	$\frac{1}{12}$	$\cdots$	$\cdots$	$\cdots$	$\frac{1}{60}$	$\frac{1}{6}$
3	$\vdots$	$\vdots$	$\ddots$			$\vdots$	$\vdots$
4	$\vdots$	$\vdots$		$\ddots$			
5	$\vdots$	$\vdots$			$\ddots$		
6	$\frac{1}{60}$	$\frac{1}{60}$				$\frac{1}{12}$	$\frac{1}{6}$
	$\frac{1}{6}$	$\frac{1}{6}$	$\cdots$	$\cdots$	$\cdots$	$\frac{1}{6}$	

Notice that

$$\mathbb{P}(U=1,V=2)=\frac{1}{60}\neq\mathbb{P}(U=1)\cdot\mathbb{P}(V=2)=\frac{1}{6}\cdot\frac{1}{6}=\frac{1}{36}.$$

So U,V are not independent.

Definition. let (X,Y) a random vector (Ω,F,P) every event $A\in F$ with form $A=\{\omega|X(\omega)\in B\}$ for $B\in\beta(\mathbb{R})$, is called **generator by X.**

Definition. X,Y are called independent if every pair of event A_1,A_2 when A_1 generaized by X and A_2 generaized by Y are independent.

In general $X_1,...,X_n$ aew independent if all events $\{A_k\}_{k=1}^{n}$ S.T A_k are generaized by X_k are independent.

Claim. the elements if a discrete random vector are independent if :

$$\mathbb{P}(X_1,...,X_n)=\mathbb{P}_{X_1}(x_1)\cdot\mathbb{P}_{X_2}(x_2)\cdots\mathbb{P}_{X_n}(x_n)$$

Proof. Assume that $X_1,...,X_n$ independent then ;

$$*\mathbb{P}(X_1,...,X_n) = \mathbb{P}(\overbrace{X_1 = x_1}^{A_1}, \overbrace{X_2 = x_2}^{A_2}, ..., \overbrace{X_n = x_n}^{A_n}) = \prod_{k=1}^{n} \mathbb{P}_{X_k}(x_k)$$

Now assuming $*$ satisfied then :
$B_2 = \{2,5\}, B_1 = \{1,2,3\}, n = 2$
we will show that $\{\overbrace{X \in B_1}^{A_1}\}, \{\overbrace{Y \in B_2}^{A_2}\}$ are independent.

Indeed,

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}((X \cap Y) \in \{(1,2),(1,5),(2,2),(2,5),(3,2),(3,5)\})$$
$$= \mathbb{P}_{X,Y}(1,2) \cdots \mathbb{P}_{X,Y}(3,5) = \mathbb{P}_X(1)\mathbb{P}_Y(2) + ... + \mathbb{P}_X(3) \cdot \mathbb{P}_Y(5)$$

$$= (\mathbb{P}_X(1) + \mathbb{P}_X(2) + \mathbb{P}_X(3)) \cdot (\mathbb{P}_Y(2) + \mathbb{P}_Y(5)) = \mathbb{P}(X \in \{1,2,3\}) \cdot \mathbb{P}(Y \in \{2,5\})$$
$$\mathbb{P}(A_1) \cdot \mathbb{P}(A_2)$$

Hence, A_1, A_2 are independent. the proof for the generalized case idenetical. □

Corollary. *if* (X,Y) *independent then knowing* $\mathbb{P}_X(x), \mathbb{P}_Y(y)$ *determine* $\mathbb{P}_{X,Y}(x,y)$.

Example. Yulia and samir are doing bernulli trials one after other and in the same time. they are doing it in independent way.

(1) what is the they both get success in the first time together.
(2) what is the probability that till first time samir got success, yulia successed only once.

Solution. (1)

we need that $\mathbb{P}(A) = \sum_{k=1}^{\infty} \mathbb{P}_{X,Y}(k,k)$ when X denote the first success trial for samir and Y denote the first success trial for yulia.

$\mathbb{P}(A) = \sum_{k=1}^{\infty} \mathbb{P}_{X,Y}(k,k) = \mathbb{P}(A) = \sum_{k=1}^{\infty} \mathbb{P}_X(k) \cdot \sum_{k=1}^{\infty} \mathbb{P}_Y(k) = \mathbb{P}(A) = \sum_{k=1}^{\infty} p^2(q^2)^{k-1} = \frac{p^2}{1-q^2} = \frac{p}{1+q}$

Solution. (2).

for the reader.

Hint : $\mathbb{P}(B) = \sum_{k=1}^{\infty} \mathbb{P}(B|C_k) \cdot \mathbb{P}(C_k)$

When C_k denote thee event in which samir success first time in the k trial.

Definition. a random vector $X = (X_1,...,X_n)$ is **absolute continious** if exist integrable function $f_X(x_1,...,x_n) \geq 0$ S.T

$$\mathbb{P}(X \in B) = \int_B \cdots \int f_X(x_1,..x_n)dx_1 \cdots dx_n$$

(B has a boundry(∂) with a measure zero), f_X is called the **density function** f X.

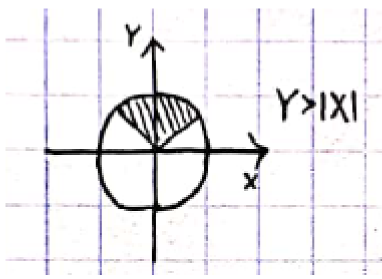
Example. if $D \subseteq \mathbb{R}^n$ a domain with volume n dimentional positive and finite.
$X = (X_1, X_2, ...X_n)$ **uniform in** D if it has density :

$$f_X(X_1, .., X_n) = \left\{ \begin{array}{cc} \frac{1}{|D|} & , x \in D \\ 0 & x \notin D \end{array} \right\}$$

Mark : $X \sim U(D)$
for instance, if D is a half circle in $\mathbb{R}^2$.
(X, Y) uniform random vector in $D, (X, Y) \sim U(D)$.
what is $\mathbb{P}(Y > |X|)$?

Solution. it's a $\frac{1}{4}$ of circle, hence :

$$\mathbb{P}(Y > |X|) = \iint\limits_{D_0} dxdy = \frac{|D_0|}{|D|} = \frac{1}{4}$$



Claim. let (X, Y) random vector with a density $f_{X,Y}(x, y)$ then also for X we have density which given by :

$$f_X(x) = \int\limits_{-\infty}^{\infty} f_{X,Y}(x, y)dy$$

Remark. this calim is true for $f(x, \cdot)$ integrable on $\mathbb{R}, \forall x$ but not for some countable number of x.

Proof. Notice that :

$$F_X(X) = \mathbb{P}(X \le x) = \mathbb{P}((X, Y) \in D_X) = \iint\limits_{D_X} f(u, v)dudv = \int\limits_{-\infty}^{x} (\int\limits_{-\infty}^{\infty} f_{X,Y}(u, v)dv)du = \int\limits_{-\infty}^{\infty} g(u)du$$

So by the definition, $g(\cdot)$ is the density of X. $\square$

QUESTION :

did we assume that $\iint\limits_{\mathbb{R}^2} f_{X,Y}(x, y)dxdy = 1$

Answer :

NO, we assumed that it's only exist, but from the definition.

$$\iint\limits_{\mathbb{R}^2} f_{X,Y}(x,y)dxdy = \mathbb{P}(X,Y)dxdy = \mathbb{P}(X,Y) \in \Omega$$

hence,

$$\mathbb{P}(\Omega) = 1 \Rightarrow integral = 1$$

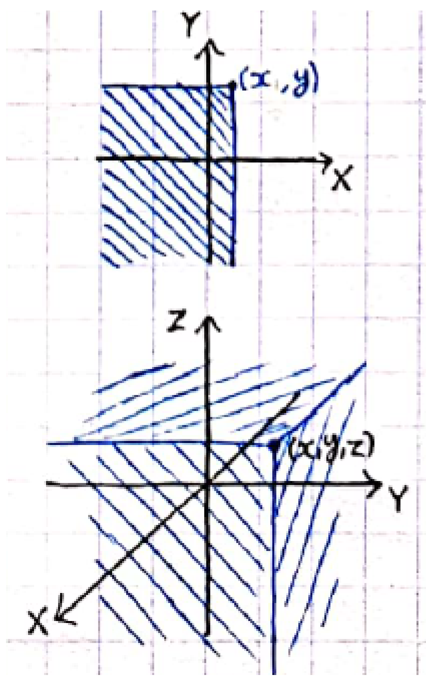
Common Distribute Functions

Definition. let $X = (X_1, X_2, ..., X_n)$ a random vector with $Dim = n$ the fucntion :

$$F_x(x_1, ..., x_n) = \mathbb{P}(X_1 \leq x_1, ..., X_1 \leq x_n) = \mathbb{P}(\bigcap_{k=1}^{n} \{X_k \leq x_k\})$$

is called the **distribution function of** X.

Graphic Intiuition :



Proberties of distribution funtion :

(1) it's not decreasing in every element separately.
(2) $\underset{x_k \to -\infty}{lim} F_X(x_1, .., x_n) = 0, \forall 1 \leq k \leq n.$
(3) $\underset{x_k \to \infty}{lim} F_X(x_1, .., x_n) = F_{X_1,..X_{k-1}}(x_1, .., x_{n-1})$

(4) let (box) $T = \prod_{i=1}^{n} (a_i, b_i)$ then :

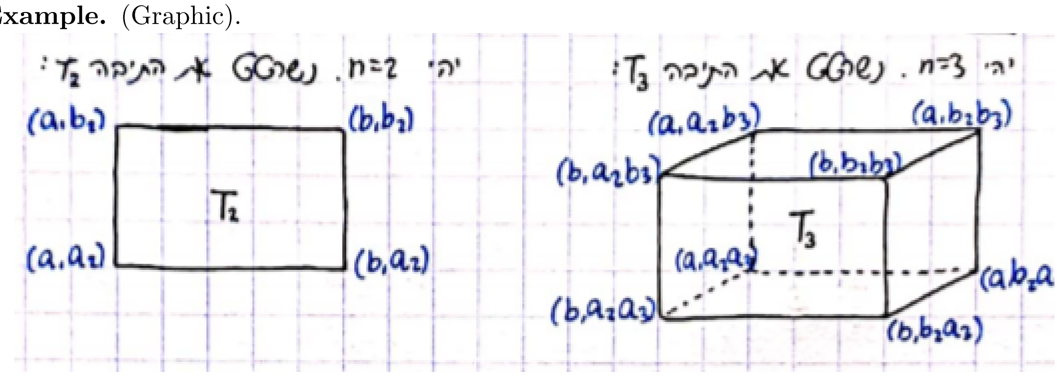
$$\mathbb{P}(X \in T) = \sum_{v \in V} \sigma_V F(v)$$

when V is a set with 2^n verices of the box T and number of a in V $\sigma_V = -1$, $(\sigma_{(a_1,b_2,b_3)} = (-1)^1 = -1$

Remark. the σ_V tell us the sign in which the vertice will appear in the formula which is given by $(-1)^{\#a}$ when $\#a$ is number of $a$in the vector.

Remark. we use that u=in order to find the the probability function.\mathbb{P}_{Y}

Example. (Graphic).

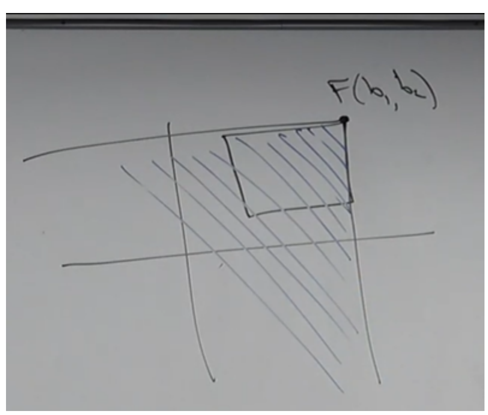


in the left picture we can find $\mathbb{P}((X,Y) \in T)$ uing the formula.

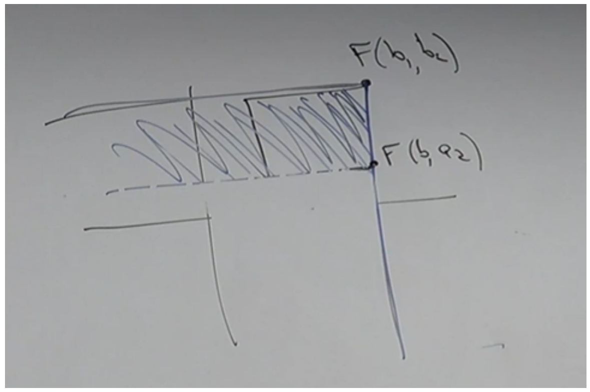
$$\mathbb{P}((X,Y) \in T) = F_{X,Y}(b_1, b_2) - F_{X,Y}(a_1, b_2) - F_{X,Y}(b_1, a_2) + F_{X,Y}(a_1, a_2)$$

Intituiton for proof in case $n = 2$

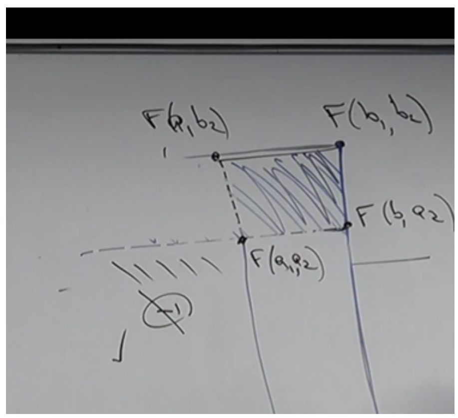
first we notice that in the box, this vertice denote the probability that we fall on all the marked volume



now it's to much because we want that box and not all of this, so we can add "remove extras" by adding other vertice.



Still to much so we remove more by susbstracting with new "vertice" (distribution function) $F_{X,Y}(b_1, a_2)$



NOW notice that we removed the marked volume which marked by ✓ Twice so in order to fix it we add the new "Vertice" which is distribution function in the removed twice area in order to fix the error so we add $F_{X,Y}(a_1, a_2)$

proof proberty 3 :

we are assuming that $F_X(x_1, ..., x_n)$ is monotonic not-decreasing as function of x_n.

it's enough to show that :

$$F_{X_1,..,X_n}(x_1, .., x_n) = \lim_{m\to\infty} F_X(x_1, .., x_{n-1}, m)$$

Check :wr

$$F_X(x_1, .., x_{n-1}, m) = \mathbb{P}(\underbrace{X_1 \le x_1, .., X_{n-1} \le x_{n-1}, X_n \le m}_{C_m}) \underset{m\to\infty}{\to} \mathbb{P}(\bigcup_{m=1}^{\infty} \{X_1 \le x_1, .., X_{n-1} \le x_{n-1}, X_n \le m\})$$

$$= \mathbb{P}(\{X_1 \le x_1, ..., X_{n-1} \le x_{n-1}\}) = F_{X_1,..,X_{n-1}}(x_1, .., x_{n-1})$$

In General:

$$\lim_{\substack{x_1 \to \infty \\ x_2 \to \infty \\ x_3 \to \infty}} F_X(x_1,..,x_n) = F_{X_1,..,X_n}(x_4,..,x_n)$$

(true for every choosing of index).

Theorem. *let $X_1,..,X_n$ independent random variable. if and only if*

$$F_X(x_1,..x_n) = F_{X_1}(x_1)\cdots F_{x_n}(x_n), \forall x_1,..,x_n$$

Proof. we need to show that $\{X_k \in (-\infty, x_k]\}_{k=1}^{n}$ are independent $\forall x_1,..,x_n \Rightarrow$every event $\{X_k \in B_k\}$ independent.

thats true by the construction of borel sets (we can build borel set as union/intersection/complements of the intervals $(-\infty, x_k)$, we will not write the full proof.

Now assume that $(x_1,...,x_n)$ is absolute continious with density $f_{X_1,.X_n}(x_1,..,x_n)$

then giving the density we can create probability function by :

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} (\int_{-\infty}^{y} f_{X,Y}(u,v)dv)du$$

in general,

$$F_X(x_1,..,x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f_X(u_1,...u_n)du_1...du_n$$

□

Theorem. *let X with a density f_X(n dimentional) then in all $x = (x_1,..,x_n)$ in which f_X continious also in it's neighberhoood*

$$f_X(x) = \frac{\partial^n F_X(x_1,..,x_n)}{\partial x_1 \cdots \partial x_n}$$

Remark. it's exactly the inverse of the previous Theorem,

Proof. we will show for $n = 2$ and for $n > 2$ it's identical.

reminder : Leibniz Theorem ("deffrentiaion under the integral sign") if $g(u,v)$ is define on the box $D = [a,b]x[c,d]$ and if $\frac{\partial g}{\partial v}(u,v)$ exists and continious in neighberhood of (x,y) then,

$$G(v) = \int_{a}^{b} g(u,v)du$$

deffrentiable in $v = y$ and in it's neighberhood $G'(y) = \int_{a}^{b} \frac{\partial g}{\partial v}(u,y)du$.

if yes, then denote,

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} (\underbrace{\int_{-\infty}^{y} f_{X,Y}(u,v)dv}_{g(u,y)})du$$

Notice that:

$$\frac{\partial g}{\partial y}(u,v) = f_{X,Y}(u,y)$$

exist and continious.
So by Leibniz,

$$\frac{\partial F}{\partial x}(x,y) = \int_{-\infty}^{y} f_{X,Y}(x,v)dv = g(u,y)$$

So again by Leibniz,

$$\frac{\partial}{\partial y}[\underbrace{\frac{\partial}{\partial x}F_{X,Y}(x,y)}_{g(u,y)}] = f_{X,Y}(x,y)$$

In general n case we can apply Leibniz many times ro get the required. □

Function of Random Vector

Definition. if $X_1,..,X_n$ random vector and $h:\mathbb{R}^n \to \mathbb{R}$ borel funciton i.e

$$\forall B \in \beta(\mathbb{R}), h^{-1}(b) \in \beta(\mathbb{R}^n)$$

then

$$\underbrace{Y}_{Random-Variable} = \underbrace{h(X)}_{Random-Vector}$$

$$\underbrace{Y^{-1}(B)}_{\in \mathbb{R}^1} = (h \circ X)^{-1}(B) = X^{-1}(\underbrace{\underbrace{h^{-1}(B)}_{\in \mathbb{R}^n})}_{\in F}$$

i.e Y is a random variable (it's a function on which the inverse image of B is a event in F).

Example. let $(X,Y) \sim U(Unit-Circle)$ denote the unit circle in D.

$$f_{X,Y}(x,y) = \begin{Bmatrix} \frac{1}{\pi}, & (x,y) \in D \\ 0, & (x,y) \notin D \end{Bmatrix}$$

let $R = \sqrt{X^2+Y^2}$ what is the distribution of R?

Solution. we will find $F_R(r)$,$0 \le r \le 1$

$$F_R(r) = \mathbb{P}(R \le r) = \mathbb{P}((X,Y) \in \underbrace{D_r}_{*}) = \frac{\pi r^2}{\pi 1^2} = r^2$$

*- circle in radius r into the unit circle.
hence,

$$F_R(r) = \left\{ \begin{array}{cc} 0, & r \leq 0 \\ r^2, & 0 \leq r \leq 1 \\ 1, & r > 1 \end{array} \right\}$$

General Case:

let $X = (X_{\text{i}}, .., X_n)$ random vector, $h : \mathbb{R}^n \to \mathbb{R}$ borel function, let $Y = h(x)$,

if X Diacrete then,

$$Eh(X) = \sum_{x \in \mathbb{R}^n} h(x)\mathbb{P}_X(x_1, .., x_n)$$

if X absolute continious then,

$$Eh(X) = \int_{\mathbb{R}^n} \cdots \int f_X(x)dx$$

We can say tht if 1 of side exist then the other exist as well.

Remark. in the absolute continious case we also assume on h that not only $\{x \in \mathbb{R}^n | h(x) > y\} = h^{-1}(y, \infty)$ also it's is a borel set in $\mathbb{R}^n$ on which riemann integral of the continious function exist.

Proof. (discrete case)
$Y = h(x_1, .., x_n), (x_1, ..., x_n)$ discrete random vector$\Rightarrow$ Y also discrete.

$$\mathbb{P}_Y(y) = \mathbb{P}(Y = y) = \mathbb{P}(x \in \{x | h(x) = y\}) = \sum_{\{x|h(x)=y\}} \mathbb{P}_X(x)$$

Now,

$$EY = \sum_{y \in \mathbb{R}} y\mathbb{P}_Y(y) = \sum_{y \in \mathbb{R}} \sum_{\{x|h(x)=y\}} \mathbb{P}_X(x) = \sum_{x \in \mathbb{R}^n} \mathbb{P}_X(x) \sum_{\{y|h(x)=y\}} y = \sum_{x \in \mathbb{R}^n} \mathbb{P}_X(x)h(x)$$

□

Claim. if $x_1, .., x_n$ independent random variables with expectation and if $Y = X_1 \cdot ... \cdot X_n$ is also random variable with expectation then $EY = E[X_1] \cdot E[X_2] \cdots E[X_n]$

Remark. $E(X_1 \cdots X_n) = EX_1 \cdots EX_n \nRightarrow X_1, ..X_n$ independent.

Proof. assume that $(x_1, ...x_n)$ is absolute continious.

$$EY = \int_{\mathbb{R}^n} \cdots \int x_1 \cdots x_n f_X(x_1, .., x_n)dx_1 \cdots dx_n = \int_{\mathbb{R}^n} \cdots \int x_1 \cdots x_n f_{X_1}(x_1) \cdots f_{X_n}(x_n)dx_1 \cdots dx_n$$

$$\int_{-\infty}^{\infty} x_1 f_{X_1}(x_1)dx_1 \cdot \int_{-\infty}^{\infty} x_2 f_{X_2}(x_2)dx_2 \cdots \int_{-\infty}^{\infty} x_n f_{X_n}(x_n)dx_n = EX_1 \cdots EX_2$$

□

Claim. if X_1, X_2 independent with finite expectation then :

$$Var(X_1 + X_2) = VarX_1 + VarX_2$$

Proof. using definition of Var we get that :

$$Var(X_1+X_2) = E(X_1+X_2)^2-[E(X_1+X_2)]^2 =_{linear} EX_1^2+2EX_1X_2+EX_2^2-(EX_1)^2-(EX_2)^2-2EX_1X_2$$
$$= VarX_1 + 2\underbrace{(EX_1EX_2 - EX_1EX_2)}_{=0(X_1,X_2(independent)} + VarX_2 = VarX_1 + VarX_2$$

□

Remark. it was enough to assume that $EX_1X_2 = EX_1EX_2$

Definition. pair of random variables are statistically independent, if they has no-correlation i.e

$$EX_1X_2 = EX_1EX_2$$

Remark. X_1, X_2 are independent $\Rightarrow X_1, X_2$ are statistically independent.
the other direction not necessarily true.

Example. $Y = X^2, X \sim Norm(0,1)$
$$EXY = EX^3 = 0(ODD - FUNCTION)$$

$$EXEY = 0 \cdot 1 = 0$$
i.e X, Y are statistically independent, but they are not independent.

Definition. A random variable X which get only the values $0, 1$ only, is called **indicator.**

Example. we are rolling coin with probability p for tree

$$X = \left\{ \begin{array}{cc} 1, & tree \\ 0, & otherwhise \end{array} \right\}$$
$$EX = P = \mathbb{P}(X = 1)$$
$$Var(X) = EX^2 - (EX)^2 = p - p^2 = p(1-p) = pq$$

Example. person do n bernoulli trials with paramerter success p.

$$X = number - success$$

$$EX = np$$

$$VarX = npq$$

Define n indicators :

$$X_i = \left\{ \begin{array}{cc} 1, & Success - in - trial - i \\ 0, & Failure - in - trial - i \end{array} \right\}$$
$$X = X_1 + .. + X_n$$
moreover $X_1, .., X_n$ are indpenedent so by formula we showed before which says if $X_1, ...X_n$ are independent and $Y = X_1 + X_2...X_n$ so $E[Y] = E[X_1 + X_2...X_n] = E[X_1] + ... + E[X_n]$ hence,

(1) $E[X] =_{linear-of-expectation} \sum_{i=1}^{n} \underbrace{E[X_i]}_{p} = np$

(2) $Var[X] =_{independent} \sum_{i=1}^{n} \underbrace{Var(X)}_{pq} = npq$

Remark. also if the trials are independent but we success all of the with probability p then

$$E[number - success] = np$$

but not necessarily that

$$Var(number - of - success)npq$$

Example. look at

$$X_1 = \left\{ \begin{array}{cc} 1, & p \\ 0, & q \end{array} \right\}$$

$$X_1 = X_2 = ... = X_n$$

$$X_k = \left\{ \begin{array}{cc} 1, & p \\ 0, & q \end{array} \right\}, k = 1, 2....$$

$$S = X_1 + ... + X_n = \left\{ \begin{array}{cc} 0, & p \\ n, & q \end{array} \right\}$$

$$ES = 0 \cdot q + np = np$$

Notice that the expectation will always be np .

$$Var(S) = n^2 p = (np)^2 = n^2 pq \neq npq$$

Example. The is n different vertices in unit circle $V_1, ..., V_n$, between every pair of vertices we connect by edge.

what is the expectation of produces triangle ? (the is potential $\binom{n}{3}$ as most).

N =number of triangles, denote $\underbrace{(i, j, k)}_{triangle}, [1 \leq j \leq k \leq n]$

$$N_{i,j,k} = \left\{ \begin{array}{cc} 1, triangle - appear & (i, j, k) \\ 0, & otherwise \end{array} \right\}$$

$$N = \sum_{1 \leq j \leq k \leq n} N_{i,j,k}$$

$$EN = \sum_{n} \underbrace{EN_{i,j,k}}_{p^3} = \binom{n}{3} p^3$$

Sum of Independent Random Variable

Example. $X \sim Pois(\lambda_X), Y \sim Pois(\lambda_Y)$, X, Y independent.

<u>Denote</u> : $Z = X + Y$

What is the $\mathbb{P}_Z(n), n = 0, 1....$

$$\mathbb{P}_Z(n) = \sum_{k=0}^{n} \mathbb{P}(X = k, Y = n-k) = \sum_{k=0}^{n} \mathbb{P}(X = k)\mathbb{P}(Y = n-k) = \sum_{k=0}^{n} \mathbb{P}_X(k)\mathbb{P}_Y(n-k)$$

$$\sum_{k=0}^{n} \frac{e^{-\lambda_X} \cdot \lambda_X^k}{k!} \cdot \frac{e^{-\lambda_Y} \cdot \lambda_X^{n-k}}{(n-k)!} = \frac{e^{-(\lambda_X+\lambda_Y)} \cdot}{n!} \cdot \underbrace{\sum_{k=0}^{n} \binom{n}{k} \lambda_X^k \cdot \lambda_Y^{n-k}}_{(\lambda_X+\lambda_Y)^n}$$

Corollary. $Z \sim Pois(\lambda_X + \lambda_Y)$

Convolution

Definition. assuming there is two density functions the function

$$h(x) = (f * g)(x) = \int_{-\infty}^{\infty} f(t)g(x-t)dt$$

is called the ***Convolution*** *of f and g.*

Proberties :

$f * g$ is commutative proberty and accosiative i.e

$$f * g = g * f$$

$$(f_1 * f_2) * f_3 = f_1 * (f_2 * f_3)$$

Is $(f_2 * f_3)$ is density funtion?

Answer : yes.

Claim. $(f * g)$ is a density funtion.

$$\int_{-\infty}^{\infty} (f * g)(x)dx = \int_{-\infty}^{\infty} (\int_{-\infty}^{\infty} f(t)g(x-t)dt)dx = \int_{-\infty}^{\infty} (f(t) \underbrace{\int_{-\infty}^{\infty} g(x-t)dx)dt}_{=1(*\underbrace{explain}_{=1})} = 1$$

*Explain : The integral of density function is 1.

Theorem. *let X, Y 2 independent random variables with a denisty f_X, f_Y then*

$$Z = X + Y$$

is absolute continious and it's density given by

$$f_Z(\xi) = (f_X * f_Y)(\xi)$$

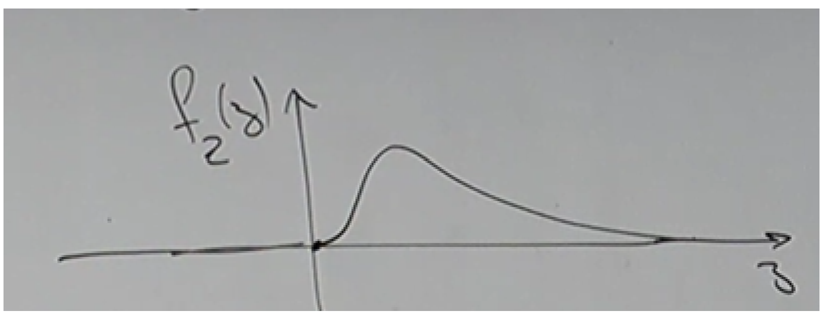
Proof. (Tutorial). □

Example. Assume $X \sim Exp(\lambda)$,$Y \sim Exp(\lambda)$, $Z = X + Y$

$$\xi > 0, f_Z(\xi) = (f_X * f_Y)(\xi) = \int_0^\xi \underbrace{\lambda e^{-\lambda x}}_{f_X(x)} {}_{.x>0} \cdot \underbrace{\lambda e^{-\lambda(\xi - x)}}_{f_Y(\xi - x), \xi - x > 0} dx = \lambda^2 \int_0^\xi e^{-\lambda\xi} dx = \lambda^2 \xi e^{-\lambda\xi}$$

i.e we conclude that :

$$f_Z(\xi) = \left\{ \begin{array}{cc} \lambda^2 \xi e^{-\lambda\xi}, & \xi > 0 \\ 0, & \xi < 0 \end{array} \right\}$$



Reminder :

X a random variable Gamma with paramerters λ, r, $r \in \mathbb{N}, \lambda > 0$ if :

$$f_X(x) = \left\{ \begin{array}{cc} \frac{\lambda^r x^{r-1}}{(r-1)!} e^{-\lambda x}, & x > 0 \\ 0, & x < 0 \end{array} \right\}$$

Notice that $\Gamma(1, \lambda) = Exp(\lambda)$

$$X \sim \Gamma(r, 1)$$

Corollary. *"$Exp(\lambda) * Exp(\lambda) = \Gamma(2, \lambda)$" or we can say that*

$$\underbrace{\Gamma(1, \lambda)}_{Exp(\lambda)} * \Gamma(1, \lambda) = \Gamma(2, \lambda)$$

Claim. for $s, r \in \mathbb{N}$ then:

$$\Gamma(r, \lambda) + \Gamma(r, \lambda) = \Gamma(r + s, \lambda)$$

HOMEWORK: show for $s = 1$ and some r (The claim stem by induction $\forall s$)

Corollary. *$X_1, .., X_n$ a sequence of random variables $Exp(\lambda)$ independent.*

$$S = \sum_{k=1}^{n} X_k, S \sim \Gamma(n, \lambda)$$

Life example :

Assuming there is n busses which arrive to station and the time betweeen busses is $Exp(\lambda)$ S.T the time for arring each one is independent, so we need to wait if want to know in which time bus number n arrive, it's actually the sum of n periods of time i.e

$$Exp(\lambda) + ... + Exp(\lambda) = \Gamma(n, \lambda)$$

By the previous claims.

Example. let $X \sim U[0,1], Y \sim [0,1], Z = X + Y$ hence, $0 < \xi < 2$

$$f_Z(\xi) = \int \underbrace{\mathbf{1}_{[0,1]}(x)}_{0<x<1} \cdot \underbrace{\mathbf{1}_{[0,1]}(\xi - x)}_{0<\xi<x<1 \Rightarrow \xi-1<x<\xi} dx = \int_{max\{0,\xi-1\}}^{min\{1.\xi\}} \mathbf{1}_{[0,1]}(x)\mathbf{1}_{[0,1]}(\xi - x)dx$$

So if $\xi \in (0, 1)$ then :

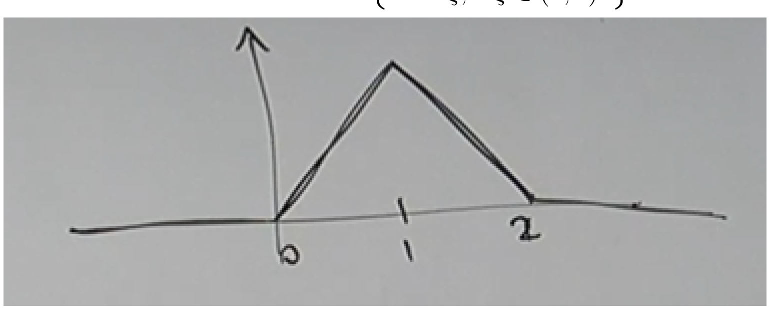
$$f_Z(\xi) = \int_0^{\xi} dx = \xi$$

So if $\xi \in (1, 2)$ then :

$$f_Z(\xi) = \int_{\xi-1}^{1} dx = 2 - \xi$$

So in total we get that :

$$f_Z(\xi) = \left\{ \begin{array}{cc} \xi, & \xi \in (0,1) \\ 2-\xi, & \xi \in (1,2) \end{array} \right\}$$



intiuition : if we think about the discrete case we can think about it as rolling cubes and we roll two cubes in independent way and we are looking for the sum of the two results $2-12$ it's hard to get 2 and 12 Since it's hard to get $1-1=2$ or in symmertic way $6-6=12$ So in the discrete case the graph above (triange) describe that, so it's easie to get 4 or easier to get 7 so we can thinkabout 7 as the max point in the triangle.

Conditional Probability Function

Definition. let (X,Y) a Discrete Random Vector, the funciton (of y)

$$\mathbb{P}_{Y|X}(y|x)=\mathbb{P}(Y=y|X=x)$$

which defined $\forall x$ S.T $\mathbb{P}(X=x)>0$ and $\forall y\in\mathbb{R}$ is called the conditional probability function of Y by the conition X.

Remark. it's a function of y, $\underline{x \text{ is paramater}}$.

Remark. $\mathbb{P}_{X,Y}(y|x)=\frac{\mathbb{P}_{X,Y}(x,y)}{\mathbb{P}_X(x)}$ Since, $\mathbb{P}(Y=y|X=x)=\frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(X=x)}=\frac{\mathbb{P}_{X,Y}(x,y)}{\mathbb{P}_X(x)}$

Example. if $X=$ number of success of N bernulli trials with probability success p when N is random $N\sim Poiss(\lambda)$ then $X\sim Poiss(\lambda,p)$,

$$\mathbb{P}_{X|N}(k|n)=\binom{n}{k}p^kq^{n-k},k=0,1,..,n(for-n-only)$$

$$\mathbb{P}_X(k)=e^{-\lambda p}\frac{(\lambda p)^k}{k!},k=0,1...(till-\infty)$$

Claim. the funciton $\mathbb{P}_{Y|x}(y|x)$ (of y) is probability function.
indeed $\mathbb{P}_{Y|X}(y|x)\geq 0$ and we need to check that $\sum\limits_y\mathbb{P}_{Y|X}(y|x)\underset{=}{?}1$

Notice that :

$$\sum_y\mathbb{P}_{Y|X}(y|x)=\frac{\overbrace{\sum\limits_y\mathbb{P}_{Y|X}(x,y)}^{\mathbb{P}_X(x)}}{\mathbb{P}_X(x)}=1$$

Claim. $\mathbb{P}_Y(y)=\sum\limits_{\substack{x\\ \{x:\mathbb{P}_x(x)\neq 0\}}}\mathbb{P}_{Y|X}(y|x)\mathbb{P}_X(x)$

Proof. Notice that

$$\mathbb{P}_Y(y)=\mathbb{P}(Y=y)=\sum_{x:\mathbb{P}_X(x)\neq 0}\mathbb{P}(Y=y|X=x)\mathbb{P}(X=x)=\sum_{\substack{x\\ \{x:\mathbb{P}_x(x)\neq 0\}}}\mathbb{P}_{Y|X}(y|x)\mathbb{P}_X(x)$$

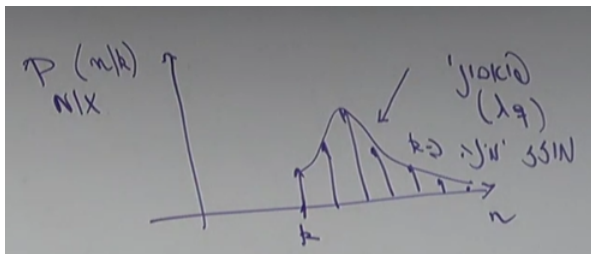
When we used the full probaility formula. ∎

Claim. $\mathbb{P}_Y(x|y) = \frac{\mathbb{P}_{X|Y}(y|x)\mathbb{P}_X(x)}{\mathbb{P}_Y(y)} \forall x, y$ S.T $\mathbb{P}_X(x) > 0, \mathbb{P}_Y(x) > 0$

Proof. (bayes formula- proof like before).

Example. $N \sim Pois(\lambda)$, $X =$ number of success in N bernulli trials.
find $\mathbb{P}_{N|X}(n|k)$

$$\mathbb{P}_{N|X}(n|k) = \frac{\mathbb{P}_{X|N}(k|n)\mathbb{P}_N(n)}{\mathbb{P}_X(k)} = \frac{(\binom{n}{k}p^k q^{n-k})(e^{-\lambda}\frac{\lambda^n}{n!})}{e^{-\lambda p}\frac{(\lambda p)^k}{k!}} = \frac{e^{-\lambda(1-p)} \cdot (\lambda q)^{n-k}}{(n-k)!}, n = k, k+1, ...$$



Definition. let (X, Y) Discrete random vector $\mathbb{P}_X(x) > 0$ then :

$$E(Y|X = x) = \sum_y y\mathbb{P}_{Y|X}(y|x)$$

in assumption that the series well defined i.e (not ∞ or ∞)

In the previous example

(1) $E(X|N = n) = np$
(2) $E(N|X = k) = k + \lambda q$ Since $\mathbb{P}_{N|X}(n|k) = Poiss(\lambda q)$ moved by k so from expectation linearity we get the following.

Definition. let (X, Y) Discrete random vector and $E[Y] < \infty$ (i.e $E[Y]$ exist and finite)

$$E[Y|X] = E[Y|\underbrace{X = x}_{h(x)}]|_{x=X}$$

Remark. $E[Y|X]$ is not $E[Y|X = x]$

Remark. $E[Y|X]$ <u>is a random variable.</u>

Example. (CONTINUE OF PREVIOUS EXAMPLE).
what is $E(X|N)$?

Reminder :

$$E(X|N = n) = np$$

So now by the definition $E(X|N) = Np$ we just plug-in N instead of n as in the definition $|_{x=X}$

Reminder :

$$E(N|X = k) = k + \lambda q$$

So now by the definition we have that :

$$E(N|X) = X + \lambda q$$

why we do this?

Notice that we plug in n assuming $p = 0.5$ in order to find the expectation we plug in every time, for example if number of trials is $n = 12$ we have that $E(X|N = n) = np = 12 \cdot 0.5 = 6$ and for $n = 10$ we get that $E(X|N = n) = np = 10 \cdot 0.5 = 5$ So we get a table of conditional expectation for every value of n so we are tired to say it as a infinite table of every value of n so we summarize it by saying that expectation of X condintioned by number of trials N is number of trials $\cdot p$ So we have the general formula and doesn't depend on which n we plug in into $E(X|N = n)$.

Theorem. *(Law of total expectation)*
let (X, Y) a discrete random vector with $E|Y| < \infty$ (expectation of Y exist and finite) then :

$$EY = E[\underbrace{\quad E(Y|X) \quad}_{random-variable}]$$

Example. (CONTINUE OF LAST EXAMPLE).

Assuming that we didn't do that number of success is $Poiss(\lambda p)$ and we need to know $E[X]$?

$$E[X] = E[E(X|N)] = E[Np] = pE[N] = \lambda p$$

Since $E[N] = \lambda$ (we showed before that for $(X \sim Poiss(\lambda p), EX = \lambda p)$)

<u>Ranodom absolute continious vector</u>

Assuming that $E|Y| < \infty$ we define $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$ for all parameter x in which $f_X(x) > 0$ and $\forall y \in \mathbb{R}$

why we define it like that ?

As the same definition of a discrete random in which we define :

$$\mathbb{P}_{Y|X}(y|x) = \frac{\mathbb{P}_{X,Y}(x,y)}{\mathbb{P}_X(x)}$$

WE WILL CHECK THAT LATER..

Claim. In those assumption :

(1) $f_{Y|X}(y|x)$ (funtion of y) is a density funciton i.e ($\int\limits_{-\infty}^{\infty} = 1 \wedge \geq 0$)

(2) $f_Y(y) = \int\limits_{-\infty}^{\infty} f_{Y|X}(y|x)f_X(x)dx$

(3) $f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(X)}{f_Y(y)}$

Proof. the proof of those requierments are easy and identical to the discrete case. □

We will justify the definition.

The naive idea is that the definition need to be :

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

then we have problem Since $\{X = x\}$ event has probability 0.

Solution to the problem.

$$F_{Y|X}(y|x) = \lim_{\epsilon\to 0}\mathbb{P}(Y \leq y|x-\epsilon < X < x+\epsilon)$$

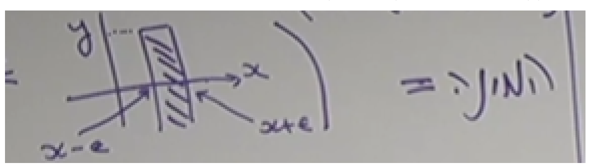
and we define the density as the derivative of F

$$f_{Y|X}(y|x) = \frac{\partial F_{Y|X}(y|x)}{\partial y}$$

We derivative only by y sice we know that x is parameter.

Check :

$$*\mathbb{P}(Y \leq y|x-\epsilon < X < x+\epsilon) = \frac{\mathbb{P}(X \in (x-\epsilon, x+\epsilon), Y \leq y)}{\mathbb{P}(X \in (x-\epsilon, x+\epsilon))}$$

The numerator = 

So we use previous formula $\mathbb{P}(X \in T) = \sum\limits_{v\in V} \sigma_V F(v)$ then :

$$Numerator = F_{X,Y}(x+\epsilon, y) - F_{X,Y}(x-\epsilon, y)$$

hence,

$$* = \frac{F_{X,Y}(x+\epsilon, y) - F_{X,Y}(x-\epsilon, y)}{F_X(x+\epsilon) - F_X(x-\epsilon)}$$

now for $\epsilon \to 0$ we can use L'Hôpital's rule;

$$* \underset{\epsilon\to 0}{\to} \frac{\frac{\partial F_{X,Y}(x,y)}{\partial x} + \frac{\partial F_{X,Y}(x,y)}{\partial x}}{2\frac{\partial F_X(x)}{\partial x}} = \frac{2\frac{\partial F_{X,Y}(x,y)}{\partial x}}{2f_X(x)} = \frac{\frac{\partial F_{X,Y}(x,y)}{\partial x}}{f_X(x)}$$

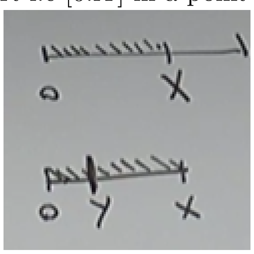
So we get that :

$$f_{Y|X}(y|x) = \frac{\frac{\partial F_{X,Y}(x,y)}{\partial x}}{f_X(x)}$$

$$\Downarrow$$

$$f_{Y|X}(y|x) = \frac{\partial}{\partial y}[\frac{\frac{\partial F_{X,Y}(x,y)}{\partial x}}{f_X(x)}] = \frac{\frac{\partial^2}{\partial x \partial y}F_{X,Y}(x,y)}{f_X(x)} = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

Example. we break a stick $[0,1]$ in uniform point X after that we break the left part i.e $[0.X]$ in a point Y uniformly (like in the photo)



find $f_Y(y)$,$f_{X|Y}(x|y)$?

Solution. we will find first $f_Y(y)$

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

Remember :

$$f_{Y|X}(y|x) = \left\{ \begin{array}{cc} \frac{1}{x}, & 0 < y < x < 1 \\ 0, & otherwise \end{array} \right\}$$

hence,

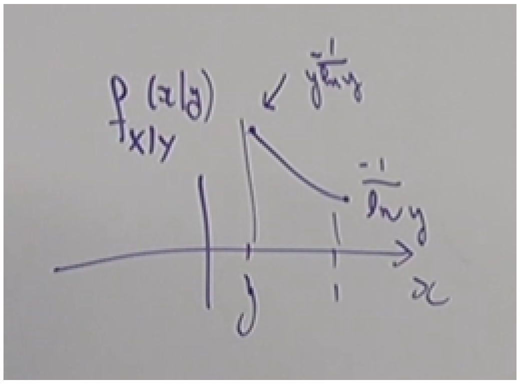
$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x) f_X(x) dx = \int_{y}^{1} \frac{1}{x} dx = -ln(y)$$

Now we find $f_{X|Y}(x|y)$?

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)} = \frac{\frac{1}{x} \cdot 1}{-ln(y)} = -\frac{1}{xln(y)}$$

hence,

$$f_{X|Y}(x|y) = \left\{ \begin{array}{cc} -\frac{1}{xln(y)}, & 0 < y < x < 1 \\ 0, & otherwise \end{array} \right\}$$



Definition. let (X, Y) random absolutely continioud vector $E|Y| < \infty$ the conditional expectation of Y given $X = x$ defined as :

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

Moreover,

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

then we define :

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

(all of this as in the descrete case).

In the previous example what is $E(Y|X = x)$?

$$E(Y|X = x) = \frac{x}{2}$$

(Since we break in the begging at X and the expectation of uniform random variable which is Y in this case i.e $Y \sim U[0, X]$ we know that for $X \sim U[a, b]$ imply $E[X] = \frac{b+a}{2}$ so we got from the first break that $Y \sim [0, X]$ hence, $E[Y] = \frac{0+x}{2} = \frac{x}{2}$).

Now for $E(Y|X)$ as in the discrete case we know that is Randomm varible i.e

we just plug in X instead of X and we get that :

$$E(Y|X) = \frac{X}{2}$$

Theorem. *(Law of total expectation-Absouletly Continious case)*
let (X, Y) absouletly Continious random vector and $E|Y| < \infty$ then :

$$E \underbrace{\quad E[Y|X] \quad}_{random-variable} = E[Y]$$

Example. (In the stick example).

by the definition :

$$EY = \int_0^1 y(-lny)dy = ..(integration - by - parts).. =$$

by the Law of total expectation Theorem :

$$EY = E(E[Y|X]) = E[\frac{x}{2}] = \frac{1}{2}EX = \frac{1}{4}$$

which is very rational that the break of every time would be $\frac{1}{2}$ from the previous if we do the integral we can check that the answer is identical.

Proof. (for the absolutely continious case)

Denote the function $\varphi(y) = E[X|Y = y]$ that how it's defined and we used that $E[X|Y]$ to denote the function $\varphi(Y)$ as in the previous definiotions, it's a rendom variable, Moreover The law of total expectation/law of iterated expectation/tower rule states that, for any random variables X and Y.

$$E[X] = E[E[X|Y]] = \left\{ \begin{array}{cc} \sum_y E[X|Y = y]p_Y(y), & Discrete - Y \\ \int E[X|Y = y]f_Y(y)dy, & Continious - Y \end{array} \right\}$$

$$E[E[X|Y]] = E[\varphi(Y)] = \int_{-\infty}^{\infty} \varphi(y) \cdot f_Y(y)dy = \int_{-\infty}^{\infty} f_Y(y)(\int_{-\infty}^{\infty} xf_{X|Y}(x|y)dx)dy$$

$$= \int_{-\infty}^{\infty} (\int_{-\infty}^{\infty} xf(x,y)dx)dy = \int_{-\infty}^{\infty} x(\int_{-\infty}^{\infty} f(x,y)dy)dx \int_{-\infty}^{\infty} xf_X(x)dx = E[X]$$

□

Claim. Assume that $E|Y| < \infty$.

- $E[aY|X] = aE[Y|X]$
- $E[h(X)Y|X] = h(X)E[Y|X]$ if $E|h(X)y| < \infty)$

This claims yields also for discrete random variable and absolutely continious random variable Moreover the proof is on the website.

Covariance

Introduction.

let $(X_1, .., X_n)$ random vector, we will say that the vector is from order 2 if $EX_k^2 < \infty$, $\forall k = 1, .., n$.

Remember:

$$E|X^2| < \infty \Rightarrow E|X| < \infty$$

Since $|X| \leq 1 + X^2$ (true for $0 \leq X \leq 1$ and $X \geq 1$) hence,

$$|X| \leq 1 + |X^2| \to E|X| \leq 1 + EX^2$$

Definition. let (X, Y) random vector with order 1 the **covariance** between X, Y defined as :

$$\sigma_{X,Y} = cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

Remark. The covariance defined well and finite.$(xy) \leq \frac{x^2+y^2}{2}$ hence,

$$E|(X - \mu_X)(Y - \mu_Y)| \leq \frac{E(X - \mu_X)^2 + E(Y - \mu_Y)^2}{2}$$

Remark. if X, Y independent then :

$$cov(X, Y) = 0$$

Since the $cov(X, Y) = E[(X-\mu_X)E(Y-\mu_Y)]$ now notice that $(X-\mu_X), (Y-\mu_Y)$ are idnependent then :

$$cov(X, Y) = E[(X - \mu_X)E(Y - \mu_Y)] = 0 \cdot 0 = 0$$

Some proberties :

- $cov(X, X) = Var(X)$ i.e $\sigma_{X,X} = \sigma_X^2$
- $cov(aX + b, Y) = a \cdot cov(X, Y)$
- $cov(X_1 + X_2, Y) = cov(X_1 + Y) + cov(X_2, Y)$

In term of probability :

Assuming that $\mu_X = \mu_Y = 0$, $cov(X, Y) = EXY$

Intiuition for cavarince:

if in population we are looking at random variable of X = weight of person in population and the choosing a preson is randomly, Moreover the random variable Y = the height of person in population, then is the $cov(X, Y) > 0 \vee cov(X, Y) < 0$ out intiuition is that the asnwer is $cov(X, Y) > 0$, Since if we assume that $E[Y] = 0.70$ and $E[X] = 78$ then the statistical tendency is the person with weight bigger than the expectation will be also with height bigger than the expactation, so if there is deviation from the expectation is more likely to have deviation of the two values to be more bigger or smaller.

Example. X = height of random person, Y = weight of random person we expect that $cov(X, Y) > 0$

Back to some independent proberties :

Claim. if X, Y are independent random variables then $g(X), h(X)$ are also indpendent for every borel funcitons $g.h$ explicitly, $Y - \mu_Y, X - \mu_X$ are independent.

why the claim is true?

We need to show that if $\{g(X) \in B_1\}, \{h(Y) \in B_2\}$ are independent events for all pair of borel sets B_1, B_2 **but** the event

$$\{g(X) \in B_1\} = (g \circ X)^{-1}(B_1) = (X^{-1} \circ g^{-1})(B_1) = X^{-1}(\underbrace{\quad g^{-1}(B_1) \quad}_{Borel-set(\in \beta(\mathbb{R}^1)})$$

$$= event - by - X = \{X \in g^{-1}(B_1))$$

identical way we get that $\{h(Y) \in B_2\} = \{Y \in h^{-1}(B_2)\}$ event produced by Y and we can conclude that $\{g(X) \in B_1\}, \{h(Y) \in B_2\}$are independent Since the first event produced by X and the second by $Y \Rightarrow h(Y), g(X)$ are independent random variables.

Important proberty :

Define $V = \{$ all the random variable on (Ω, F, P) with expectation $0\}$, V is vector space with dimentions ∞.

Claim. $\forall X, Y \in V$ satisfy :

$$< X, Y >= EXY = cov(X, Y)$$

is inner product.

Remark. it's obvious that

$$< X, Y >=< Y, X >$$

The operation with X separately and in Y separetely.

Remark. $< X, X >= VarX \geq 0$

Remark. if $< X, X >= 0 \Rightarrow VarX = 0 \Rightarrow X$ Constant $\Rightarrow X = 0$ in probability 1.

Corollary. *for all (X, Y) with order 2 we have that ;*

$$|con(X, Y)| \leq \sigma_X \sigma_Y$$

Proof. inequality Caushy-Shwartz which say that

$$| < \tilde{X}, \tilde{Y} > | \leq < \tilde{X}, \tilde{X} >^{\frac{1}{2}} < \tilde{Y}, \tilde{Y} >^{\frac{1}{2}}$$

then we have that :

$$| < \tilde{X}, \tilde{Y} > | = |Cov(X, Y)| \leq < \tilde{X}, \tilde{X} >^{\frac{1}{2}} < \tilde{Y}, \tilde{Y} >^{\frac{1}{2}} = \sigma_X \cdot \sigma_Y$$

when :

$$\tilde{X} = X - \mu_X \in V, \tilde{Y} = Y - \mu_Y \in V$$

□

Reminder :

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

Claim. $cov(X, Y) = E[XY] - E[X]E[Y]$.

Proof. Notice that :

$$cov(X, Y) = E(X - \mu_X)(Y - \mu_Y) = E[XY] - \mu_X E[Y] - \mu_Y E[X] + \mu_X \mu_Y$$
$$= E[XY] - E[X]E[Y]$$

Remark : $\mu_X \mu_Y = E[X]E[Y], \mu_X E[Y] = E[X]E[Y], \mu_Y E[X] = E[Y]E[X]$ □

Corollary. *if X, Y independent then $cov(X, Y) = 0$ Since,* $E[XY] = E[X]E[Y]$

Definition. if (X,Y) random vector with order 2, X,Y statistically independent if $cov(X,Y)=0$ i.e $E[XY]=E[X]E[Y]$

Remark. X,Y are statistically independent.$\underset{\not\Leftarrow}{\Rightarrow}$ X,Y are independent.

Reminder :

$$|cov(X,Y)| \leq \sigma_X \sigma_Y$$

Remark. $Cov(aX,bY)=abCov(X,Y)$ we can't know the relation since a,b can be very very small and it depend in which scalar we are talking about so we can solve it by defining coorelation which give us intiuition about the realtion between those random variables.

Definition. The (corellation) between X,Y defined as

$$p_{X,Y}=\frac{cov(X,Y)}{\sigma_X \cdot \sigma_Y}$$

Here we assume the (X,Y) is with order 2 and X,Y are not trivial i.e $\sigma_X \neq 0, \sigma_Y \neq 0$.

By Caushy-Schwartz ($|cov(X,Y)| \leq \sigma_X \cdot \sigma_Y$ make sure that $|p_{X,Y}| \leq 1$ and $-1 \leq p_{X,Y} \leq 1$

Example. $N \sim Poiss(\lambda), X \sim Bin(N,p) \to \mathbb{P}_{X|N}(k|n)=\binom{n}{k}p^k q^{n-k}, k=0,1....$

Notice that $X \sim Bin(N,p) \Rightarrow X \sim pois(\lambda p)$ find $p_{N,X}=?$

Possible answer :

$$p_{N,X}>0$$

Since our iniuition that if there is success and number of trials than the expectation from the inverse , little success with a lot of trials or vice versa.

Now we will calculate $p_{N,X}$.

$$p_{N,X}=\frac{E[NX]-E[N]E[X]}{\sigma_N \sigma_X}=\frac{E[NX]-\lambda(\lambda p)}{\sqrt{\lambda}\cdot\sqrt{\lambda p}}$$

Now we will use Law of total expectation Theorem in order to find $E[NX]$:

$$E[NX]=EE[N|X|N]=ENE[X|N]=EN\underbrace{E[X|N]}_{Np}=pEN^2=p(\sigma_N^2+(EN)^2)$$

$$p_{N,X}=\frac{p\lambda+p\lambda^2-p\lambda^2}{\lambda\sqrt{p}}=\sqrt{p}$$

<u>The Covariance Matrix Of Random Vector</u> :

Definition. let $X=(X_1,...,X_n)$ Random vector with order 2 the matrix $\sum_X \in \mathbb{R}^{nxn}$ which defined as :

$$(\sum_X)_{i,j}=cov(X_i,X_j)$$

Is called the Covariance matrix of X.

$$\sum_X = \begin{pmatrix} \sigma_{X_1,X_1} & \cdots & \sigma_{X_1,X_n} \\ \vdots & \ddots & \vdots \\ \sigma_{X_n,X_1} & \cdots & \sigma_{X_n,X_n} \end{pmatrix}$$

it's a symmetric matrix i.e $cov(X,Y) = cov(Y,X)$ and it's defined not negative.

Claim. let X random vector with order 2 with $dim = n$,$\vec{a} \in \mathbb{R}^n$

$$Var(\underbrace{a_1X_1 + .. + anX_n}_{a^TX}) = \sum_{i,j=1}^{n} (\sum_X)_{i,j} a_i a_j$$

Proof. Notice that :

$$Var(a^TX) = cov(a_1X_1 + ...a_nX_n, a_1X_1 + ...a_nX_n) = \sum_{i,j=1}^{n} \underbrace{Cov(X_i, X_j)}_{(\sum_X)_{i,j}} a_i a_j$$

□

Corollary. $\sum\limits_{i,j=1}^{n} (\sum_X)_{i,j} a_i a_j$ *is always non-negative the fact stem from that* $(Var > 0)$ *i.e* $(\sum_X)_{i,j}$ *is always non-nagative.*

Theorem. *if* $M \in \mathbb{R}^{nxn}$ *symmetric defined positive (The Quadratic form*<0 $\forall(a_1, ..., a_n) \neq 0$*) if and only if :*

(1) $\lambda_1, ..., \lambda_n > 0$
(2) All the main minors of M matrix are positive.

Proberties of Covariance Matrix :

(1) $\sum_X$ is symmetric defined non-Negative.
(2) $\sum_{X+v} = \sum_X$ Since $cov(X_i + v_i, X_j + v_j) = cov(X_i, X_j), \forall v_i, v_j$.
(3) $\forall M \in \mathbb{R}^{nxn}$, $\underbrace{\sum_{MX}}_{\in \mathbb{R}^{mxm}} = \underbrace{M}_{mxn} \underbrace{\sum_X}_{nxn} \underbrace{M^t}_{nxm}$

Proof. (3).
the $\sum_X$ matrix we can define it as :

$$\sum_X = E(X - \mu_X)(X - \mu_X)^T$$

when :

$$\mu_X = \begin{pmatrix} \mu_{X_1} \\ \vdots \\ \mu_{X_n} \end{pmatrix}$$

So we have :

$$\sum_X = E(X-\mu_X)(X-\mu_X)^T = E\begin{pmatrix} X-\mu_{X_1} \\ \vdots \\ X-\mu_{X_n}\end{pmatrix}\begin{pmatrix} X-\mu_{X_1} & \cdots & X-\mu_{X_n}\end{pmatrix}$$

$$= E\begin{pmatrix} (X_1-\mu_1)^2 & \cdots & (X_1-\mu_1)(X_n-\mu_n) \\ \ldots & \ddots & \vdots \\ (X_n-\mu_n)(X_1-\mu_1) & \cdots & (X_n-\mu_n)^2 \end{pmatrix}$$

So we got that Covariance matrix.

Now in order to show the following Notice that :

$$\sum_{MX} = EM(X-\mu_X)(M(X-\mu_X))^T = EM(X-\mu_X)(X-\mu_X)^T M^T$$

$$= ME[(X-\mu_X)(X-\mu_X)^T]M^T = M\sum_X M^T$$

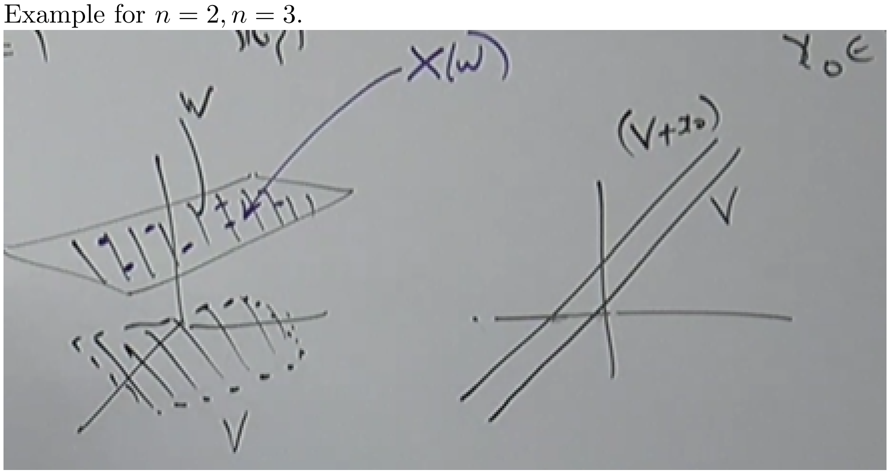
□

Claim. if X Random vector with order 2 and $\sum_X$ is not defined positive i.e exist $0 \neq a \in \mathbb{R}^n$ S.T $a^T \sum_X a = 0$ thats equivalent to $(det \sum_X = 0))$ then exist affine space in $W \subset \mathbb{R}^n$ S.T $\mathbb{P}(X \in W) = 1, dimW < n$.

Definition. Affine space W is a subset of $\mathbb{R}^n$ in form $W = V + x_0$ when V is a vector linear subspace and $x_0 \in \mathbb{R}^n$.

Remark. we can think about it as a subset of $\mathbb{R}^n$ when we say $\mathbb{P}(X \in W) = 1$ i.e in all that space we can think about is as trivial random variable that get always 1 .

Example for $n = 2, n = 3$.



Proof. assume that $a \sum_X a^T = 0 \vee Var(a^T X) = 0$ for $0 \neq a \in \mathbb{R}^n \Rightarrow c \in \mathbb{R}$ S.T $\mathbb{P}(a^T X = C) = 1 \Rightarrow \mathbb{P}(X \in \underbrace{\{x \in \mathbb{R}^n : a^T x = x\}}_{affine-set}) = 1$ It's a exercise to check that it's indeed affine set. □

Gaussian Vector

(1) Standard Gaussian Vector : Random Vector Z with dimention n is a random vector with density.

$$x \in \mathbb{R}^n . f_X(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(x_1^2 + \ldots + x_n^{2)}} = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}|x|^2}$$

in other words $Z_1, .., Z_n$are independent vectors and all of the distribute $N(0, 1)$.

$$f_Z(\zeta) = \frac{1}{(2\pi)^{\frac{\pi}{2}}} e^{-\frac{1}{2}\zeta^T I_n \zeta}, \zeta \in \mathbb{R}^n$$

(2) Gaussian vector with expectation 0 and density : Random Gaussian vector with expectation 0 and density is a random vector X which has representation $X = MZ$ when Z standard gaussian vector with $dim = n$ amd $M \in \mathbb{R}^{nxn}$ is invertible matrix.

Remark. if z standard gaussian

$$\mu_Z = \begin{pmatrix} \mu_{Z_1} \\ \vdots \\ \mu_{Z_n} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \sum_Z = I_n = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Proberties of $\mu_X = MZ$

(1) $\mu_X = \mu_{MZ} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ Since $X_1 = m_{11}X_1 + ... + m_{1n}Z_n \Rightarrow E[X_1] = m_{11}E[Z_1] + ... + m_{1n}E[Z_n] = 0.$

(2) $\sum_X = M \underbrace{\sum_Z}_{I_n} M^T = \underbrace{MM^T}_{defined-positive}$

(3) for X we have density and it's given by :

$$f_X(x) = \frac{1}{(2\pi)\sqrt{\underbrace{|det \sum_X|}_{\neq 0}}} \underbrace{e^{-\frac{1}{2}x^T(MM^T)^{-1}x}}_{e^{-\frac{1}{2}x^t \sum_X^{-1} x}}$$

Example. for $n = 1$ we have :

standard $\rightarrow \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$.

general $\rightarrow \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{x^2}{\sigma^2}}$ notice that $\frac{1}{\sigma^2}$ here is $\sum_X^{-1}$.

Theorem. *let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ injective transformation with $\overline{\text{jacobian non zero}}$ and let X random vector with density $f_X(x)$ then for the random vector $\overline{T = T(X)}$ also have density and it's given by :*

$$f_Y(y) = |det \frac{\partial(x_1, .., x_n)}{\partial(y_1, ..., y_n)}| f_X(T^{-1}(y))$$

$\forall y \in Img(T)$ *and* 0 *otherwise.*

$$\frac{\partial(x_1, .., x_n)}{\partial(y_1, ..., y_n)} = (\frac{\partial x_i}{\partial y_j})_{i,j=1}^n$$

In our case $X = MZ$ s.t $T(\zeta) = \underset{\in \mathbb{R}^{nxn}}{M} \underset{\in \mathbb{R}^n}{\zeta}$ So $T^{-1}(x) = M^{-1}x$

$$\frac{\partial(x_1, .., x_n)}{\partial(y_1, ..., y_n)} = (M^{-1})^T$$

Since $X_1 = m_{11}^{-1}\zeta_1 + m_{12}^{-1}\zeta_2 + ... + m_{1n}^{-1}\zeta_n$ when

$$\frac{\partial x_1}{\partial z_i} = (M^{-1})_{1i}$$

$$det\frac{\partial(x_1, .., x_n)}{\partial(y_1, ..., y_n)} = \frac{1}{detM}$$

Son in the Theorem we can see that :

$$f_X(x) = \frac{1}{(2\pi)\underbrace{|detM|}_{\neq 0}} e^{-\frac{1}{2}|T^{-1}x|^2}$$

$$|T^{-1}(x)|^2 = |M^{-1}x|^2 =_{norm-of-vector} (M^{-1}x)^T(M^{-1}x) = x^T(M^{-1})^T(M^{-1})x$$

$$= x^T(MM^T)^{-1}x$$

hence,

$$f_X(x) = \frac{1}{(2\pi)\underbrace{|detM|}_{\neq 0}} e^{-x^T(MM^T)^{-1}x} = \frac{1}{(2\pi)\underbrace{\sqrt{|det\sum_X|}}_{\neq 0}} e^{-x^T(\sum_X)^{-1}x}$$

Proof. Notice that :

$$\mathbb{P}(Y \in \underbrace{B}_{Borel-set-in-\mathbb{R}^n}) = \mathbb{P}(T(X) \in B) = \mathbb{P}(X \in T^{-1}(B)) = \int\limits_{T^{-1}(B)} f_X(x)dx$$

Now denote $y = T(x)$ So we have :

$$= \int\limits_B \underbrace{|det\frac{\partial(x_1, .., x_n)}{\partial(y_1, ..., y_n)}|f_X(T^{-1}(y))}_{g(y) = f_Y(y)}dy$$

□

Proberties of Gaussian vector with density and expectation 0 (continue).

4. if $X = MZ$ like before then every subvector of X is also Gaussia vector with density and expectation 0 explicitly every element is also like this (scalar).

Proof. only for $n = 2$.

Assume that $f_{X,T}(x, y) = Ce^{-\frac{1}{2}(ax^2+2bxy+cy^2)}$ when a quadratic form is defined positive :

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}, a > 0.ac - b^2 > 0$$

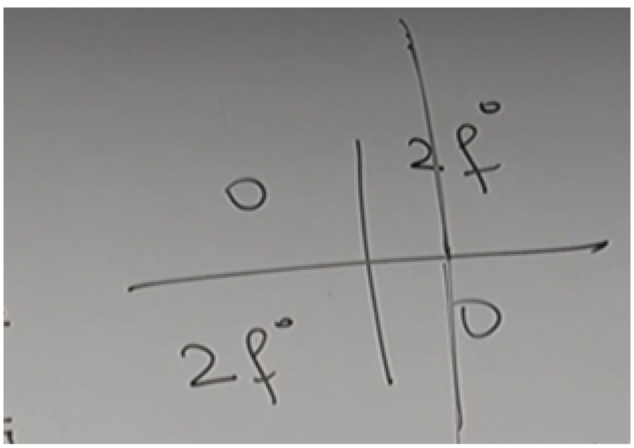
$$f_Y(y) = C \underbrace{\int_{-\infty}^{\infty} e^{-\frac{a}{2}(x+\frac{b}{a}y)^2} dx}_{C'} \cdot e^{-\frac{1}{2}y^2(c-\frac{b}{a}y)^2} = C' e^{-\frac{1}{2}\frac{y^2}{(\frac{a}{ac-b^2})^2}}$$

Notice that $C' e^{-\frac{1}{2}\frac{y^2}{(c-\frac{b}{a}y)^2}}$ is exactly the gaussian i.e $(\frac{a}{ac-b^2})^2 = \sigma^2$ □

Remark. if (X, Y) is gaussian vector with density $\Rightarrow X$ Gaussian and Y Gaussian but the inverse is not right.

Example. As in the photo Denote :

$$f_{X,Y}(x, y) = \frac{e^{-\frac{1}{2}(x^2+y^2)}}{2\pi}$$



We notice thats it's not Gaussian density

$$f_X(x) = \int_0^{\infty} 2f^0(x, y)dy = \int_{-\infty}^{\infty} f^0(x, y)dy = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

This density not Gaussain but the marginal density are Gaussian.

Claim. if (X, Y) gasussian vector (expectation 0 and density) then if X, Y statistically independent i.e $(E[XY] = E[X]E[Y]) \Rightarrow X, Y$ are independent.

Proof. if (X, Y) statistically independent $\Rightarrow \sum_{X,Y} = \begin{pmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_Y^2 \end{pmatrix} \Rightarrow \sum_{X,Y}^{-1} = \begin{pmatrix} \frac{1}{\sigma_X^2} & 0 \\ 0 & \frac{1}{\sigma_Y^2} \end{pmatrix}$ is also diagonal matrix hence,

$$f_{X,Y}(x, y) = Ce^{-\frac{1}{2}(x,y)^T \sum_{X,Y}^{-1}(x,y)} = Ce^{-\frac{1}{2}(\frac{x^2}{\sigma_X^2}+\frac{y^2}{\sigma_Y^2})} = Ce^{-\frac{1}{2}(\frac{x^2}{\sigma_X^2})} e^{-\frac{1}{2}(\frac{y^2}{\sigma_Y^2})}, \forall x, y$$

Hence, X, Y are independent. □

General Question :

if $f_{X,Y}(x, y) = g(x)h(y)$ are X, Y independent? yes but not true 100%

(In this case $g(x) = f_X(x), h(y) = f_Y(y)$).
what if $f_{X,Y}(x, y) = g7(x)\frac{1}{7}h(y)$?
Answer: X, Y are indepedent and $\exists\alpha > 0$ S.T

$$f_X(x) = \alpha g(x), f_Y(y) = \frac{1}{\alpha}g(y)$$

Proberty :

if $X = MZ$ a gaussian vector with expectation 0 and $Y = AX$ when $A \in \mathbb{R}^{mxn}$ $(m \leq n)$ then Y is a gaussian random vector.

Proof. We will split the probert into cases :
Case 1 :
$m = n \Rightarrow Y = AMZ$ when $AM \in \mathbb{R}^{nxn}$
Case 2 :
if $m \leq n$ then define $\tilde{A} \in \mathbb{R}^{nxn}$ which defined as :

$$\tilde{A} = \begin{pmatrix} A_{mxn} \\ A'_{(n-m)xn} \end{pmatrix} \Rightarrow \tilde{Y} = \tilde{A}Z = \begin{pmatrix} Y_{mx1} = AZ \\ Y'_{(n-m)x1} = A'Z \end{pmatrix}_{nx1} (vector)$$

Notice that $\tilde{Y}$Gaussian from case 1 and Y is a subvector of gaussian vector hence, Y is gaussian itself (in dim=n). □

Remark. if for X has density and $A \in \mathbb{R}^{mxn}$ with degree m ($degA \leq min(m, n) = m$), here we assume that $degA = m$ hence, $Y = AX$ has density Since we can choose A' to be matrix with m independent rows from A till we get row space with dim=n i.e $\tilde{A}$ is invertible matrix hence $\tilde{Y} = \tilde{A}Z$ is a gaussian vector in $\mathbb{R}^n$ with a density $\Rightarrow Y$, as a sunvector of $\tilde{Y}$ with density also has density.

Private Case :
if X has a density $0 \neq a = (a_1, .., a_n) \in \mathbb{R}^n$ then $Y = (a_1X_1, ... + a_nX_n)$ is a random variable with density $X \sim N(0, \underbrace{\sigma^2}_{>0})$ Since in this case $m = 1$ $A = (a_1, .., a_n) \in \mathbb{R}^{1xn} \Rightarrow Y$ has density.

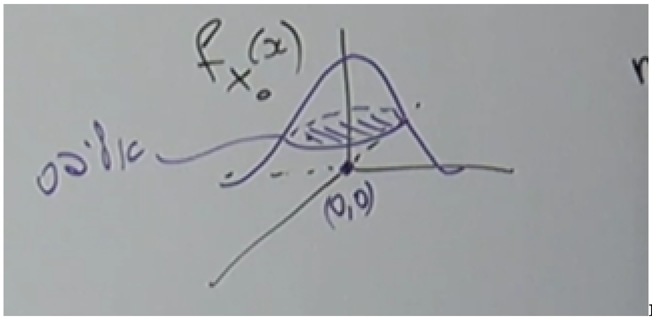
Remark. if $Y = AZ$ when $A \in \mathbb{R}^{mxn}$ and $deg(A) < m$ then Y has no density it stem from the fact that :

$$\sum_Y = A \sum_X A^T$$

is not defined positive $deg(\sum_Y) < m \Rightarrow Y$ is in affine space with dim $< m$.

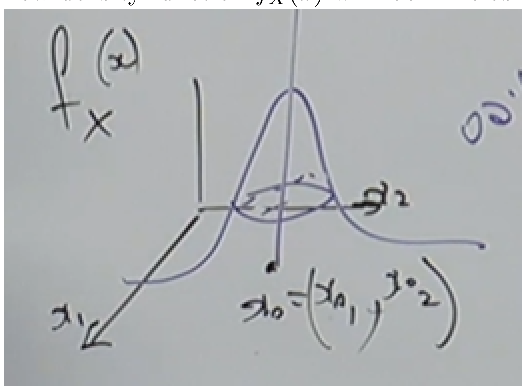
Definition. A general gaussain random vector (dim=n) is a random vector $X = X_0 + x_0$ when X_0 is a gaussian random vector with expectation 0 and $x_0 \in \mathbb{R}^n$.

Example. for $n = 2$ and $X_0 \in \mathbb{R}^n$



notice we have Ellipse

now density function $f_X(x)$ will look like as :



The density of X when ($\sum_X$ invertible) is given by $f_X(x) = Ce^{-\frac{1}{2}(x-x_0)^T(\sum_X^{-1})(x-x_0)}$
We change x in $(x - x_0)$ as in case $n = 1$.

$$f_X(x) = Ce^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

All the previous proberties true also if the expactation vector μ_X is not 0 vector.

$$\begin{pmatrix} \mu_{X_1} \\ \vdots \\ \mu_{X_n} \end{pmatrix} = \begin{pmatrix} X_{0_1} \\ \vdots \\ X_{0_n} \end{pmatrix}$$

(1) Subvector of Gaussian vector is also Gaussian.
(2) Gaussian vector with statistically independent elements it's elements independent.
(3) if X Gaussian vector then AX is gaussian vector $\forall A \in \mathbb{R}^{mxn}, m \leq n$.

Exercise. Given $f_{X,Y,X}(x, y, z) = Ce^{-\frac{1}{2}(x^2+5y^2+\zeta^2-4y\zeta+2y)}$ what is $f_{U,V}(u, v)$ when $U = X + Y, V = Y + Z$.

(x, y, z) is a gaussian vector (Dimension 3) $\sum_{X,Y,Z}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & -2 \\ 0 & -2 & 1 \end{pmatrix} \Rightarrow$
Notice that we have two blocks on diagonal so the invertible matrix $\sum_{X,Y,X} =$

$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 5 \end{pmatrix}$ Check that :

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & -2 \\ 0 & -2 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 5 \end{pmatrix} = I_3$$

Now we need to find $\sum_{U,V}$ Notice that

$$\begin{pmatrix} U \\ V \end{pmatrix} = \overbrace{\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}}^{M}\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

hence we get that :

$$\sum_{U,V} = M \sum_{X,Y,Z} M^T$$

in the tutorial we will learn how to find $\begin{pmatrix} \mu_X \\ \mu_Y \\ \mu_Z \end{pmatrix}$ so we can find $\begin{pmatrix} \mu_U \\ \mu_V \end{pmatrix} = \begin{pmatrix} \mu_X + \mu_Y \\ \mu_Y + \mu_Z \end{pmatrix}$

Corollary. *We can find the Gaussian density $f_{U,V}(u,v)$ Since we have* $\begin{pmatrix} \mu_U \\ \mu_V \end{pmatrix}, \sum_{U,V}$.

Chebyshev's and markov inequality

Question :

What is the probability of "Big Deviation" ("the "tail probability").

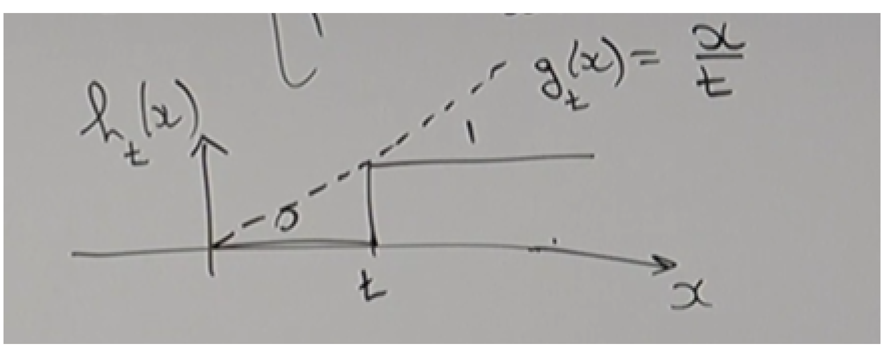
Markov Inequality :

Let X random variable with finite expectation of $|X|$ i.e $(E[X] < \infty)$ then $\forall t > 0$

$$\mathbb{P}(X > t) \leq \frac{E[X]}{t}$$

Proof. Denote $\forall t > 0$ the function $h_t : [0,\infty) \to [0,\infty)$ S.T

$$h_t(x) = \left\{ \begin{array}{ll} 0, & x < t \\ 1, & x \geq t \end{array} \right\}$$



Notice that $h_t(x) \leq g_t(x)$ hence $E[\underbrace{h_t(x)}_{indicator}] \leq E[g_t(x)]$ ($h_t(x)$ is indicator since it's random variable which get onlt $0, 1$ values) hence,

$$\mathbb{P}(X \geq t) \leq E(\frac{X}{t}) =_* \frac{E[X]}{t}$$

Where in $*$ we used the expectation linear proberty. □

Corollary. *let $\phi : [0, \infty) \to [0, \infty)$ strictly incresing function and X random variable S.T $E[\phi(x)] < \infty$ then*

$$\mathbb{P}(|X| > t) \leq \frac{E(\phi(x))}{\phi(t)}$$

Remark. $\mathbb{P}(|X| \geq t) \leq \frac{E(\phi(|X|)}{\phi(t)}$

Proof. Notice that :

$$\mathbb{P}(|X| > t) = \mathbb{P}(\underbrace{\phi(|X|)}_{\tilde{X} \geq 0} \geq \phi(t)) \leq_{makov} \frac{E[\tilde{X}]}{\phi(t)} = \frac{E|\phi(X)|}{\phi(t)}$$

□

Private Case :
$\phi(x) = (X - \mu_X)^2$

Chebyshev's inequality :

Let X random variable with order 2 with a finite expectation μ_X and finite variance σ_X^2 then :

$$\mathbb{P}(|X - \mu_X| \geq t) \leq \frac{\sigma_X^2}{t^2}$$

Proof. Denote $Y = X - \mu_X$ and define $\phi(y) = y^2, y > 0$ then by narkov inequality :

$$\mathbb{P}(|Y| \geq t) \leq \frac{EY^2}{t^2} = \mathbb{P}(|X - \mu_X| \geq t) \leq \frac{EY^2}{t^2}$$

□

Sequence of infinite events

Let (Ω, F, P) is probability space, and $\{A_n\}$ a sequence of events in it.

Definition. $\overline{lim} A_n = \bigcap\limits_{k=1}^{\infty} \bigcup\limits_{n=k}^{\infty} A_n \to$ the event for all ω S.T $\omega \in A_n$ for ∞ of n

Definition. $\underline{lim} A_n = \bigcup\limits_{k=1}^{\infty} \bigcap\limits_{n=k}^{\infty} A_n \to$ the event for all the ω S.T $\omega \in A_n$ for $n \geq n_0$

Proberties :

(1) $\underline{lim} A_n \subset \overline{lim} A_n$
(2) $(\overline{lim} A_n)^c = \underline{lim} A_n^c$
(3) $(\underline{lim} A_n)^c = \overline{lim} A_n^c$

Lemma. *(Borel cantelli 1).*

Assume that there is $\sum\limits_{n=1}^{\infty}\mathbb{P}(A_n)<\infty$ *then* $\mathbb{P}(\overline{lim}A_n)=0$.

Proof. Notice that :

$$\mathbb{P}(\overline{lim}A_n)=\mathbb{P}(\bigcap_{k=1}^{\infty}\underbrace{\bigcup_{n\geq k}^{\infty}A_n}_{B_n})=\lim_{k\to\infty}\mathbb{P}(B_k)=\lim_{k\to\infty}\mathbb{P}(\bigcup_{n\geq k}A_n)\leq\lim_{k\to\infty}\sum_{k=n}^{\infty}\mathbb{P}(A_n)=_{*}0$$

Where in * we used the fact the it;s a tail of converges series. □

Remark. We used the fact that $B_n$is non-decreasing.

Lemma. *(Borel cantelli 2).*

if A_n*as before and independent and if* $\sum\limits_{n=1}^{\infty}\mathbb{P}(A_n)=\infty$ *then* $\mathbb{P}(\overline{lim}A_n)=1$ *explicitly if* A_n*independent then :*

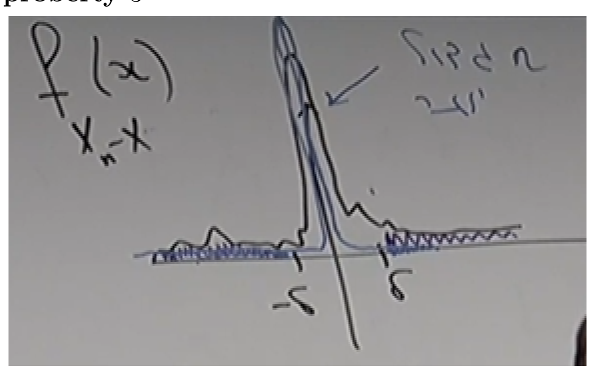
$$\mathbb{P}(\overline{lim}A_n)=\left\{\begin{array}{c}0\\1\end{array}\right\}$$

Convergence of random variables

Definition. let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables defined on (Ω,F,P) and X is other random variable on (Ω,F,P) then :

- $X_n\to X$ almost surely (a.s) if $\mathbb{P}(X_n(\omega)\to X(\omega))=1$
- if $\forall p>0$ we say that $X_n\to X$ in L^p if $E|X_n|^p<\infty,E|X|^p<\infty$ (i.e there is moment of order p for all random variables) and $\lim\limits_{n\to\infty}E(|X_n-X|^p)=0$
- $\lim\limits_{n\to\infty}X_n=X$ **in probability** if $\forall\delta>0$ we have that $\lim\limits_{n\to\infty}\mathbb{P}(|X_n-X|>\delta)=0$

Intiuition for proberty 3

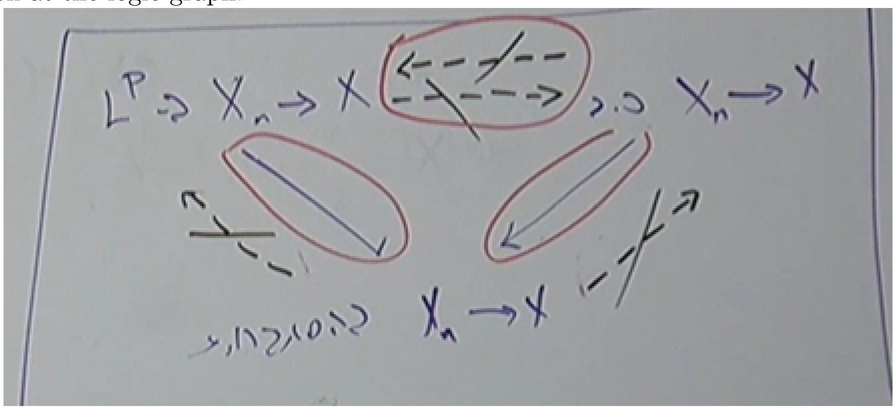


Remark. In the second proberty notice that $E(|X_n-X|^p)<\infty$ the fact stem from that

$$(|X_n-X|^p)\leq(|X_n|+|X|)^P\leq C(|X_n|+|X|)$$

Theorem. ***if*** $X_n\to X$ ***a.s then*** $X_n\to X$ ***in probability and if*** $X_n\to X$ ***in*** L^p ***then*** $X_n\to X$***.***

Remark. if we assume that $X_n \to X$ in probability $\to X_n \to X$ a.s then we get contiadaiction since if $X_n \to X$ a.s then $X_n \to X$ in L^p and it's not true we can look at the logic graph.



Proof. We will show that if $X_n \to X$ a.s (almost surely) then $X_n \to X$ in probability.

Assume thhat $X_n \to X$ a.s then $\forall \delta > 0$

$$0 = \mathbb{P}(|X_n - X| > \delta(i.o)) = \mathbb{P}(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} \{|X_n - X| > \delta\}) = \lim_{k\to\infty} \mathbb{P}(\bigcup_{n=k}^{\infty} \{|X_n - X| > \delta\})$$

$$\geq \lim_{k\to\infty} \mathbb{P}(\{|X_n - X| > \delta\})$$

i.e we have that $\lim_{k\to\infty} \mathbb{P}(\{|X_n - X| > \delta\}) = 0$ and it's exactly the definition of $X_n \to X$ in probability.

Since $(|X_n - X|)$ it's not bigger then delta in infinitly times so at some place it will always be less hence, we have probability 0

$i.o$ mean infinitly often.

Now we will show that $X_n \to X$ in L^p then $X_n \to X$ in probability.

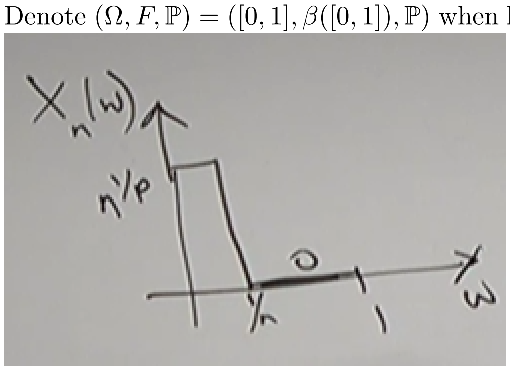
$\forall \delta > 0$ then :

$$\mathbb{P}(|X_n - X| > \delta) =_{markov} \mathbb{P}(|X_n - X|^p > \delta^p) \leq \frac{E[|X_n - X|^p}{\delta^p} \underset{n\to\infty}{\to} \frac{0}{\delta^p} = 0$$

i.e $X_n \to X$ in probability. $\square$

Remark. if $X_n \to X$ a.s $\not\Rightarrow X_n \to X$ in L^p. Counterexample :

Denote $(\Omega, F, \mathbb{P}) = ([0,1], \beta([0,1]), \mathbb{P})$ when $\mathbb{P}([a,b]) = b - a$ (Lebesgue measure).



$$X_n(\omega) = \left\{ \begin{array}{ll} n^{\frac{1}{p}} & 0 < \omega < \frac{1}{n} \\ 0 & \frac{1}{n} < \omega \end{array} \right\}$$

$X_n(\omega) \to 0, \forall \omega > 0$ So $X_n \to \underbrace{X}_{0}$ a.s **BUT**

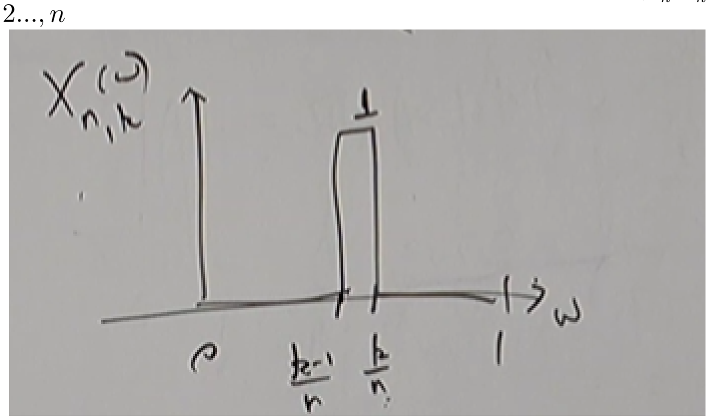
$$|X_n - X|^p - \left\{ \begin{array}{ll} n, & \omega < \frac{1}{n} \\ 0, & \omega > \frac{1}{n} \end{array} \right\}$$

$$|X_n - \underbrace{X}_{0}|^p = n(\frac{1}{n}) + 0 \cdot (1 - \frac{1}{n}) = 1 \underset{n\to\infty}{\nrightarrow} 0$$

i.e $X_n \to X$ in L^p.

Remark. if $X_n \to X$ in $L^p \nRightarrow X_n \to X$ a.s, Counterexample :

The same probability space $([0,1], F, \mathbb{P})$, $X_{n,k} = \mathbf{1}_{(\frac{k-1}{n}, \frac{k}{n})}(\omega)$, $n \in \mathbb{N}, k = 1, 2..., n$



$*$ $X_n \to 0$ in L^p.

$$E|X_{n,k}|^p = 1^p \cdot \frac{1}{n} \underset{(n,k)\to\infty}{\to} 0$$

BUT $X_n \to 0$ a.s

Now notice that $\{X_n(\omega)\} = 1, 0, 1, 010, ... \nrightarrow 0$ Since we heve infinite 1, i.e

$$X_{n,k}(\omega) \nrightarrow 0$$

(The sqeunce will always include 1 for some k)

Law Of Large Numbers

Assume that $\{X_n\}$ a sequence of random variablrs on $(\Omega, F, \mathbb{P})$ are **i.i.d** (independent, identically distributed) assuming that $E[X^2] < \infty$ and denote $E[X] = \mu$, $Var[X] = \sigma^2$ define :

$$S_n = \sum_{k=1}^{n} X_k$$

$$"the - running - avg" = \frac{S_n}{n}$$

$$E\frac{S_n}{n} = \mu$$

$$Var\frac{S_n}{n} = \frac{1}{n^2}Var(S_n) = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

The Weak Law Of Large Numbers :

$\frac{S_n}{n} \to \mu$ in probability i.e given $\{X_n\}_{n=1}^{\infty}$ i.i.d random variables $Var(X_1) = \sigma^2 < \infty$,$E[X_1] = \mu \in \mathbb{R}$then $\underset{n\to\infty}{lim} \frac{S_n}{n} = \mu$ in probability i.e $\forall\delta > 0$ $\underset{n\to\infty}{lim} \mathbb{P}(|\frac{S_n}{n} - \mu| > \delta) = 0$

The Strong Law Of Large Numbers :

$\frac{S_n}{n} \to \mu$ a.s (almost surely) i.e given $\{X_n\}_{n=1}^{\infty} i.i.d$, $E[X_1] = \mu \in \mathbb{R}$ then :

$\underset{n\to\infty}{lim} \frac{S_n}{n} = \mu$ a.s .

Remark. The law is true without the assumption that $\sigma^2 < \infty$ but we will show it with the assumption.

Proof. (**The Weak Law Of Large Numbers**).
let $\delta > 0$

$$\mathbb{P}(|\frac{S_n}{n} - \underbrace{\mu}_{E\frac{S_n}{n}} | > \delta) \leq_{Chebyshev's} \frac{Var\frac{S_n}{n}}{\delta^2} =$$

We know that :

$$Var\frac{S_n}{n} = \frac{1}{n^2}Var(S_n) = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

hence,

$$* = \frac{\sigma^2}{n\delta^2} \underset{n\to\infty}{\to} 0$$

□

Proof. **(The Strong Law Of Large Numbers**). □

Remark. the proof in the tutorial assume additional assumptions $\forall n, E[X_n^4] \leq m < \infty$ (without assume that $|X_n|$ has identical distribution) but $\forall n, EX_n = \mu$.

Exercise. let $\{X_n\}, i.i.d$ a sequence of random variable and assume that $E[X_1] < \infty$.

Claim. $\underset{n\to\infty}{lim} \frac{X_n}{n} = 0$ a.s.

(Background : if X_n get a value in the bounded interval $[a, b]$ in probability 1, $X_n \sim U[-8, 200]$

$$\frac{a}{n} \leq \frac{X_n}{n} \leq \frac{b}{n}(a.s) \Rightarrow \frac{X_n}{n} \to 0$$

otherwise the true condition is $E[X_1] < \infty$.

Solution. First a remark from calculus 1 .

A sequence a_n not converges to 0 $\iff$ $\forall\epsilon > 0, (|a_n| > \epsilon, i.o), \exists\epsilon > 0, \exists\epsilon = \frac{1}{m}, m \in \mathbb{N}$.

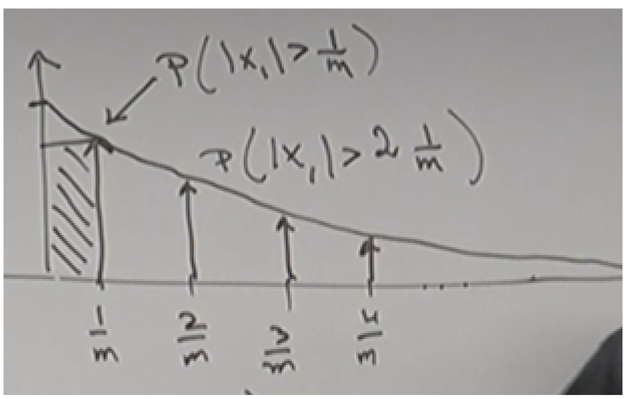
We need to show that :

$$\underbrace{\mathbb{P}(|\frac{X_n}{n}| > \frac{1}{m}(i,o), \exists m)}_{\frac{X_n}{n} \nrightarrow 0} = 0$$

We will show for all m ,$\mathbb{P}(\frac{|X_n|}{n} > \frac{1}{m}, (i.o)) = 0 = \mathbb{P}(\overline{lim}\{|\frac{X_n}{n}| > \frac{1}{m}\})$.

By Borel-Cantelli (1) we need to show that :

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > n \cdot \frac{1}{m}) < \infty$$



$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > n \cdot \frac{1}{m}) \leq \int_0^{\infty} (1 - F_{|X_1|}(x))dx = mE[X_1] < \infty$$

QUESTION :

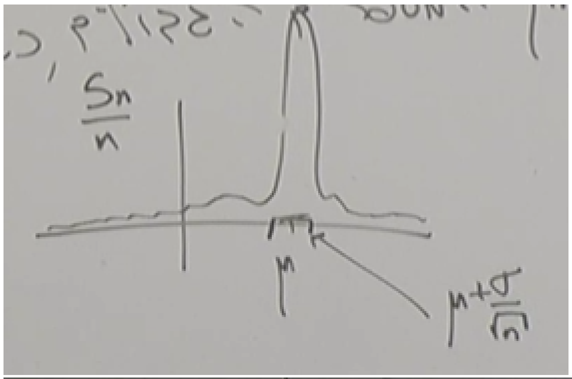
Why $\{|\frac{X_n}{n}| > \epsilon, (i.o)\} = \overline{lim}\{\underbrace{|\frac{X_n}{n}| > \epsilon}_{A_n}\}$ Because, $\underbrace{\overline{lim} A_n}_{\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k} = \{\omega; \omega \in A_n$ for ∞ valuse of n$\}$.

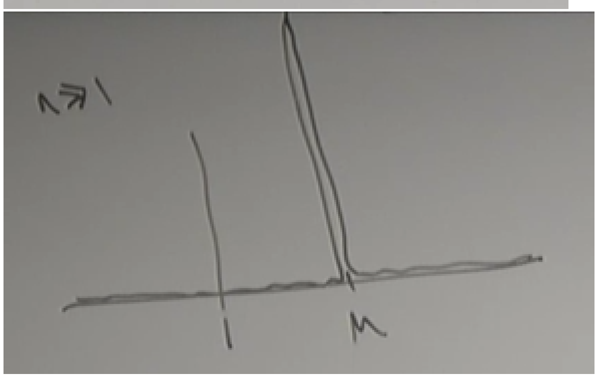
Central Limit Theorem

Theorem. *(Introduction).*

$\mu = \mu_{X_1} \in \mathbb{R}$ *and* $0 < \sigma_X^2 < \infty$ *(*X_1 *is not trivial).*

then by The Strong law of large numbers when n big enough .

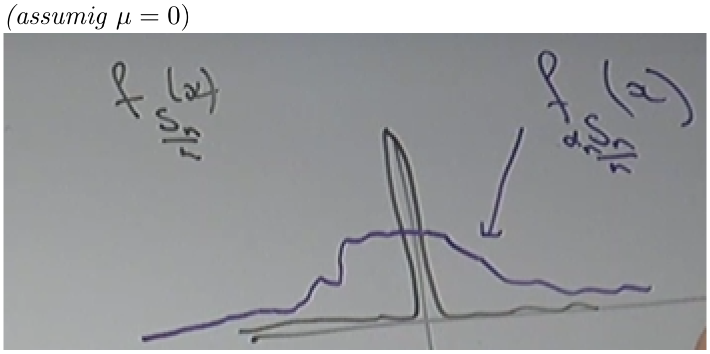




What should be α_n in order to imply

$$Var[\alpha_n \frac{S_n}{n}] = 1$$

(assumig $\mu = 0$)



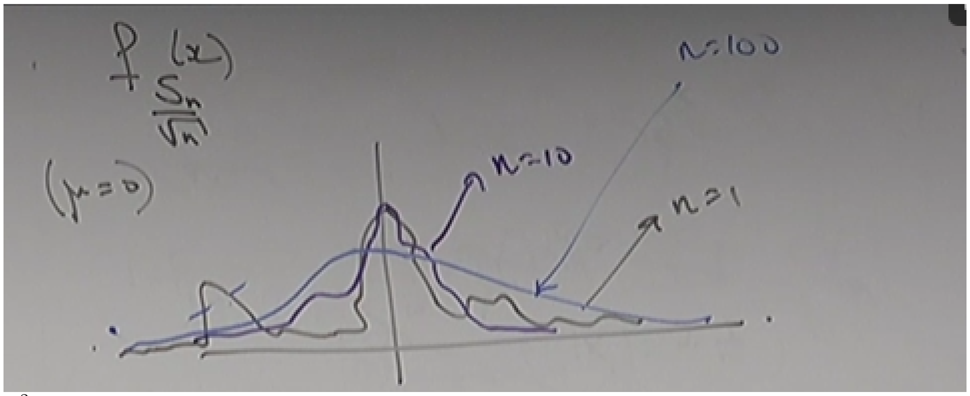
$$Var(\alpha_n \cdot \frac{S_n}{n}) = 1 = \alpha_n^2 \cdot \underbrace{Var(\frac{S_n}{n})}_{\frac{\sigma^2}{n}} = \frac{\alpha_n^2}{n}\sigma^2$$

if we define $\alpha_n = \sqrt{n}$ then $\forall n, Var(\alpha_n \frac{S_n}{n}) = \sigma^2$.
In total *when $(\mu = 0)$ we are intrested in :*

$$\frac{S_n}{\sqrt{n}}, n \in \mathbb{N}$$

Moreover,

$$E[\frac{S_n}{\sqrt{n}}] = 0, Var[\frac{S_n}{\sqrt{n}}] = \sigma^2, \forall n$$



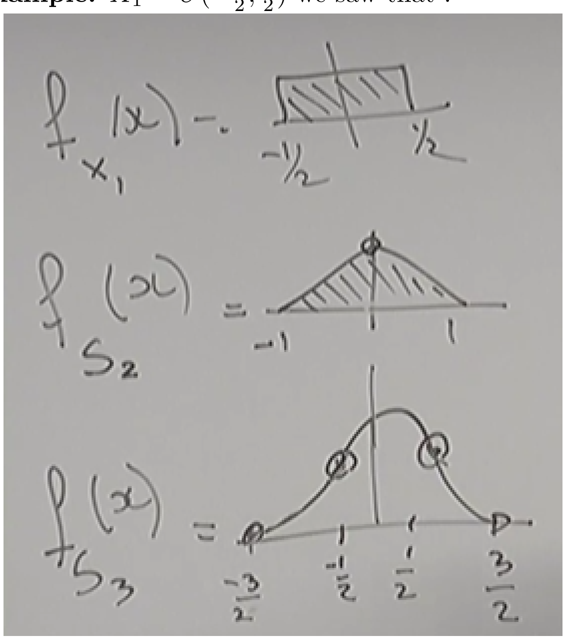
$\sigma^2_{\frac{S_n}{\sqrt{n}}} = \sigma^2$ *is constant.*

Theorem. *(The Central Limit Theorem).*
(assumig $\sigma = 1, \mu = 0$*)*

$$\lim_{n\to\infty} F_{\frac{S_n}{\sqrt{n}}}(x) = \phi(x)$$

When $\phi(x)$ *is a distribution of* $N(0,1)$

Example. $X_1 \sim U(-\frac{1}{2}, \frac{1}{2})$ we saw that :



Intituition : the CTL theorem tell us that for any family of distribution in the end it will be like the gaussian distribution for n good enough.

Reminder :

$\{X_k\}_{k\in\mathbb{N}}$ of random variables on $(\Omega, F, \mathbb{P})$ this sequence is i.i.d assume that X_1 from order 2, $EX_1 = \mu \in \mathbb{R}$, $Var(X_1) = \sigma^2 \in (0, \infty)$ and $S_n = \sum_{k=1}^{n} X_k$.

(the running avg): $\frac{S_n}{n}$ $Var\frac{S_n}{n} = \frac{\sigma^2}{n}, E\frac{S_n}{n} = \mu$.

Theorem. *(CTL)*

(1) if $\mu = 0, \sigma = 1$ then

$$\forall x \in \mathbb{R}, \lim_{n\to\infty} (\frac{S_n}{\sqrt{n}} \le x) = \lim_{n\to\infty} F_{\frac{S_n}{\sqrt{n}}}(x) = \phi(x)$$

(2) if $\mu \in \mathbb{R}$, $\sigma^2 \in (0, \infty)$ general then

$$\forall x, \lim_{n\to\infty} \mathbb{P}(\frac{S_n - n\mu}{\sqrt{n}\sigma} \le x) = \phi(x)$$

Remark. Notice that $1 \to 2$ and $2 \to 1$ Since we denote $\tilde{X_k} = \frac{X_k - \mu}{\sigma}$ (we make it standard distribution) So

$$E[\tilde{X_k}] = E[\frac{X_k - \mu}{\sigma}] = \frac{1}{\sigma}E[X_k - \mu] = \frac{1}{\sigma}(E[X_k] - \mu) = 0$$

$$Var(\tilde{X_k}) = Var(\frac{X_k}{\sigma}) = \frac{1}{\sigma^2} \cdot Var(X_k) = \frac{1}{\sigma^2} \cdot \sigma^2 = 1$$

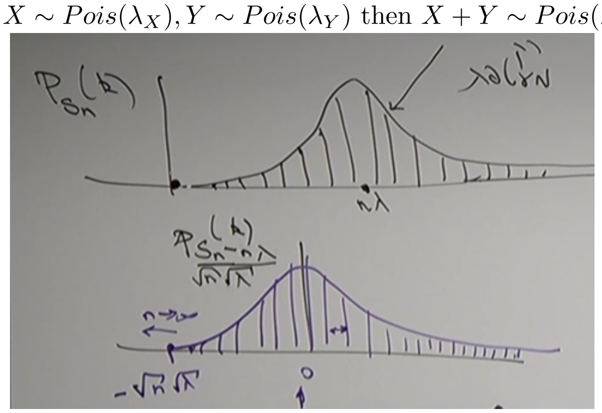
$$\tilde{S_n} = \sum_{k=1}^{n} \tilde{X_k} = \frac{S_n - n\mu}{\sigma}$$

From 1 we know that :

$$\lim_{n\to\infty} (\frac{\tilde{S_n}}{\sqrt{n}} \le x) \underset{n\to\infty}{\to} \phi(x) \Rightarrow \lim_{n\to\infty} \mathbb{P}(\frac{S_n - n\mu}{\sqrt{n}\sigma} \le x) = \phi(x)$$

X_k are independent and distribute identecally.
In the proof of CTL it's enough to show (1).

Example. $X_k \sim Pois(\lambda)$, $S_n \sim Pois(n\lambda)$(we saw before that if we have independent random variables distrubute $Pois(\lambda)$ then the sum disribute $Pois(n, \lambda)$) Since if $X \sim Pois(\lambda_X), Y \sim Pois(\lambda_Y)$ then $X + Y \sim Pois(\lambda_X + \lambda_Y)$.



We normalized the first graph in the second and we can see that the expectation $= 0$ and Var=1

Definition. a sequence of random variables U_n converges in distribution to $Z \sim N(0,1)$ if $\forall x$, $\underset{n\to\infty}{lim} F_{U_n}(x) = \phi(x)$ and we will write $U_n \underset{\to}{D} Z$ (sometime we will write it as $U_n \underset{\to}{D} \phi$) if yes, then CLT in this tyrmonology $\underset{n\to\infty}{lim}(\frac{S_n - n\mu}{\sqrt{n}\sigma}) \underset{=}{D} \phi$

Remark. X_n are not necessarily defined in the same probability space for the definition $X_n \to X \sim N(0,1)$ in probability Since we need that $F_{X_n}(x) \to \phi$.

Definition. We will say $X_n \underset{\to}{D} X$ in distribution (X not necessiarly $N(0,1)$) if

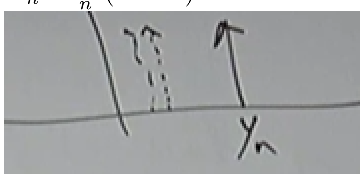
$$\underset{n\to\infty}{lim} F_{X_n}(x) = F_X$$

for all x in which F_X continious.

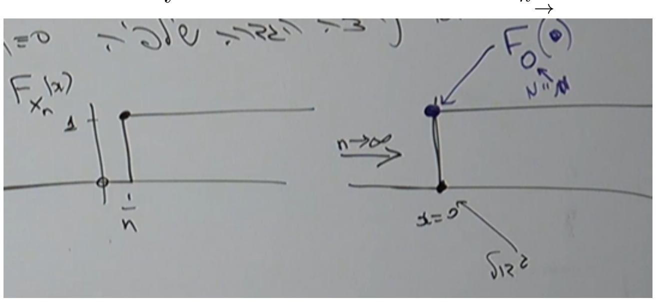
Since $\phi(x)$ continious in every x it's not neccessary to add reservation in case $X \sim N(0,1)$.

Example. (trivial).

$X_n = \frac{1}{n}$ (trivial)



We absouletely need a definition in which $X_n \underset{\to}{D} X = 0$



Theorem. *Given a sequence of random variables* $\{X_n\}_{n\in\mathbb{N}}$ *and* X.
then the two requierments are equivalent.

(1) $X_n \underset{\to}{D} X$
(2) $\underset{n\to\infty}{lim} E[e^{itX_n}] = E[e^{itX}]$

In CLT $(\mu = 0, \sigma = 1)$ we need to show that $\frac{S_n}{\sqrt{n}} \underset{\to}{D} Z \sim N(0,1)$ we will show that :

$$E[e^{it\frac{S_n}{\sqrt{n}}}] \underset{n\to\infty}{\to} E[e^{itZ}], \forall t$$

What is $E[\underbrace{e^{itX}}_{Complex-Random-variable}]$ **?**

$e^{itX} = costX + isintX$ **(sum of of random variables)**

$$EcosX_n \to EcostX$$

$$EsinX_n \to EsintX$$

We will define for every random variable the affine function of X:

$$\varphi(\underset{\in\mathbb{R}}{x}) = Ee^{itX} \in \mathbb{C}$$

Now we will translate requierment (2).

$$\varphi_{\frac{S_n}{n}}(t) \underset{n\to\infty}{\to} \varphi_Z(t), Z \sim N(0,1)$$

(showing that is equivalent to show the convergency in distribution).

Challenging Questions

Tutorial 1:

Exercise. In a cards box there is 52 cards, we take randomly 3 cards one after other :

(1) what is the probability tha we choose exactly 1 king ?
(2) what is the probability that we choose at least one king?

Solution. (1).

Denote Ω a colloection of threes with different inputs, in which every input is random number $1-52$ noitce that :

$$|\Omega| = 52 \cdot 51 \cdot 50$$

Denote A the event that we got exactly one king, $A \subseteq \Omega$.

$A =$ a collection of different threes, which imply one input from it is constant (king).

$$\mathbb{P}(A) = \mathbb{P}(king, something, something) + \mathbb{P}(sometihg, king, something) + \mathbb{P}(something, something, king)$$

$$= \frac{4}{52}\cdot\frac{48}{51}\cdot\frac{47}{50} + \frac{48}{52}\cdot\frac{4}{51}\cdot\frac{47}{50} + \frac{48}{52}\cdot\frac{47}{51}\cdot\frac{4}{50} = 3\cdot\frac{4\cdot 48\cdot 47}{52\cdot 51\cdot 50} = \frac{1128}{5525} \approx 20\%$$

Solution. (1-other-way).

Denote $\Omega = \{\text{subsets from size 3 out of 52 elements}\}$, $|\Omega| = \binom{52}{3}$.

$A = \{\text{all the subsets from order 3 S.T exist one king (cont number from } 1-4)\}$

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{\overbrace{options-choose-which-king}^{4} \cdot \overbrace{options-to-add-for-king}^{\binom{48}{2}}}{\binom{52}{3}} \approx 20\%$$

Solution. (2).

Denote B the event in which the event in which at least 1 king taken i.e B obtian inside it the two events the first one in which exactly 1-king, exactly 2-king, exactly 3-king :

$$\mathbb{P}(B^c) = \mathbb{P}(no{-}king{-}taken) = \frac{\overbrace{3 - cards - out - any - king - taken}^{\binom{48}{3}}}{\binom{52}{3}} = \frac{\frac{48!}{3!\cdot 45!}}{\frac{52!}{3!\cdot 49!}} = \frac{48\cdot 47\cdot 46}{52\cdot 51\cdot 50}$$

hence,

$$\mathbb{P}(B) = 1 - \mathbb{P}(B^c)$$

Exercise. Two persons deal to meet and they arrive in a random time and in independent way between hours $17:00-18:00$ and who arrive first wait 20min, and left if the other didn't arrive.

what is the probability that they meet :

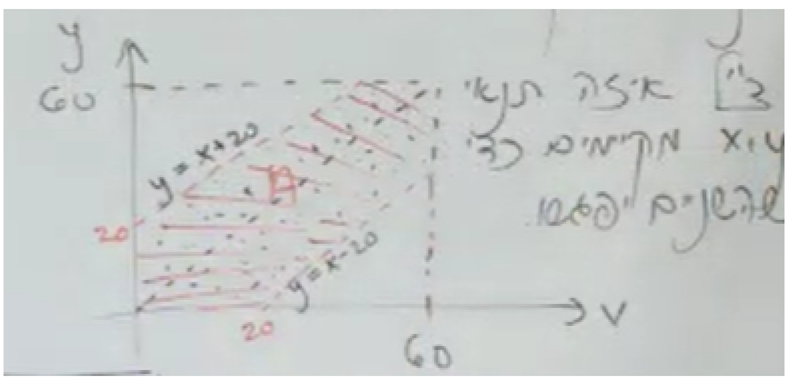
Solution. Denote $\Omega = \{(x,y) \in [0,60]^2\}$

x- arriving time of first person.

y- arriving time of seccond person.

The condition that the both meet is

$$y \le x \le y + 20 \vee x \le y \le x + 20 \iff |x - y| \le 20$$



So this condition represent this Volume from the box $[0,60]^2$.

Denote the volume in the graph by A hence,

$$\mathbb{P}(meet) = \mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{60^2 - 40^2}{60^2} = \frac{5}{9}$$

Tutorial 2 :

Exercise. a book store do in every day give a present randomly to a customer from customers who reach everyday, givent that the probability that k come to the store in some day is $\frac{1}{2^k}, k \in \mathbb{N}$ what is the probability that a person come to the store win the present:

Solution. defiene the event $A_j = \{j$ costumers went to the store $\}$, $B = \{$win the present$\}$

$$\mathbb{P}(B) = \sum_{j\in\mathbb{N}} \mathbb{P}(B|A_j)\mathbb{P}(A_j) = \sum_{j\in\mathbb{N}} \frac{1}{j}\cdot\frac{1}{2^j}$$

Remeber :

$$ln(1+x) = \sum_{j\in\mathbb{N}} \frac{x^j(-1)^{j+1}}{j}, \forall x > -1$$

So we plug in $x = -\frac{1}{2}$ hence,

$$ln(\frac{1}{2}) = \sum_{j\in\mathbb{N}} \frac{-2}{j\cdot 2^j} \Rightarrow -ln(2) = -\sum_{j\in\mathbb{N}} \frac{2}{j\cdot 2^j} \Rightarrow ln(2) = \sum_{j\in\mathbb{N}} \frac{2}{j\cdot 2^j}$$

So in total we have that

$$\mathbb{P}(B) = ln(2)$$

Exercise. A airplane disappear during a flight, through searching, we are given that the airplane exist in two area $1, 2$ in probability p, q, $(q + p = 1)$, given during day searching that the probability for finding in area 1 is α and in area 2 is β.

(1) what is the probability to find the airplane in searching day 1.
(2) Given that we didn't find the airplane in the first day, what the probability that the airplane is in area 1.

Solution. (1).
Denote the events :
$B = \{$ airplane found in day 1 $\}$,
$A = \{$airplane in area 1$\}$
$A^c = \{$airplane in area 2$\}$

$$\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c) = \alpha p + \beta q$$

Solution. (2).
we need to find $\mathbb{P}(A|B^c)$.

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

Now it's not obvious how to find $\mathbb{P}(A \cap B^c)$ but we can use bayes :

$$\mathbb{P}(A|B^c) = \frac{\mathbb{P}(A\cap B^c)}{\mathbb{P}(B^c)} = \frac{\mathbb{P}(B^c|A)\cdot\mathbb{P}(A)}{\mathbb{P}(B^c)} = \frac{(1-\alpha)p}{1-(\alpha p + \beta q)}$$

Exercise. A mokey type randomly on keyboard with no time limit, what is the probability that at some point the monkey will type that word supercalifragilistic-expialidocious is 1.

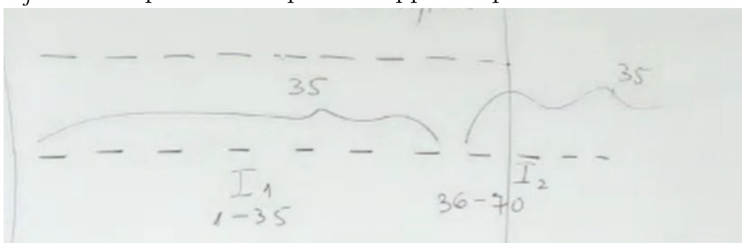
Remark. In reality thats not true Since time is limited.

Solution. we have 100 option in the keyboard, a option for events is :
$A_j = \{$ a sequence appear in place j , $j + N$ $\}$
Denote in A the event at some point the monkey typed the sequence :

$$A = \bigcup_{j=1}^{\infty} A_j$$

Denote I_j the "sequence of letters in keyboead" S.T $I_j \cap I_m = \emptyset, \forall m \neq j$ in length 35 for all j, denote $B_j = \{$ the sequnce appear in interval $I_j\}$

Remark. we are saying the if a sequence appear in B_j so it should be in A Since in B_j the sequence appear Since for example if a sequence appear in A deosn't appear in B_j for example if the sequence happen in place 19 it's in A but not in B_j



So in B_j denote the I_j event in which the sequence "supercalifra...." appear but we can notice that it's appear for cell 1 to 35 after that from 35 to 70 but it can never appear in 19 which mean the sequence in which appear is always in A but not always in B_j so we conclude that :

$$\bigcup_{j=1}^{\infty} B_j \subseteq A$$

We will show that $\mathbb{P}(B_j) = 1$

Solution. Notice that :

$$1 = \mathbb{P}(\bigcup_{j=1}^{\infty} B_j) \leq \mathbb{P}(A) \leq 1$$

Notice that :

$$\mathbb{P}(\bigcup_{j=1}^{\infty} B_j) = \lim_{n\to\infty} \mathbb{P}(\bigcup_{j=1}^{\infty} B_j) = \lim_{n\to\infty} (1 - \mathbb{P}(\bigcup_{j=1}^{\infty} B_j)^c) = \lim_{n\to\infty} (1 - \mathbb{P}(\bigcap_{j=1}^{\infty} B_j^c)) = *$$

We show in the Lecture that if we have independent events in our case $\{B_j\}$Since, I_j are disjoint so $\{B^c\}$ are disjoints hence,

$$* = \lim_{n\to\infty} 1 - \prod_{j=1}^{n} \mathbb{P}(B_j^c))$$

when

$$\mathbb{P}(B_j^c) = 1 - \mathbb{P}(B_j) = 1 - (\frac{1}{100})^{35}$$

So in total we got that :

$$= \lim_{n\to\infty} 1 - \prod_{j=1}^{n} \mathbb{P}(B_j^c)) = \lim_{n\to\infty} (1 - (1 - \underbrace{(\frac{1}{100})^{35}}_{\alpha <!}))^n$$

Now from calculus 1 we know that for $0 < \alpha < 1$ hence,

$$\alpha^n \underset{n\to\infty}{\to} 0$$

So we get that $\mathbb{P}(\bigcup_{j=1}^{\infty} B_j) = 1$ hence,

$$1 = \mathbb{P}(\bigcup_{j=1}^{\infty} B_j) \leq \mathbb{P}(A) \leq 1 \Rightarrow \mathbb{P}(A) = 1$$

Toturial 3:

Exercise. In every tine shimi visit the forest and try to steal orange he success in probability p after every try (event though he success before or not) he exit the forest the is probability of α to get caught, and he do this in independent way till he get caught, and it's happen sometime, can't doit forever.

What is the probability to eat k oranges before he get caught in the first time?

Solution. Denote
$A_j = \{\text{the event in which he got caught in the day } j\ \}$
$B_k = \{\text{Success to eat } k \text{ oranges till first time he caught}\}$

$$\mathbb{P}(B_k) = \sum_{k=1}^{\infty} \underbrace{\mathbb{P}(B_k|A_j)}_{j-benulli-trials-with-success-p} \mathbb{P}(A_j) =_* \sum_{j\geq k}^{\infty} \binom{j}{k} p^k (1-p)^{j-k} \cdot (1-\alpha)^{j-1} \cdot \alpha$$

$$= (\frac{p}{1-p})^k \frac{1}{k!} \frac{\alpha}{1-\alpha} \sum_{j\geq k} j(j-1)(j-k+1) \cdot [(1-p)(1-\alpha)]^j = *$$

Remark. $(1-\alpha)^{j-1} \cdot \alpha = \mathbb{P}(A_j)$ since we want the he didn't caught in the $j-1$ day and caught in j so it's exactly $(1-\alpha)^{j-1} \cdot \alpha$

Remark. The sum only obtain $\sum\limits_{j\geq k}$ Since for $j \leq k$ we have $\mathbb{P}(B_k|A_j)$ because we are looking at the event in which he got caught and then he eat k oranges but we know that it never can be since when he get caught he can't steal again and $j \leq k$ So we have that $\mathbb{P}(\mathbb{P}(B_k|A_j) = 0$.

Denote $f_j(x) = x^j$ hence,

$$\sum_{j\geq k} j(j-1)(j-k+1) \cdot [(1-p)(1-\alpha)]^j = \sum_{j\geq k} j(j-1)(j-k+1) \cdot x^j$$

$$= \sum_{j\geq k} f_j^{(k)}(x) \cdot x^k = x^k \sum_{j=0}^{\infty} f_j^{(k)}(x) =_{\forall x+1>0} x^k (\sum_{j=0}^{\infty} f_j(x))^k) = x^k \cdot (\frac{1}{1-x})^k = x^k \cdot \frac{k!}{(1-x)^{k+1}}$$

We plug in into $x = (1-p)(1-\alpha)$ then:

$$\sum_{j\geq k} j(j-1)(j-k+1) \cdot [(1-p)(1-\alpha)]^j = \frac{(1-p)^k (1-\alpha)^k k!}{(1-(1-p)(1-\alpha))^{k+1}}$$

So in total we get that :

$$* = \frac{p^k (1-\alpha)^{k-1} \alpha}{(1-(1-p)(1-\alpha))^{k+1}}$$

Exercise. A drunk cat go for a trip from the origin point in the Cartesian coordinate system, he go up in probability q and right in probabillity p (in independent way).

(1) What is the probability to reach point $(3,4)$ in 7 steps.
(2) What is the probability to reach point $(3,4)$ at some point.

Solution. (1).

we will translate the problem into bernulli trials, every step is a trial when a success is a step right hence,

$$\mathbb{P}(pass(3,4)) = \mathbb{P}(3 - success - of - 7 - trials) = \binom{7}{3} \cdot p^3 \cdot q^4$$

Solution. (2).

The minimal number of steps to reach $(3,4)$ is 7 steps hence,

$$\mathbb{P}(reach(3,4) - in - 7 - steps) = \mathbb{P}(reach(3,4) - without - pass - in(1,2))$$

we have $\binom{7}{3}$ to arrive $(3,4)$ and we have $\underbrace{\binom{3}{2}}_{to-reach(1,2)} \cdot \underbrace{\binom{4}{2}}_{reach(3,4)-through(1,2)}$

In total we have

$$\binom{7}{3} - \binom{3}{1}\binom{4}{2} = 17$$

(number of ways to rach $(3,4)$ without passing in $(1,2)$ hence,

$$\mathbb{P}(required) = 17p^3q^4$$

Since in anyway we have p^3q^4 to reach and order not matter.

Tutorial 4:

Exercise. number of people which enter a resturant distribute $Poiss(\lambda), \lambda > 0$, given that the probability of person who visit the resturant satisfied in prabability p and not satisfied in probability $(1-p)$.

Denote X random variable which present number of people which satisfied from a meal in someday, show that $X \sim Poiss(\lambda p)$.

Solution. We need to show that

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

Denote $A_j = \{j \text{ person arrive to resturant}\}$

$$\mathbb{P}(X = k) = \sum_{j=0}^{\infty} \mathbb{P}(X = k|A_j)\mathbb{P}(A_j) = \sum_{j \geq k}^{\infty} \binom{j}{k} p^k q^{j-k} \cdot (\frac{\lambda^j}{j!}) \cdot e^{-\lambda}$$

$$= \frac{1}{k!}(\frac{p}{q})^k e^{-\lambda} \sum_{j \geq k}^{\infty} \frac{1}{(j-k)!}(q\lambda)^j = \frac{1}{k!}(\frac{p}{q})^k e^{-\lambda}(q\lambda)^k \underbrace{\sum_{j=0}^{\infty} \frac{(q\lambda)^j}{(j)!}}_{e^{-\lambda q} - taylor - expansion}$$

$$= \frac{1}{k!}(\lambda p)^k e^{-\lambda(1-q)} = \frac{1}{k!}(\lambda p)^k e^{-\lambda p}$$

Are the events :

$A_k = \{k \text{ costumers are satisfied from the meal}\}$

$B_j = \{\ j \text{ costumers wern't satisfied from the meal}\}$

are they independent ?

Solution. First we need to find $\mathbb{P}(A_k)$?

$$\mathbb{P}(A_k) = \mathbb{P}(X = k) = \frac{1}{k!}(\lambda p)^k e^{-\lambda p}$$

$$\mathbb{P}(B_j) = \frac{1}{j!}(\lambda q)^j e^{-\lambda q}$$

Denote Y the random variable which represent number of people which arrived to the store?

$$\mathbb{P}(A_k \cap B_j) = \mathbb{P}(X = k | Y = k+j) = \mathbb{P}(X = k|Y = k+j) \cdot \mathbb{P}(Y = k+j) = \binom{k+j}{k} p^k q^j$$

$$= \binom{k+j}{j} p^k q^j e^{-\lambda} \frac{(\lambda)^{k+j}}{(k+j)!} = \frac{(k+j)!}{j!(k!)} p^k q^j e^{-\lambda} \frac{(\lambda)^{k+j}}{(k+j)!} = p^k q^j e^{-\lambda} \frac{(\lambda)^{k+j}}{j!(k!)}$$

Notice that :

$$e^{-\lambda p} \cdot e^{-\lambda q} = \frac{1}{e^{\lambda p} \cdot e^{\lambda q}} = \frac{1}{e^{\lambda(p+q)}} = e^{-\lambda}$$

hence,

$$p^k q^j e^{-\lambda} \frac{(\lambda)^{k+j}}{j!(k!)} = \underbrace{\frac{1}{k!}(\lambda p)^k e^{-\lambda p}}_{\mathbb{P}(A_k)} \cdot \underbrace{\frac{1}{j!}(\lambda q)^j e^{-\lambda q}}_{\mathbb{P}(B_j)} = \mathbb{P}(A_k) \cdot \mathbb{P}(B_j)$$

Exercise. let $X \sim Exp(\lambda)$ Denote $[X] = Z$ show that $Z \sim Geo(1 - e^{-\lambda})$

Solution. We need to show that :

$$\mathbb{P}(Z = k) = (e^{-\lambda})^k (1 - e^{-\lambda})$$

By definition of the floor function : $\{X = k\} = \{k \leq X \leq k+1\}$ (same events).

$$\mathbb{P}(Z = k) = \mathbb{P}([X] = k) = \mathbb{P}(k \leq X \leq k+1)$$

Remark. $\mathbb{P}(k \leq X \leq k+1) = \mathbb{P}(\{\omega \in \Omega : k \leq X(\omega) \leq k+1\})$

$$\mathbb{P}(k \leq X \leq k+1) = \int_k^{k+1} f_X(t)dt = \lambda e^{-\lambda t} dt = \lambda(\frac{e^{-\lambda t}}{-\lambda})|_{t=k}^{k+1} = -e^{-\lambda(k+1)} + e^{-\lambda k} = (e^{-\lambda})^k (1 - e^{-\lambda})$$

Denote $Y = X - Z$ find the distribution of Y.

Solution. Notice that $Y \in [0.1]$ Since $X - [X] \in [0, 1]$

Remark. notice that $\mathbb{P}(Y = k) = 0$ Since we will conclude that Y is absulote continious random variable.

$$F_Y(t) = \mathbb{P}(Y \leq t) = \left\{ \begin{array}{cc} \mathbb{P}(\emptyset) = 0, & t < 0 \\ ? & 0 \leq t \leq 1 \\ \mathbb{P}(\Omega) = 1, & t > 1 \end{array} \right\}$$

We need to find ?..

$$\mathbb{P}(Y \leq t) = \mathbb{P}(X - [x] \leq t)$$

Now notice that we can use the same event in easier way,

$$\{X - [X] \leq t\} = \bigcup_{k=0}^{\infty} \{k \leq X \leq k + t\}$$

this union is union of disjoint sets hence,

$$\mathbb{P}(Y \leq t) = \mathbb{P}(X - [x] \leq t) = \sum_{k=0}^{\infty} \mathbb{P}(k \leq X \leq k + t) = \sum_{k=0}^{\infty} \int_{k}^{k+t} \lambda e^{-\psi} d\psi$$

$$= \sum_{k=0}^{\infty} (e^{-\lambda})^k (1 - e^{-\lambda t}) = (1 - e^{-\lambda t}) \sum_{k=0}^{\infty} (e^{-\lambda})^k =_{Geometric-Series} \frac{(1 - e^{-\lambda t})}{(1 - e^{-\lambda})}$$

hence,

$$F_Y(t) = \mathbb{P}(Y \leq t) = \left\{ \begin{array}{cc} \mathbb{P}(\emptyset) = 0, & t < 0 \\ \frac{(1-e^{-\lambda t})}{(1-e^{-\lambda})} & 0 \leq t \leq 1 \\ \mathbb{P}(\Omega) = 1, & t > 1 \end{array} \right\}$$

So to find the density function

$$f_Y(t) = F_y'(t) = \left\{ \begin{array}{cc} 0, & t < 0 \\ \frac{\lambda e^{-\lambda t}}{1-e^{-\lambda}} & 0 \leq t \leq 1 \\ 0, & t > 1 \end{array} \right\}$$

Remark. notice that $\int_{-\infty}^{\infty} f_Y(t)dt = 1$

Tutorial 6 :

Theorem. *Given a random variable X exists Discrete random variable Z and continious random variable Y, $p \in [0,1]$ S.T :*

$$F_X(t) = pF_Z(t) + (1 - p)F_Y(t)$$

Exercise. let X random variable with a distribution function :

$$F_X(t) = \left\{ \begin{array}{cc} \frac{2-t}{(3-t)^2} & t < 1 \\ C & 1 \leq t \leq 2 \\ \frac{3t-1}{16} & 2 < t < 5 \\ 1 & t \geq 5 \end{array} \right\}$$

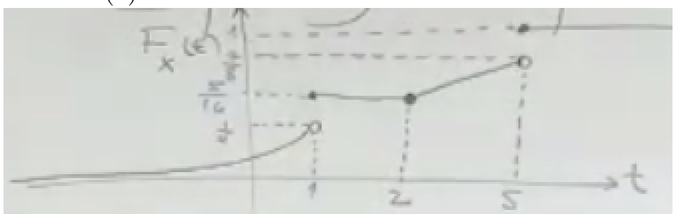
(1) Find $C \in \mathbb{R}$.
(2) Which type of random variable is X.
(3) Find the factorization which satisfied by the Theorem.

Solution. (1).

We know that the distribution function is continious in every right neighberhood hence,

$$\lim_{t \to 2^+} F_X(t) = F_X(2) = C \Rightarrow \lim_{t \to 2^+} \frac{3t - 1}{16} = \frac{5}{16} = C$$

Solution. (2).



By the graph we coclude that X is mixed random variable.

Solution. (3).
Define function :

$$\tilde{F}_Y(t) = \left\{ \begin{array}{cc} \frac{2-t}{(3-t)^2}, & t < 1 \\ \frac{1}{16}, & 1 \leq t \leq 2 \\ \frac{3t-1}{16} - \frac{1}{16} & 2 < t < 5 \\ \frac{13}{16} & t \geq 5 \end{array} \right\}$$

This function is not distribution function Since for $t \geq 5$ is not 1 so we can define :

$$F_Y(t) = \frac{16}{13}\tilde{F}_Y(t)$$

Now define

$$\tilde{F}_Z(t) = \left\{ \begin{array}{cc} 0, & t < 1 \\ \frac{1}{16}, & 1 \leq t \leq 2 \\ \frac{1}{16} & 2 < t < 5 \\ \frac{1}{16} + \frac{1}{8} & t \geq 5 \end{array} \right\}$$

Again it's not a distrubtiom function sice for $t \geq 5$ we have that $\tilde{F}_Z(t) = \frac{1}{16} + \frac{1}{8}$ So define :

$$F_Z(t) = \frac{16}{3}\tilde{F}_Z(t)$$

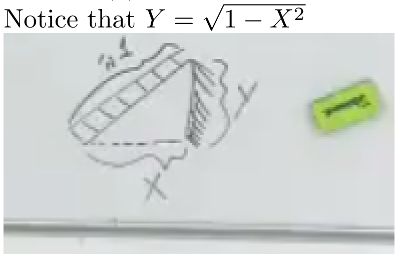
So Notice that :

$$F_X(t) = \tilde{F}_Z(t) + \tilde{F}_Y(t) = \frac{3}{16}F_Z(t) + \frac{16}{13}F_Y(t)$$

Exercise. A ladder laid on wall (1meter height) on the wall when the bottom edge is placed with random distance from the wall, in a uniform distribution on $[0, 1]$ we will denote Y as the random variable which represent the height of point in which the scale touch the wall in relation to the ground.

(1) find $E[Y]$.

Solution. (1).
Notice that $Y = \sqrt{1 - X^2}$



$$F_Y(t) = \mathbb{P}(Y \le t) = \left\{ \begin{array}{cc} 0 & t < 0 \\ ? & t \in [0.1) \\ 1 & t \ge 1 \\ \end{array} \right\}$$

let $0 \le t \le 1$ we need to find ? :

$$\mathbb{P}(Y \le t) = \mathbb{P}(\sqrt{1 - X^2} \le t) =_* \mathbb{P}(1 - X^2 \le t^2) =_{**} \mathbb{P}(X \ge \sqrt{1 - t^2}) = 1 - F_X(\sqrt{1 - t^2})$$

where in * we used the fact that $t > 0$.
where in $**$ we used the fact that X has postive values.
Notice that :

$$F'_Y(t) = f_Y(t) = \left\{ \begin{array}{cc} 0, & t < 0 \\ \frac{t}{\sqrt{1-t^2}}, & t \in [0,1) \\ 0, & t \ge 1 \end{array} \right\}$$

$$E[Y] = \int_{\mathbb{R}} f_Y(t) t dt = \int_0^1 \frac{t^2}{\sqrt{1 - t^2}} dt$$

let $t = sin(\theta)$ and we get that :

$$= \int_0^{\frac{\pi}{2}} \frac{sin^2(\theta)}{|cos(\theta)|} \cdot cos(\theta) d\theta =_{cos(\theta) > 0} \int_0^{\frac{\pi}{2}} \frac{(1 - cos(2\theta)}{2}) d\theta = \frac{0 - \frac{sin(2\theta)}{2}}{2} |_{\theta=0}^{\frac{\pi}{2}} = \frac{\pi}{4}$$

Tutorial 6 :

Definition. Variance of random variable with a moment order two finite marked as $Var(X)$ and defined :

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

Exercise. find the variance of Gaussian standard random variable :

Solution. let $Z \sim N(0, 1)$ we saw that $E[Z] = 0$
First we will find $E[Z^2]$.

$$E[Z^2] = \int_{\mathbb{R}} t^2 f_Z(t) dt = \int_{\mathbb{R}} t \cdot \frac{t}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Denote $v = t, du = \frac{t}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ hence,

$$[t \cdot \frac{1}{\sqrt{2\pi}} e^{\frac{t^2}{2}}]_{t=-\infty}^{t=\infty} - \int -1 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} = 0 + \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} = 1$$

We conclude that

$$E[Z] = E[Z^2] - E^2[Z] = 1 - 0^2 = 1$$

Exercise. find the variance of random variable $exp(\lambda)$

Solution. Denote $X \sim Exp(\lambda)$ we saw that $E[X] = \frac{1}{\lambda}$.

$$E[X^2] = \int_{\mathbb{R}} t^2 f_X(t)dt = \int_0^\infty t^2 e^{-\lambda t}dt$$

Denote $u = t^2, dv = e^{-\lambda t}$

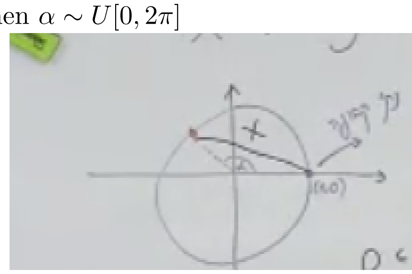
$$= [t^2 e^{-\lambda t}]_{t=0}^{\infty} - \int_0^\infty -2te^{-\lambda t} = 0 + \frac{2}{\lambda}\int_0^\infty \lambda t e^{-\lambda t}dt = \frac{2}{\lambda^2}$$

So we conclude that :

$$Var[X] = E[X^2] - E^2[X] = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Exercise. we choose a random point on circle, denote X the random variable which represent the distance of point from given point on the circle, find $Var[X]$.

Solution. Denote by α the degree between the radius with the given point with x then $\alpha \sim U[0, 2\pi]$



then $X = 2\sin(\frac{\alpha}{2})$

$$E[X] = \int_0^{2\pi} 2sin(\frac{t}{2}) \cdot \frac{1}{2\pi} = \frac{1}{\pi}(-cos(\frac{t}{2}))|_{t=0}^{t=2\pi} = \frac{4}{\pi}$$

$$E[X^2] = \int_0^{2\pi} 4sin^2(\frac{t}{2})\frac{1}{2\pi}dt = \frac{2}{\pi}\int_0^{2\pi}(1 - cos(t))dt = \frac{1}{\pi}(t - sin(t))|_{t=0}^{2\pi} = 2$$

So we conclude that :

$$Var[X] = 2 - \frac{16}{\pi} \approx 0.38$$

Exercise. if $X \sim Bin(n.p)$ we saw that $E[X] = np$

$$E[X^2] = \sum_{k=0}^{n} k^2\mathbb{P}(X = k) = \sum_{k=0}^{n} k^2\binom{n}{k}p^k q^{n-k} = \sum_{k=0}^{n} k\frac{n!}{(k-1)!(n-k)!}p^k q^{n-k}$$

$=$

$$np\sum_{k=1}^{n} k\frac{(n-1)!}{(k-1)!(n-k)!}p^k q^{n-k} = np\sum_{k=1}^{n} k\binom{n-1}{k-1}p^{k-1}q^{n-1-(k-1)}$$

$$np\sum_{k=1}^{n}\binom{n-1}{k-1}p^{k-1}q^{n-1-(k-1)} + np\sum_{k=1}^{n}(k-1)\binom{n-1}{k-1}p^{k-1}q^{n-1-(k-1)}$$

$$=_{k-1=j} np\sum_{j=0}^{n-1} j\binom{n-1}{j}p^jq^{n-1-j} + np\sum_{j=0}^{n-1} j\binom{n-1}{j}p^jq^{n-1-j} = *$$

Now notice that $np\sum\limits_{j=0}^{n-1} j\binom{n-1}{j}p^jq^{n-1-j} = E[Y]$ for $Y \sim Bin(n-1,p)$

Notice that $\sum\limits_{j=0}^{n-1} j\binom{n-1}{j}p^jq^{n-1-j} = (p+q)^{n-1} = 1^{n-1} = 1$ Since it's identical to the binomial formula.

So in total we have that :

$$* = np(n-1)p + np = np([(n-1)p+1]) = np(np+(1-p)) = np(np+q)$$

hence,

$$Var[X] = np(np+q) - E^2[X] = np(np+q) - (np)^2 = npq$$

Tutorial 7 :

Jensen inequality

Exercise. let $r \le p$ natural numbers and X is random variable show that :

$$\sqrt[r]{E[|X|^r} \le \sqrt[p]{E[|X|^p}$$

Corollary. *if $E[|X|^p]$ finite, then $E[|X|^r]$ finite $\forall p \ge r$.*

Solution. notice that :

$$E[|X|^p] \le E[(|X|^r)^{\frac{p}{r}}]$$

Look at the funcition $f(x) = x^{\frac{p}{r}}$ it's a convex function for all $x \ge 0$, So by jensen :

$$E[(f(Y)] \ge f(E[Y])$$

Denote $Y = |X|^r$ hence,b

$$E[|X|^P] \ge (E[Y])^{\frac{p}{r}} = E[|X|^r]^{\frac{p}{r}}$$

Exercise. Show Inequality of arithmetic and geometric means using jensen inequality

Solution. let $(a_i)_{i=}^n$ a positive numbers we need to show that :

$$\sqrt[n]{\prod_{i=1}^n a_j} \le \frac{\sum\limits_{j=1}^n a_j}{n}$$

Define Random variable X which satisfy $X = a_j$with probability $\frac{1}{n}$ $\forall j$ and look into the function $f(x) = log(x)$ and $f(x)$ is up convex in $x \ge 0$ and from jensen :

$$E[f(X)] \leq f(E[X])$$

$$\sum_{j=1}^{n} f(a_j)\mathbb{P}(X=a_j) \leq f(\sum_{j=1}^{n} a_j \mathbb{P}(X=a_j))$$

$$\Downarrow$$

$$\frac{1}{n}\sum_{j=1}^{n} log(a_j) \leq log(\frac{\sum\limits_{j=1}^{n} a_j}{n}) \Rightarrow log((\prod_{i=1}^{n} a_j)^{\frac{1}{n}}) \leq log(\frac{\sum\limits_{j=1}^{n} a_j}{n})$$

Since log is montotnic we get the required i.e

$$\sqrt[n]{\prod_{i=1}^{n} a_j} \leq \frac{\sum\limits_{j=1}^{n} a_j}{n}$$

Random vectors

Multi-binomial distribution :

we have a sequence of independent trials in which in every trial we can get k possible results, every result in probability p_j, $1 \leq j \leq k$ when $\sum\limits_{j=1}^{k} p_j = 1$.

(1) find the probability that in a sequence of n trials we get x_1 result 1 and x_2 trials a result 1 ... till x_k result k when $\sum\limits_{j=1}^{n} x_j = n$.
(2) find the marignal distribution of Y_2 for $k = 3$.
(3) find the marignal distribution of Y_2 for general k.

Solution. (1).
Define K random variables $(y_j)_{j=1}^{K}$ when Y_j is the random variable which represent the number of times in which we got result with type j .

$$\mathbb{P}(required) = \mathbb{P}((Y_j)_{j=1}^{K} = (x_j)_{j=1}^{K}) = \binom{n}{x_1} p_1^{x_1} \binom{n-x_1}{x_2} p_2^{x_2} \cdots \binom{n-\sum\limits_{j=1}^{K-1} x_j}{x_K} p_k^{x_k}$$

$$=\prod_{j=1}^{k} \binom{n-\sum\limits_{i=1}^{j-1} x_i}{x_j} p_j^{x_j} = \prod_{j=1}^{k} \frac{(n-\sum\limits_{j=1}^{j-1} x_i)!}{(x_j)!(n-\sum\limits_{j=1}^{j} x_i)!} p_j^{x_j} = n! \prod_{j=1}^{k} \frac{1}{(x_j)!} p_j^{x_j}$$

Solution. (2).
Notice that :

$$\mathbb{P}(Y_2 = m) = \sum_{l=0}^{n-m} \mathbb{P}((Y_j)_{j=1}^{3} = (l, m, n-m-l)) = \sum_{l=0}^{n-m} n! \frac{1}{l!} p_1^{l} \frac{1}{m!} p_2^{m} \frac{1}{(n-m-l)!} p_3^{n-l-lm}$$

$$= n!\frac{1}{m!}p_2^m\sum_{l=0}^{n-m}\frac{1}{l!}p_1^l\frac{1}{(n-l-m)!}p_3^{n-l-m} = \frac{n!}{m!(n-m)!}p_2^m\underbrace{\sum_{l=0}^{n-m}\underbrace{\frac{(n-m)!}{l!(n-m-l)!}}_{\binom{n-m}{l}}p_1^l p_3^{n-m-l}}_{=(p_1+p_3)^{n-m}}$$

$$= \binom{n}{m}p_2^m(p_1+p_3)^{n-m} = \binom{n}{m}p_2^m(1-p_2)$$

Hence,

$$Y_2 \sim Bin(p_2, n)$$

Solution. (3).

We will look at the story in different way, in every trial we will define a success as result from type 2 and otherwise we have failure, the probability to success in every single trial is p_2 and the probability for failure is $1-p_2$ or we can say $p_1+..+p_k = \sum_{j\neq 2}^{K} p_j = 1-p_2$ i.e $Y_2 = Bin(p_2, n)$

Turtorial 8:

Absolutely Continuous Random Vectors:

Exercise. let (X_1, X_2) be random vector with density,

$$f_{(X_!,X_2)}(t,s) = \begin{Bmatrix} 2e^{-s}e^{-t}, & 0<t<s<\infty \\ 0, & otherwise \end{Bmatrix}$$

find the density function (Y_1, Y_2) when $Y_1 = 2X_1, Y_2 = X_2 - X_1$ and show that they are independent.

Solution. look at the function $f:\mathbb{R}^2 \to \mathbb{R}^2$ when

$$f(x_1, x_2) = (2x_1, x_2 - x_1)$$

f is injective hence, we can use the following :

$$f_{(Y_1,Y_2)}(y_1,y_2) = f_{(x_1,x_2)}(x_1(y_1,y_2)), x_2(y_1,y_2))|\frac{\partial(x_1,x_2)}{\partial(y_1,y_2)}|$$

in our case :

$$x_1(y_1,y_2) = \frac{y_1}{2}, x_2(y_1,y_2) = \frac{y_1}{2} + y_2$$

So we plug in into the fist function and we get that :

$$f_{(Y_1,Y_2)}(y_1,y_2) = \begin{Bmatrix} 2e^{-\frac{y_1}{2}}e^{-\frac{y_1}{2}+y_2}\cdot\underbrace{|\frac{\partial(x_1,x_2)}{\partial(y_1,y_2)}|}_{Jacobian}, & 0<y_1,y_2<\infty \\ 0, & otherwise \end{Bmatrix}$$

we will calculate the jacobian :

$$\frac{\partial(x_1,x_2)}{\partial(y_1,y_2)}| = | \begin{matrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial x_2} & \frac{\partial x_2}{\partial x_2} \end{matrix} | = | \begin{matrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 1 \end{matrix} | = \frac{1}{2}$$

hence,

$$f_{(Y_1,Y_2)}(y_1,y_2) = \left\{ \begin{matrix} 2e^{-\frac{y_1}{2}}e^{-\frac{y_1}{2}+y_2} \cdot \frac{1}{2}, & 0 < y_1, y_2 < \infty \\ 0, & otherwise \end{matrix} \right\}$$

now to show independency first :

$$f_{Y_1}(y_1) = \int_0^\infty f_{(Y_1,Y_2)}(y_1,y_2)dy_2 = \int_0^\infty e^{-y_1}e^{-y_2}dy_2$$

$$e^{-y_1}\int_0^\infty e^{-y_2}dy_2 = e^{-y_1} \cdot 1$$

hence,

$$f_{(Y_1)}(y_1) = \left\{ \begin{matrix} e^{y_1}, & 0 < y_1 \\ 0, & otherwise \end{matrix} \right\}$$

in identecal way we conclude that :

$$f_{(Y_2)}(y_2) = \left\{ \begin{matrix} e^{y_2}, & 0 < y_2 \\ 0, & otherwise \end{matrix} \right\}$$

Notice that :

$$f_{(Y_2)}(y_2)f_{(Y_1)}(y_1) = f_{(Y_1,Y_2)}(y_1,y_2)$$

So we have independency as required.

Exercise. the vector (X, Y, Z) distibute uniformly in the unit ball $\{(x, y, x)|x^2 + y^2 + z^2 \leq 1\}$ is X, Y independent,

Solution. we need to find $f_{(X,Y)}(t, s)$ and $f_X(y), f_Y(t)$.
notice that :

$$f_{(X,Y)}(t,s) = \int_{-\sqrt{1-x^2-y^2}}^{\sqrt{1-x^2-y^2}} f_{(X,Y,Z)}(t,s,\zeta)d\zeta$$

Now since we have a unifom distribution we have that :

$$f_{(X,Y,Z)}(t,s,r) = \left\{ \begin{matrix} \frac{1}{\frac{4\pi}{3}}, & t^2+s^2+r^2 = 1 \\ 0, & otherwise \end{matrix} \right\}$$

$$\Rightarrow f_{(X,Y)}(t,s) = \left\{ \begin{matrix} \frac{3}{2\pi}\sqrt{1-s^2-t^2}, & t^2+s^2 \leq 1 \\ 0, & otherwise \end{matrix} \right\}$$

Now notice that :

$$f_X(t) = \left\{ \begin{matrix} \int_{-\sqrt{1-t^2}}^{\sqrt{1-t^2}} f_{(X,Y)}(t,s)ds, & t^2 \leq 1 \\ 0, & otherwise \end{matrix} \right\}$$

Now we need to calculate :

$$\int\limits_{-\sqrt{1-t^2}}^{\sqrt{1-t^2}} f_{(X,Y)}(t,s)ds = \int\limits_{-\sqrt{1-t^2}}^{\sqrt{1-t^2}} \frac{3}{2\pi}\sqrt{1-s^2-t^2}ds = \int\limits_{-\sqrt{1-t^2}}^{\sqrt{1-t^2}} \frac{3}{2\pi}\sqrt{1-(\frac{s}{\sqrt{1+t^2}})}ds$$

let $sin(\theta) = (\frac{s}{\sqrt{1+t^2}})$ now to determine the intergtal limits we plug in indstead of s we know that $s \in (-\sqrt{1-t^2}, \sqrt{1-t^2})$ so $sin(\theta)$ is between $-1 = \frac{-\sqrt{1-t^2}}{\sqrt{1-t^2}} \leq sin\theta \leq (\frac{s}{\sqrt{1+t^2}}) \leq \frac{\sqrt{1-t^2}}{\sqrt{1-t^2}} = 1$ so $-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$ hence,

$$\frac{3}{2\pi}\sqrt{1-t^2}\int\limits_{-\frac{\pi}{2}}^{\frac{\pi}{2}} cos(\theta)\cdot\sqrt{1-t^2}cos(\theta)d\theta = \frac{3}{2\pi}(1-t^2)\underbrace{\int\limits_{-\sqrt{1-t^2}}^{\sqrt{1-t^2}} cos^2(\theta)}_{\frac{\pi}{2}} = \frac{3}{2\pi}(1-t^2)\cdot\frac{\pi}{2}$$

hence,

$$f_X(t) = \left\{ \begin{array}{cc} \frac{3}{4}(1-t^2), & t^2 \leq 1 \\ 0, & otherwise \end{array} \right\}$$

In the same calculation we get $f_Y(t)$ so we have that :

$$f_Y(t) = \left\{ \begin{array}{cc} \frac{3}{4}(1-s^2), & s^2 \leq 1 \\ 0, & otherwise \end{array} \right\}$$

Moreover,

$$f_{(X,Y)}(t,s) \neq f_X(t)\cdot f_Y(s)$$

So X, Y are not independent.

Convulotion

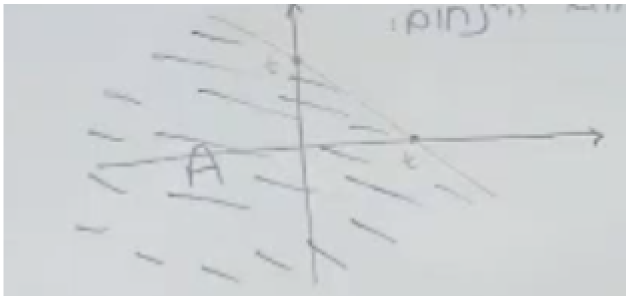
Exercise. Show the formula for convulotion let X, Y a independent random vectors and absouletely continious show that $Z = X + Y$ given by :

$$f_Z(t) = \int\limits_{\mathbb{R}} f_X(\zeta)f_Y(t-\zeta)d\zeta$$

Solution. Notice that :

$$F_Z(t) = \mathbb{P}(z \leq t) = \mathbb{P}(x+y \leq t) = \mathbb{P}((x,y) \in A)$$

When A is this Volume in the graph :



$$= \int\limits_A f_{(X,Y)}(\alpha,\beta)d\alpha d\beta$$

Remark. Since X, Y are independent then $f_{(X,Y)}(\alpha,\beta)$ is absolute continious, notice that it's not necessarily that $f_{(X,Y)}(\alpha,\beta)$ is absolute continious if X, Y are not independent Since, it can be that $X \sim U, Y \sim U$ but $f_{(X,Y)}(\alpha,\beta)$ is not uniform.

Now since they are independence we have that :

$$f_{(X,Y)}(\alpha,\beta) = f_X(\alpha)\cdot f_Y(\beta)$$

Now

$$\int\limits_A f_{(X,Y)}(\alpha,\beta)d\alpha d\beta = \int\limits_{\mathbb{R}} d\alpha(\int\limits_{-\infty}^{t-\alpha} f_X(\alpha)\cdot f_Y(\beta)d\beta) = \int\limits_{\mathbb{R}} f_X(\alpha)((\int\limits_{-\infty}^{t-\alpha} f_Y(\beta)d\beta))d\alpha$$

$$=_{\beta=\zeta-\alpha} \int\limits_{\mathbb{R}} f_X(\alpha)(\int\limits_{-\infty}^{t-\alpha} f_Y(\zeta-\alpha)d\zeta) =_{fobbini} \int\limits_{-\infty}^{t} (\underbrace{\int\limits_{\mathbb{R}} f_X(\alpha)f_Y(\zeta-\alpha)d\alpha}_{g(\zeta)})d\zeta$$

Now but the foundemettal Theorem of calculus we have that :

$$F_Z(t) = \int\limits_{-\infty}^{t} g(\zeta)\zeta \Rightarrow F'_Z(t) = g(t) \Rightarrow F_Z(t) = \int\limits_{\mathbb{R}} f_X(\alpha)f_Y(\zeta-\alpha)d\alpha)$$

<u>INDEPENDENT RANDOM VARIABLES</u>

Exercise. Given $(X_j)_{j=1}^3$ independent random variables and uniform on $[0,1]$ find the convulotion of them.

Solution. In the matrial we saw that

$$\underbrace{f_{X_1} * f_{X_2}}_{g}(t) = \left\{ \begin{array}{cc} 0, & t<0 \\ t, & t\in[0,1] \\ 2-t, & t\in[1,2] \\ 0, & t>2 \end{array} \right\}$$

$$f_{X_1} * f_{X_2} * f_{X_3} = g * f_{X_3}(t) = \int_{\mathbb{R}} g(x) f_{X_3}(t-x)dx = *$$

Now since $X_3 \sim U[0,1]$ so $t - x \in [0,1] \to 0 \le t - x \le 1 \to t - 1 \le x \le t$

$$* = \int_{\mathbb{R}} g(x)\mathbf{1}_{[t-1,t]}(x)dx = \int_{\mathbb{R}} (x \cdot \mathbf{1}_{[0,1]}(x) + (2-x)\cdot \mathbf{1}_{[1,2]}(x)) \cdot \mathbf{1}_{[t-1,t]}(x)dx$$

$$= \int_{\mathbb{R}} (x \cdot \mathbf{1}_{[0,1]\cap[t-1,t]}(x) + (2-x)\cdot \mathbf{1}_{[1,2]\cap[t-1,t]}(x))dx$$

$$= \left\{ \begin{array}{cc} 0, & t<0 \\ \int_{\mathbb{R}} x \cdot 1_{[0,t]}(x)dx, & 0 \le t \le 1 \\ \int_{\mathbb{R}} x \cdot 1_{[t-1,1]}(x)dx + (2-x)1_{[1,t]}(x)dx, & 1 \le t \le 2 \\ \int_{\mathbb{R}} (2-x)1_{[t-1,2]}(x)dx, & t \in [2,3] \\ 0, & t > 3 \end{array} \right\} = \left\{ \begin{array}{cc} 0, & t<0 \\ \frac{t^2}{2}, & 0 \le t \le 1 \\ 3t - t^2 - 1.5, & 1 \le t \le 2 \\ \frac{t^2}{2} - 3t + 4.5, & t \in [2,3] \\ 0, & t>3 \end{array} \right\}$$

Exercise. let $(X_j)_{j=1}^n$ independent random variables and let $(h_j : \mathbb{R} \to \mathbb{R})$ is borel function show that $(h_j(X_j))_{j=1}^n$ are independent random variables.

Solution. let $(A_j)_{j=1}^n$ be borel sets.

$$\mathbb{P}(\forall 1 \le j \le n, h_j(X_j) \in A_j) = \mathbb{P}(\forall 1 \le j \le n, X_j \in h_j^{-1}(A_j))$$

Now h_j is borel sets hence bu the definition $h_j^{-1}(A_j) \in \beta(\mathbb{R})$ and by independence og (X_j) stem that :

$$= \prod_{j=1}^n \mathbb{P}(X_j \in h_j^{-1}(A_j)) = \prod_{j=1}^n \mathbb{P}(X_j \in h_j(A_j))$$

.

\