

Covariance

It is a statistical measure that describe the degree to which two variables are linearly related or not.

i/p (House price dataset)

x₁	x ₂	x ₃	<u>o/p</u> price
Room	Bedroom	Bath	place

→ relation betw i/p & o/p, or 2 i/p feature

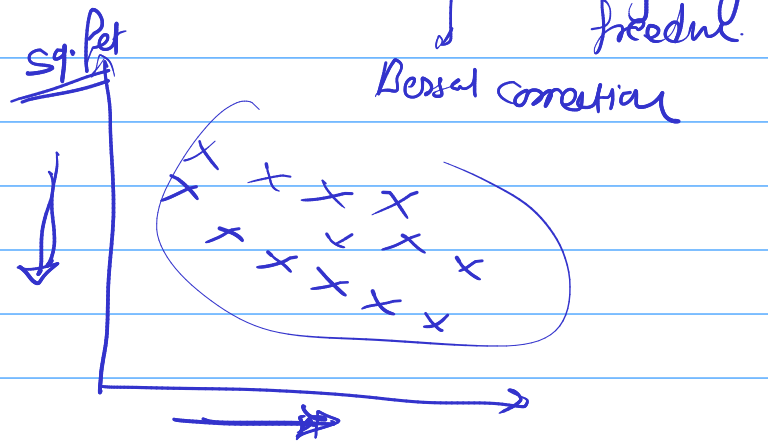
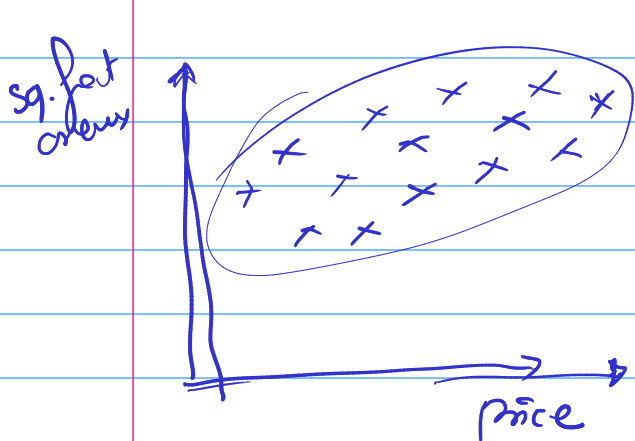
Sq. feet ↑ → price ↑ or ↓

Covariance → relation

Population

Sample

$$\text{Cov}(x, y) = \frac{\sum (x - \mu_x)(y - \mu_y)}{N} = \frac{\sum (x - \bar{x})(y - \bar{y})}{n-1}$$

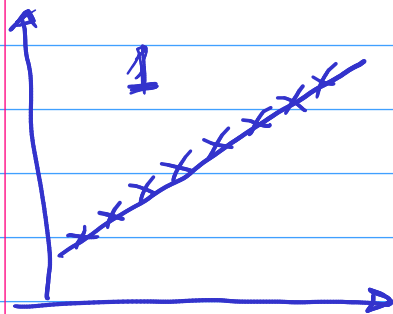


Disadvantage of Cov. → we know the relation is positive or negative but we don't know how much?

(1000) + (10) +

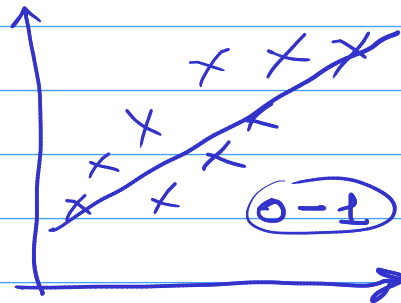
Correlation — Can Quantify the weak & strong relationship.

& it lie between -1 to 1 (range of correlation)



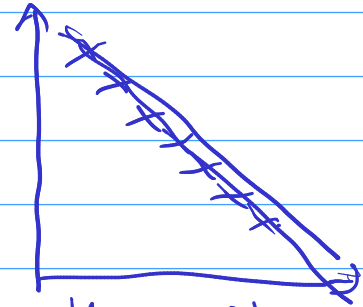
Strongly
relation
Positive Correlation

1



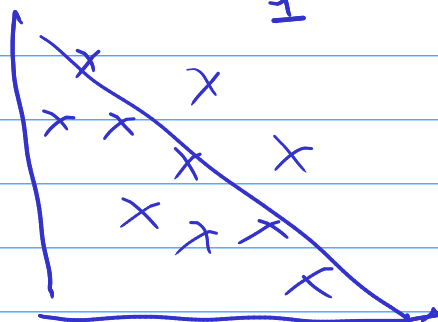
Positive,
Weak positive

0 to 1

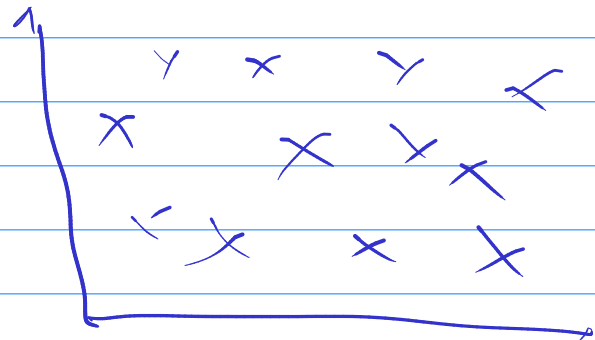


Highly Strong
negative

-1



Weak negative
 -1 to 0



No relationship
range (0) correlation

$$\text{Correlation} = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y}$$

visualise the correlation

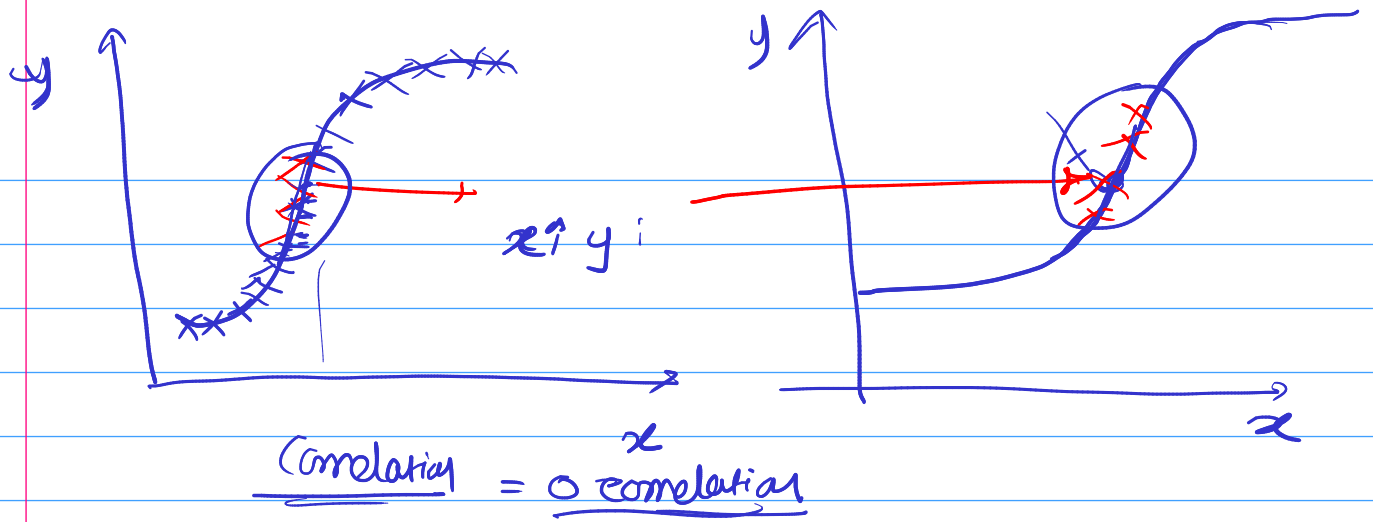
— Scatter plot (2 variable / Bivariate analysis)

— Hue parameter (For graph we use 2nd para. as hue)

— Joint plot →

— Heatmap — or

— Pairplot. →



Spearman Correlation

$$\text{Spear}(x, y) = \frac{\text{Cov}(R(x), R(y))}{R_{\sigma x} R_{\sigma y}}$$

Height weight		$R(x)$	$R(y)$
170	55	3	4
165	43	4	5
160	65	5	2
175	82	2	1
190	85	1	3

* Probability Distribution

— is a list of all of the possible outcomes of a random variable along with their corresponding probability value.

ex-coin (H, T)

Coin toss	1 (H)	0 (T)
prob	1/2	1/2

die

{1, 2, 3, 4, 5, 6}
6 outcome

1	2	3	4	5	6	Pr
1/6	1/6	1/6	1/6	1/6	1/6	

2 dice
dice

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

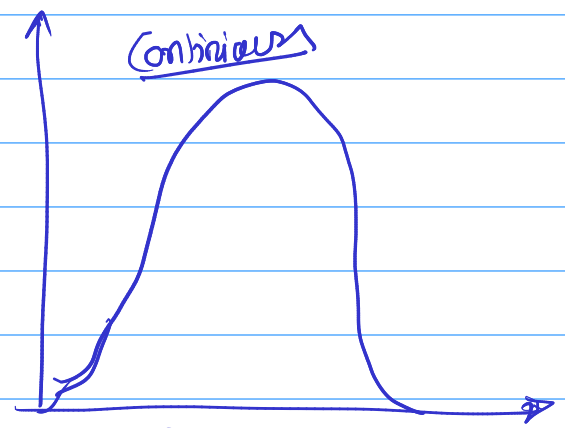
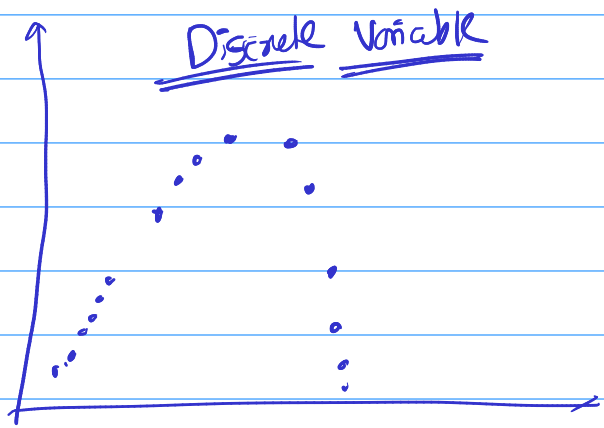
Prob	
1	0
2	1/36
3	2/36
4	3/36
5	
6	
7	6/36
...	
12	1/36

36 outcome

Probability Calculate \rightarrow if we have more dataset it is difficult to write down all the table

3
4
10

Probability Distribution (visualize the probability of function)



Probability Distribution Function

Continuous

Discrete

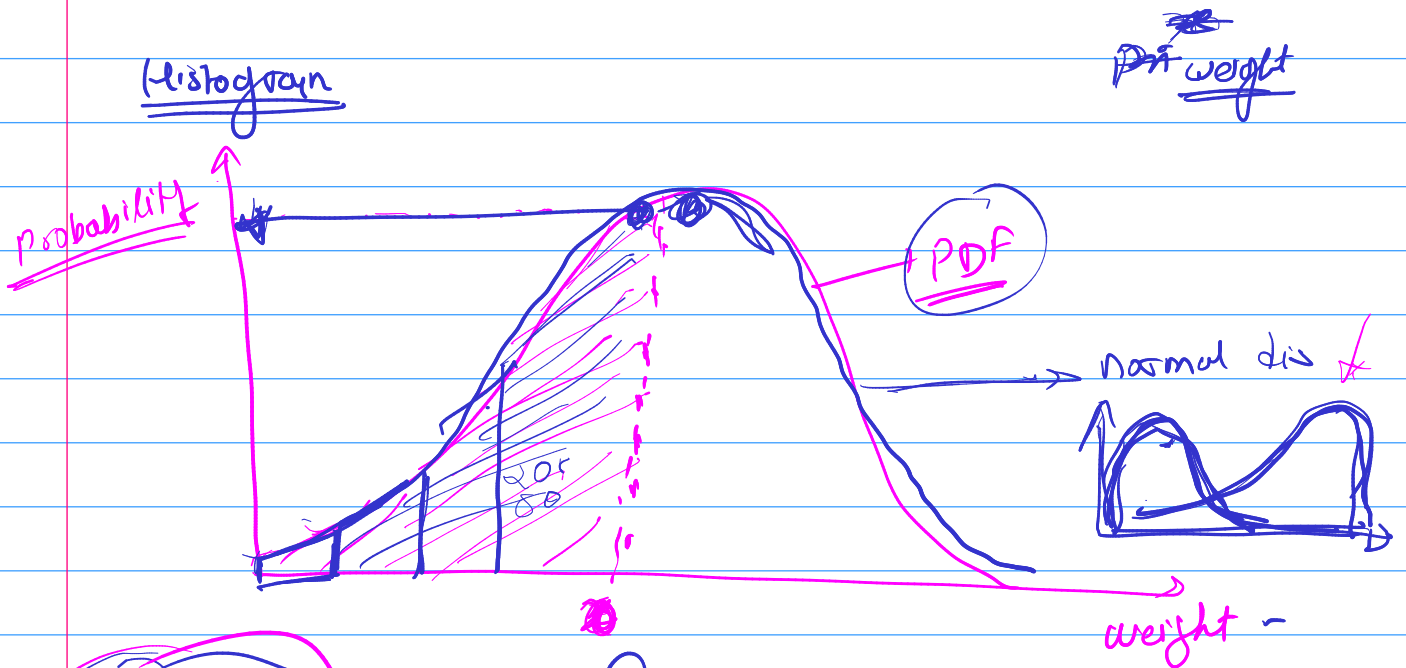
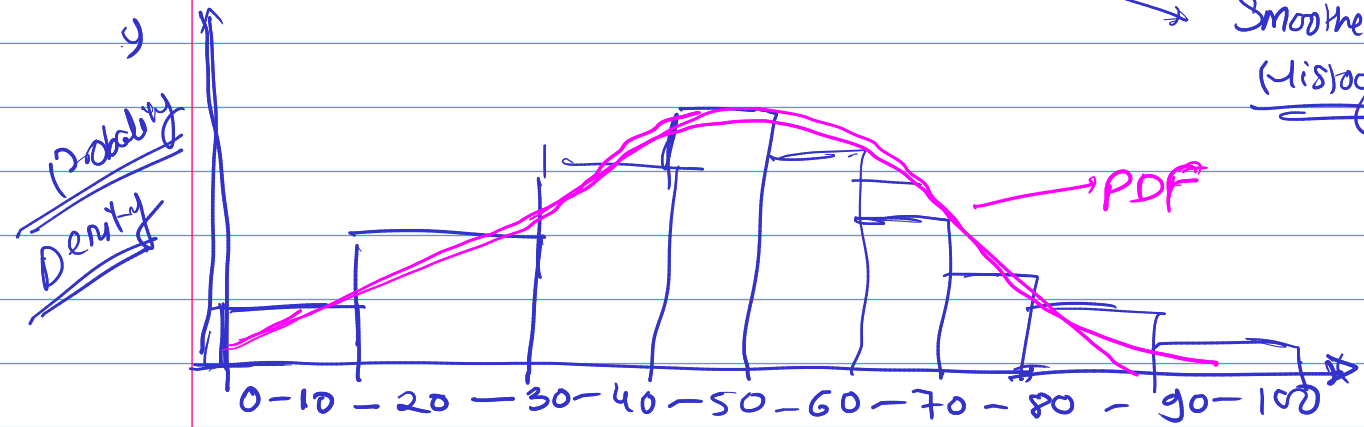
\Rightarrow Probability Density function (PDF)

Probability mass function (PMF)

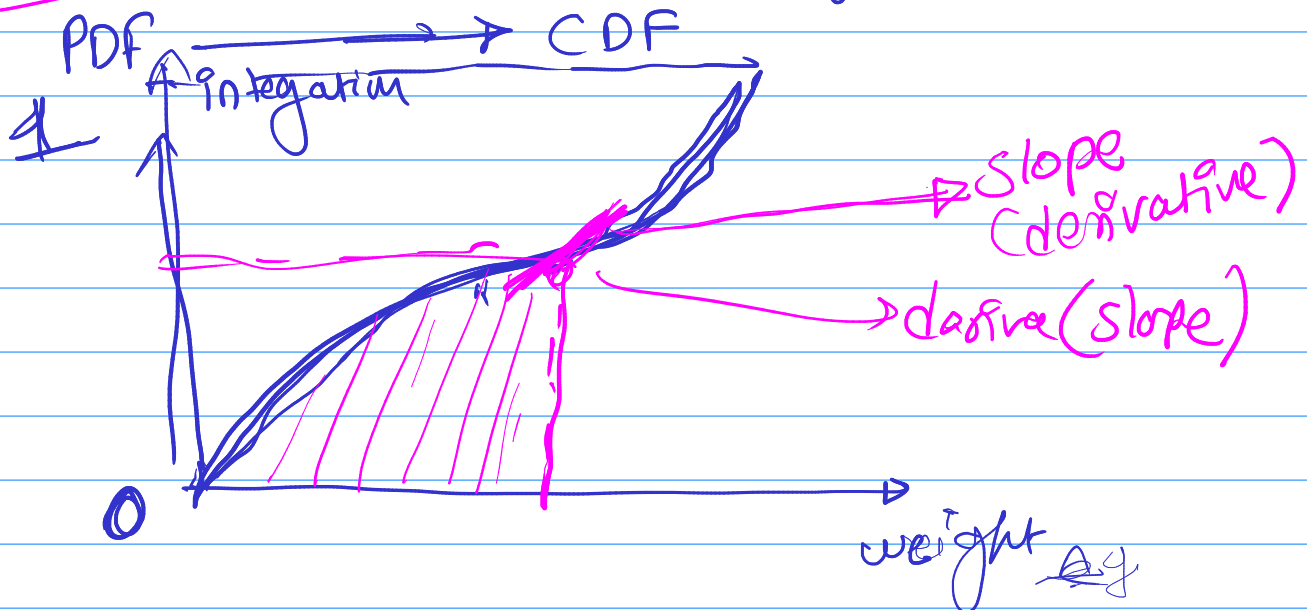
Area \rightarrow Cumulative Density function (CDF)

area \rightarrow Cumulative mass function (CMF)

9) Probability Density Function (PDF) Continuous Smoother the Histogram



Cumulative distribution function (ranges betw 0 to 1)



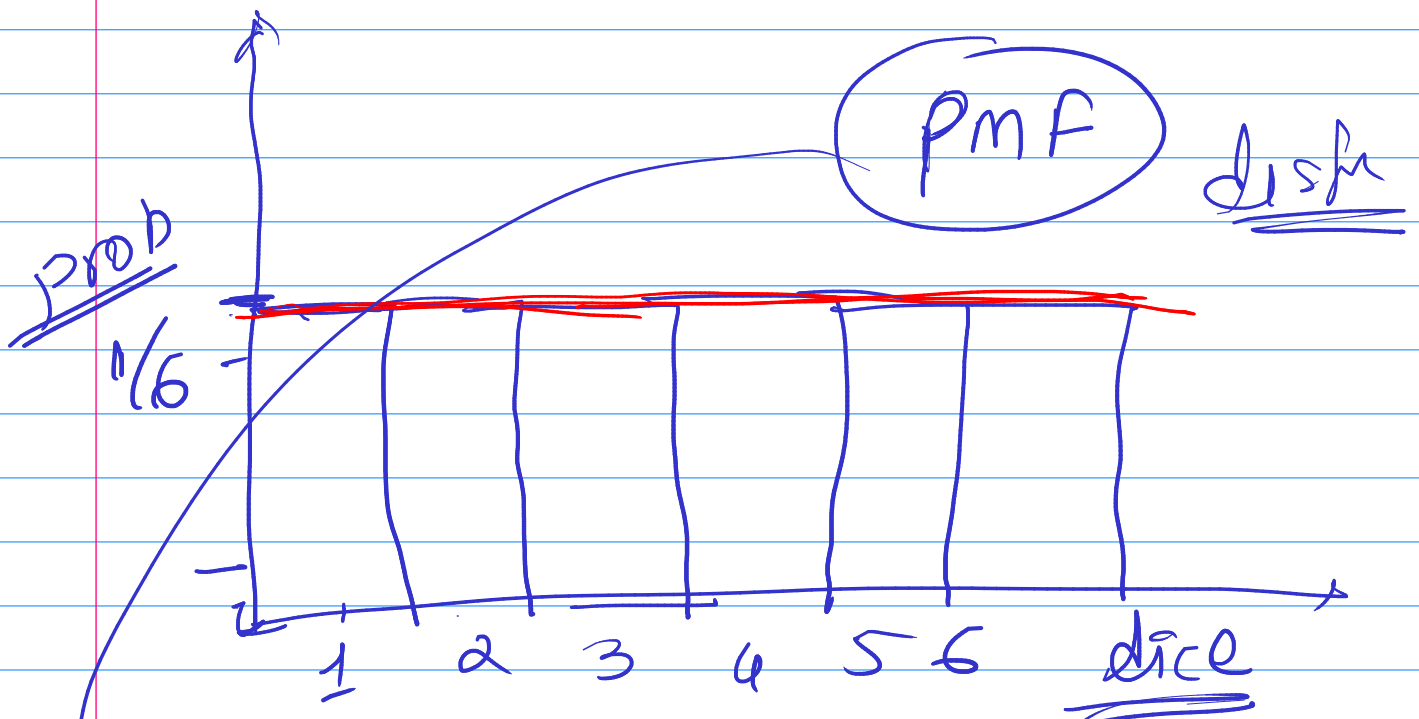
PDF \rightarrow Continuous

PMF (Probability Mass Function)

Discrete Random Variable

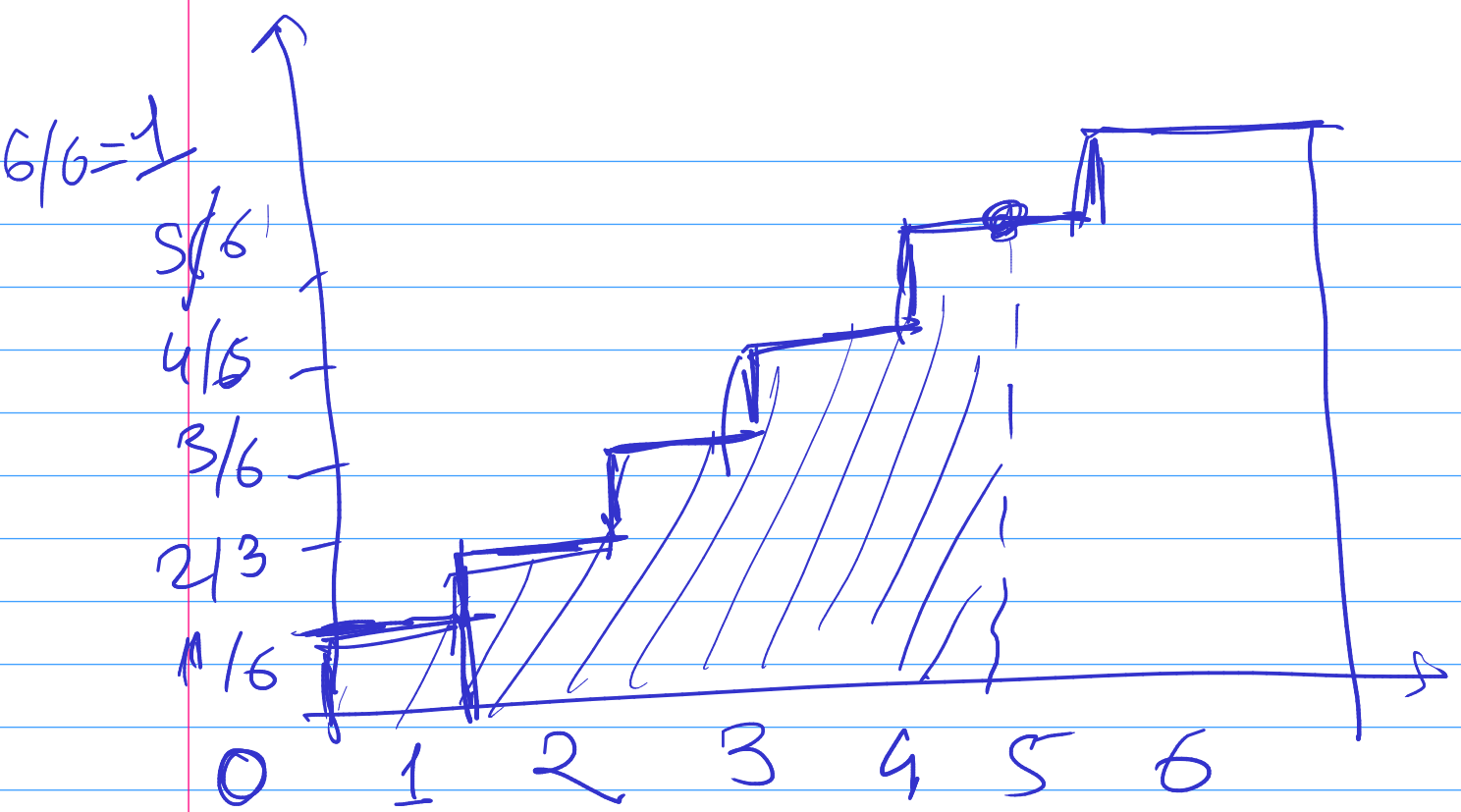
eg → dice

	1	2	3	4	5	6
<u>Prob</u>	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$



Uniform distribution

CMF → Cumulative mass function



~~CDF~~

→ area

Prob-density funct

area ^{continuous} _{fun}

PDF → CDF - contin

PMF → CMF → Dis

probab mass → commas

PDF → Probable distribution _{fun}