Day 4 → Oct 6, 2024

--→ Continuing Probability & Statistics for ML

PDF (Probability Density function)

→ describe the likelihood of a continuous random variable taking a specific value.

→ Over a given interval 0 to 1, the area under the curve of a PDF for a given interval, the probability that the random variable falls within that interval.

→ Area under curve must be 1, the total probability of all possible outcomes.

$$P(a \le X \le b) = \int_a^b f(x) \cdot dx$$

where $f(x)$ is the PDF of the random variable X.

A function $p: \mathcal{X} \to \mathbb{R}$ is probability density function if

$$\forall x \in \mathcal{X}: p(x) \geq 0$$

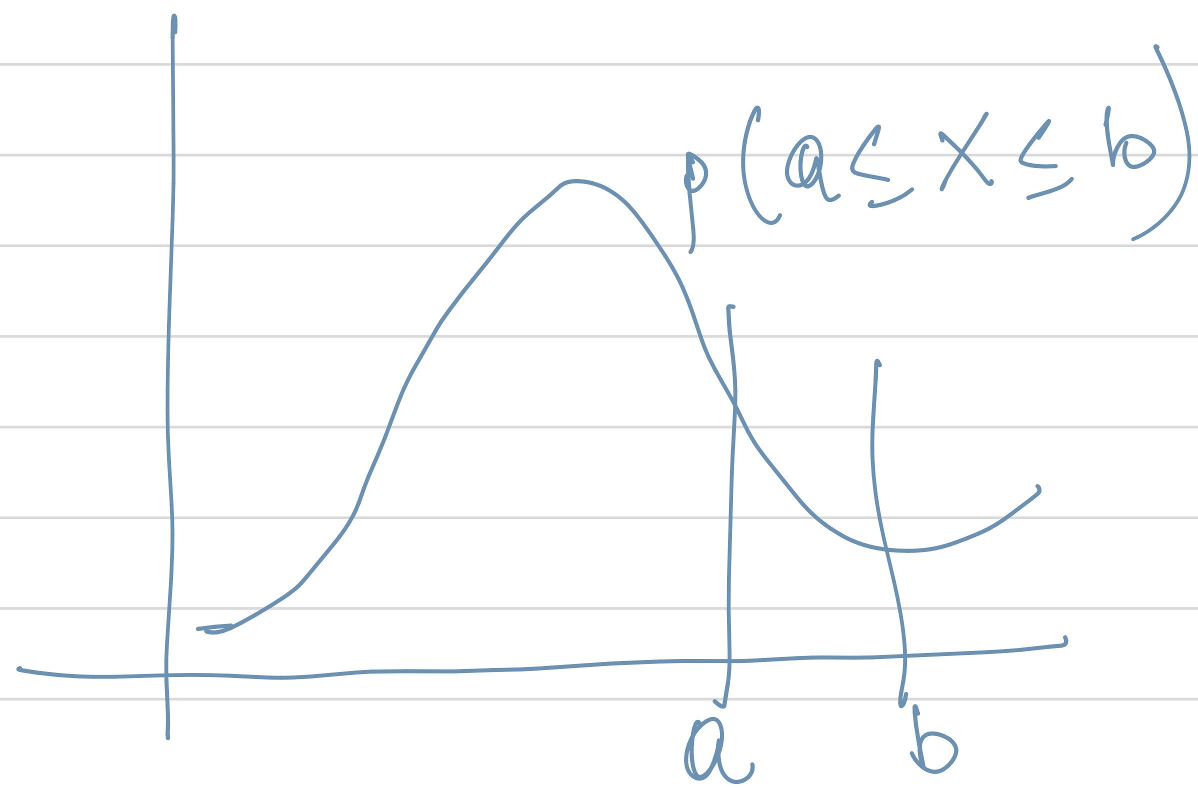
- the integral exists

$$\int_X p(x) \cdot dx = 1$$

- the domain X contains all possible x can take.

Features

- Non-Negative
- Total Area is 1
- p falls between $[a, b]$



Example

Gaussian Distribution

PDF $\rightarrow$ $p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

where $x \Rightarrow$ calculating value

$\sigma \Rightarrow$ standard deviation

$\mu \Rightarrow$ mean

So, $p(x)$ is not probability, it's density

* To obtain Probability, we need to integrate

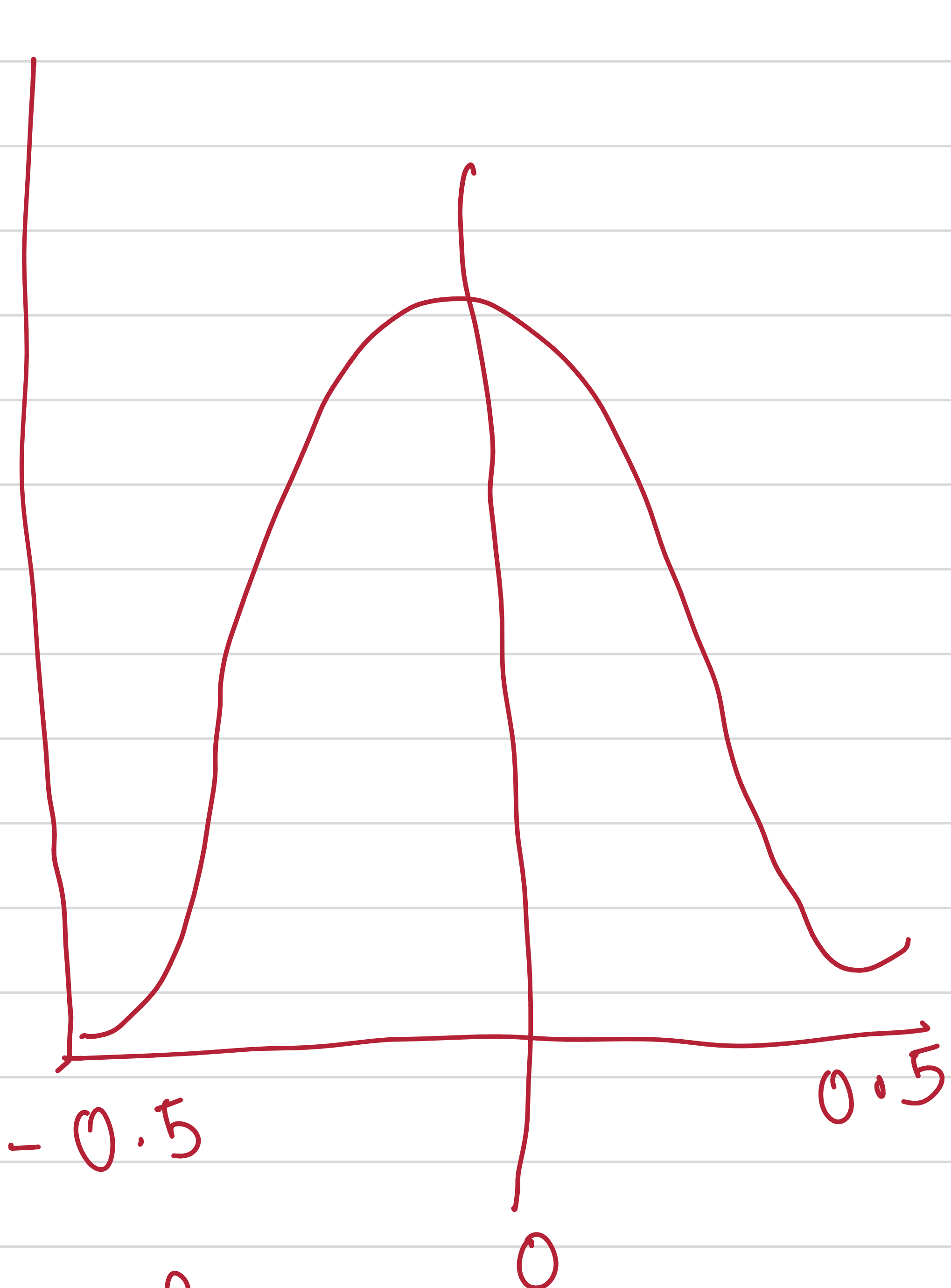
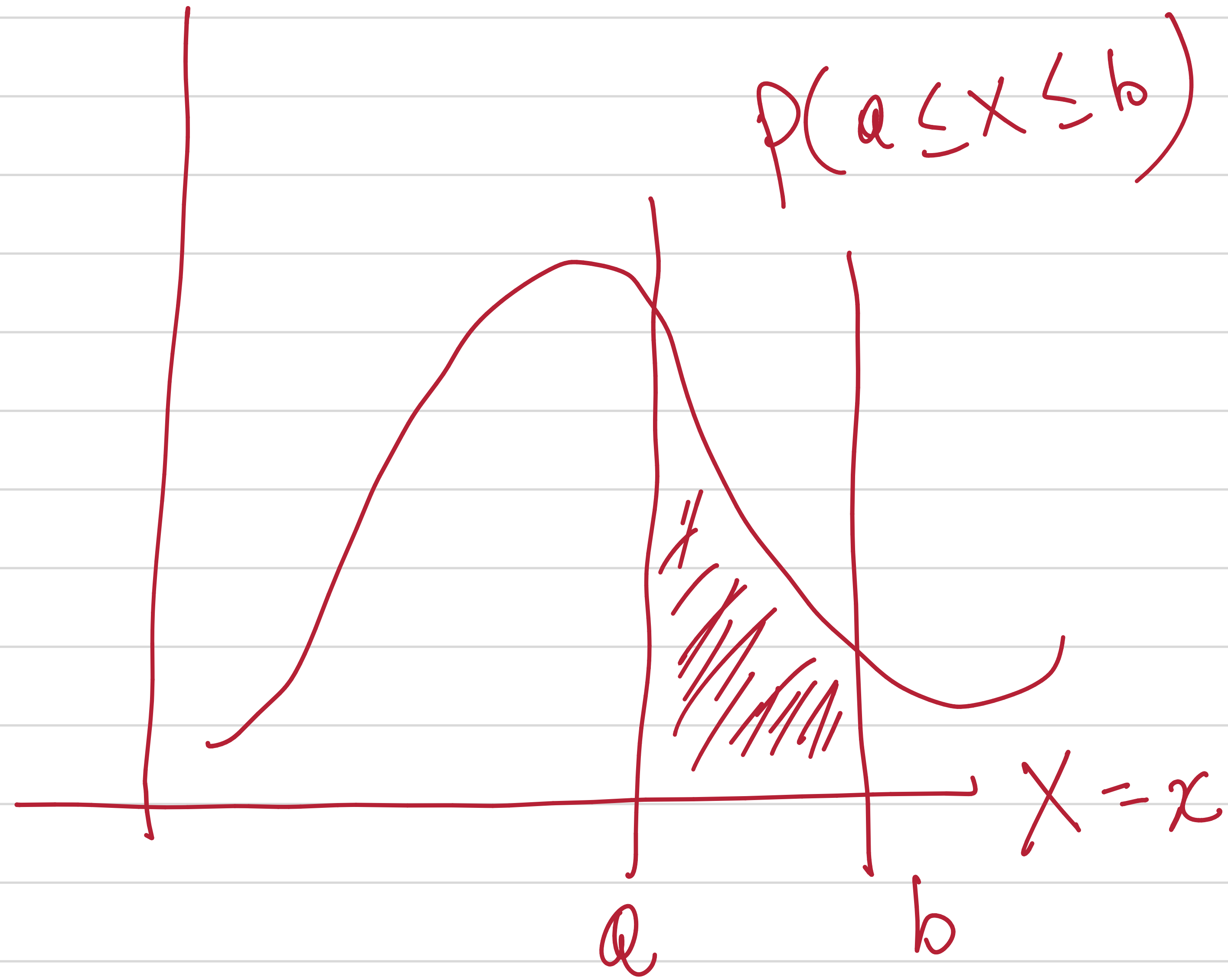


fig. Gaussian distribution



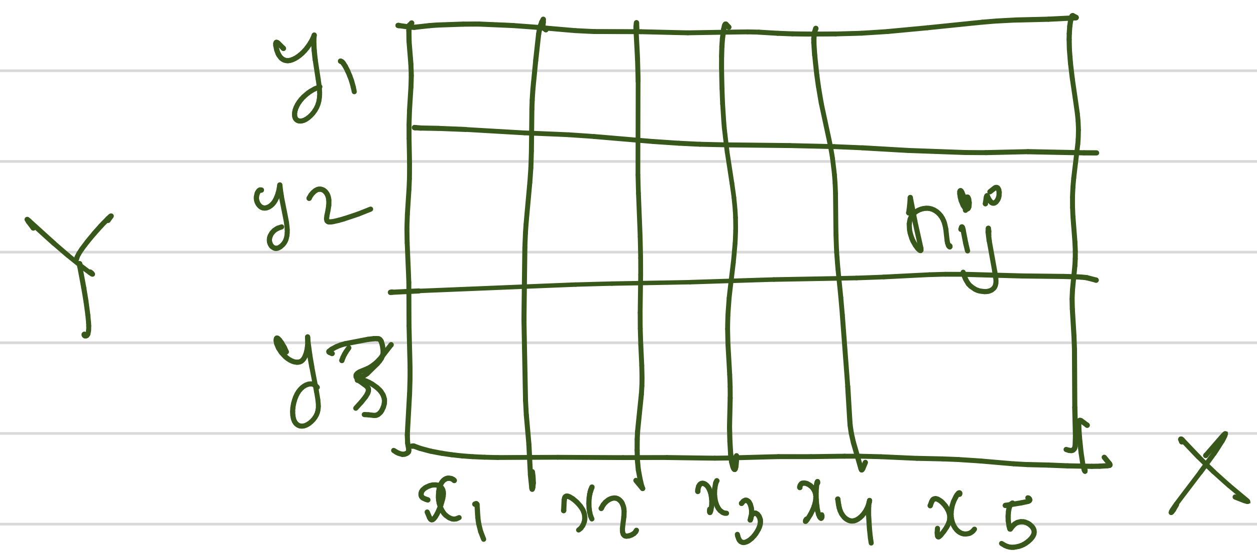
Compute Area it gives 'p' that occurs within a range

Joint Probability Distribution

→ Probability of multiple events occurs simultaneously

- Discrete Event: $p(X=x_i, Y=y_j)$ is a joint probability distribution of the event

$$X=x_i, \quad Y=y_j$$



- Continuous : $p(x,y)$ denotes the Joint Probability density

→ Example of JPD is tossing coin and stock price Simultaneously where $x \Rightarrow$ Coin toss and $Y =$ stock price

→ two events Simultaneously Occuring

→ Need to do the double Integration for the area under the Curve.

Continuous Case

$$P(a \le X \le b, c \le Y \le d) \Rightarrow \int_c^d \int_a^b f(x,y) \cdot dx \cdot dy$$

Double integration (Integral part represents the probability that X falls between a and b and Y falls between c and d.

→ P of some events at a time

Marginal probability

→ For a discrete, Marginal probability is obtained by Summing over the other variable.

Marginals of X ↳ $g(x) = \sum_{y} p(x, y)$

X \ Y	1	2	3	
1	0.32	0.03	0.01	0.36
2	0.06	0.24	0.02	0.32
3	0.02	0.03	0.27	0.32
	0.40	0.30	0.30	1

Marginals for Y → 0.40, 0.30, 0.30

$h(y) = \sum_{x} p(x, y)$

$1 \rightarrow \sum_{x} \sum_{y} p(x, y) = 1$

→ JPD talks about two Events (p) so the die and card

$x = 1$ and $y = 1$

then the event of $x=1, y=1$ (p) is

from Marginal Probability as 0.36 (At Single $x=1, y=1$)

X (rain)	Y (Umbrella)	JPD $p(x,y)$
Yes	Yes	0.4
Yes	No	0.1
No	Yes	0.2
No	No	0.3

MP for discrete

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

MP for Continuous

$$P(X=x) = \int f(x,y) \cdot dy$$

MP ~~table~~ is somehow normalized, for the JPD and sum is 1.

MP is the 'p' of one event occuring, regardless of other.

From the table Rain and Umbrella

for X = Rain, we have to 'Σ' the JPD for each case of y

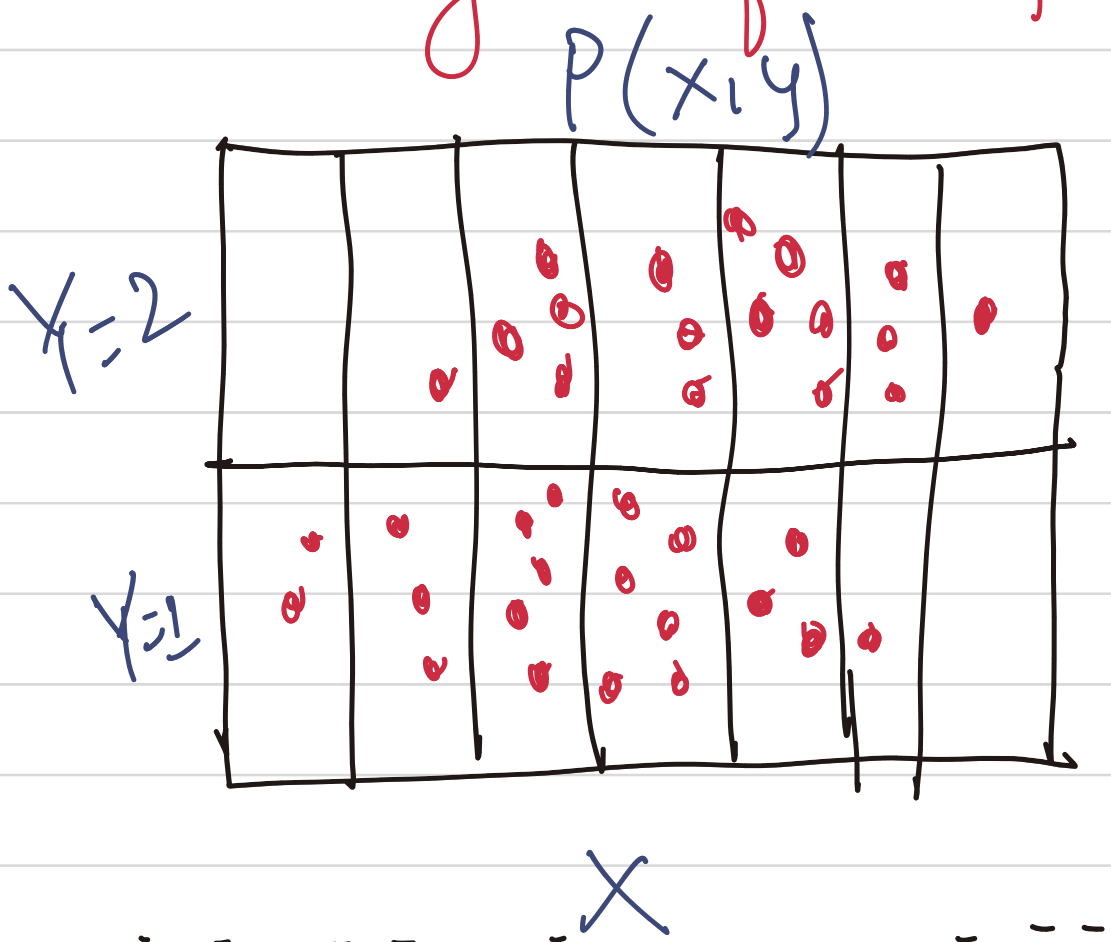
$P(X = Yes) \Rightarrow 0.4 + 0.1 \Rightarrow 0.5$

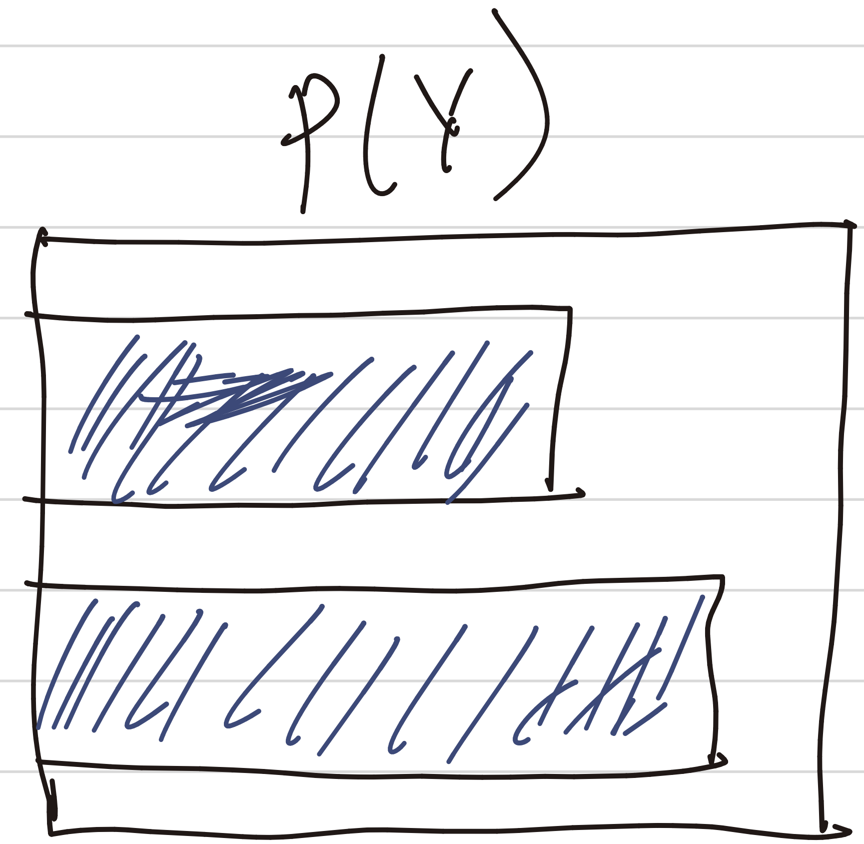
$P(X = NO) \Rightarrow 0.2 + 0.3 \Rightarrow 0.5$

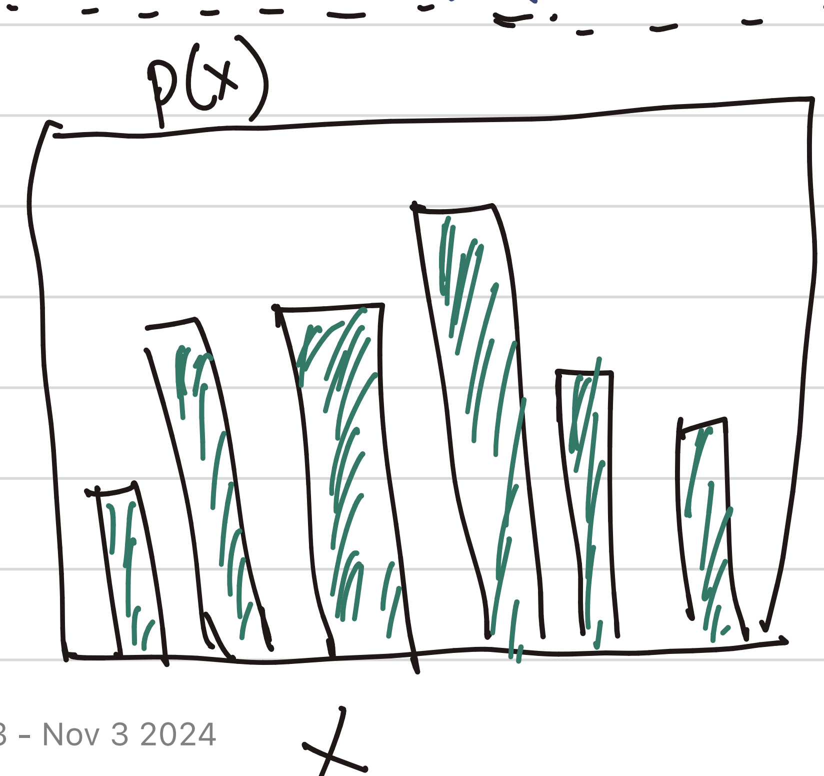
using Formula

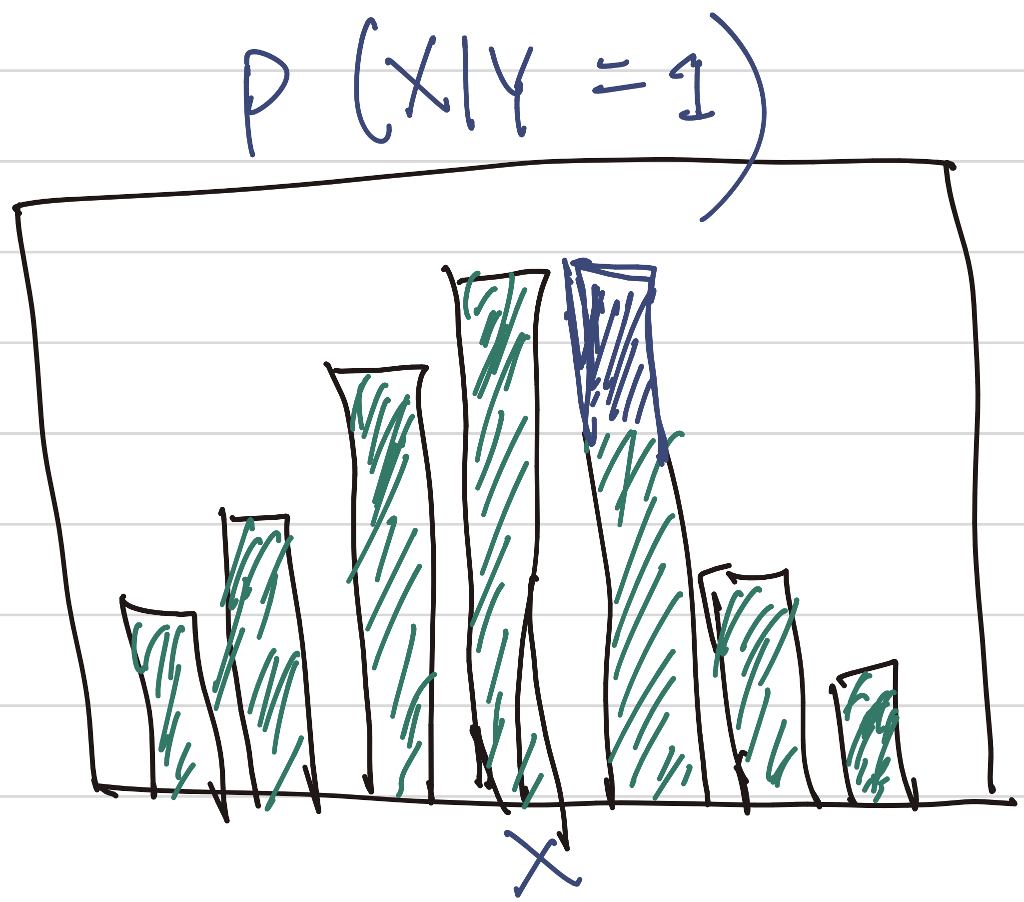
$$P(X = x) = \sum_y P(X = x, Y = y)$$

Diagrams for MP









Find X under Y is Known if it is Conditional Probability $p(x|y=1)$

- Integrate out the other variable (Continuous)

$$p(x) = \int p(x, y)\, dy$$

$$p(y) = \int p(x, y)\, dx$$

- Similarly for higher number of variables (Continuous)

$$p(x_i) = \int p(x_1, \ldots \ldots \ldots \ldots, x_D)\, dx_{\backslash i}$$

Leave x_1 & Compute for others except x_1

Multivariate Gaussian

- Probability Density function

$$p(x|\mu, \Sigma) = (2\pi)^{-\frac{D}{2}} |\Sigma|^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

we write,

$$p(x) = \mathcal{N}(x|\mu, \Sigma)$$

$$X \sim \mathcal{N}(\mu, \Sigma)$$

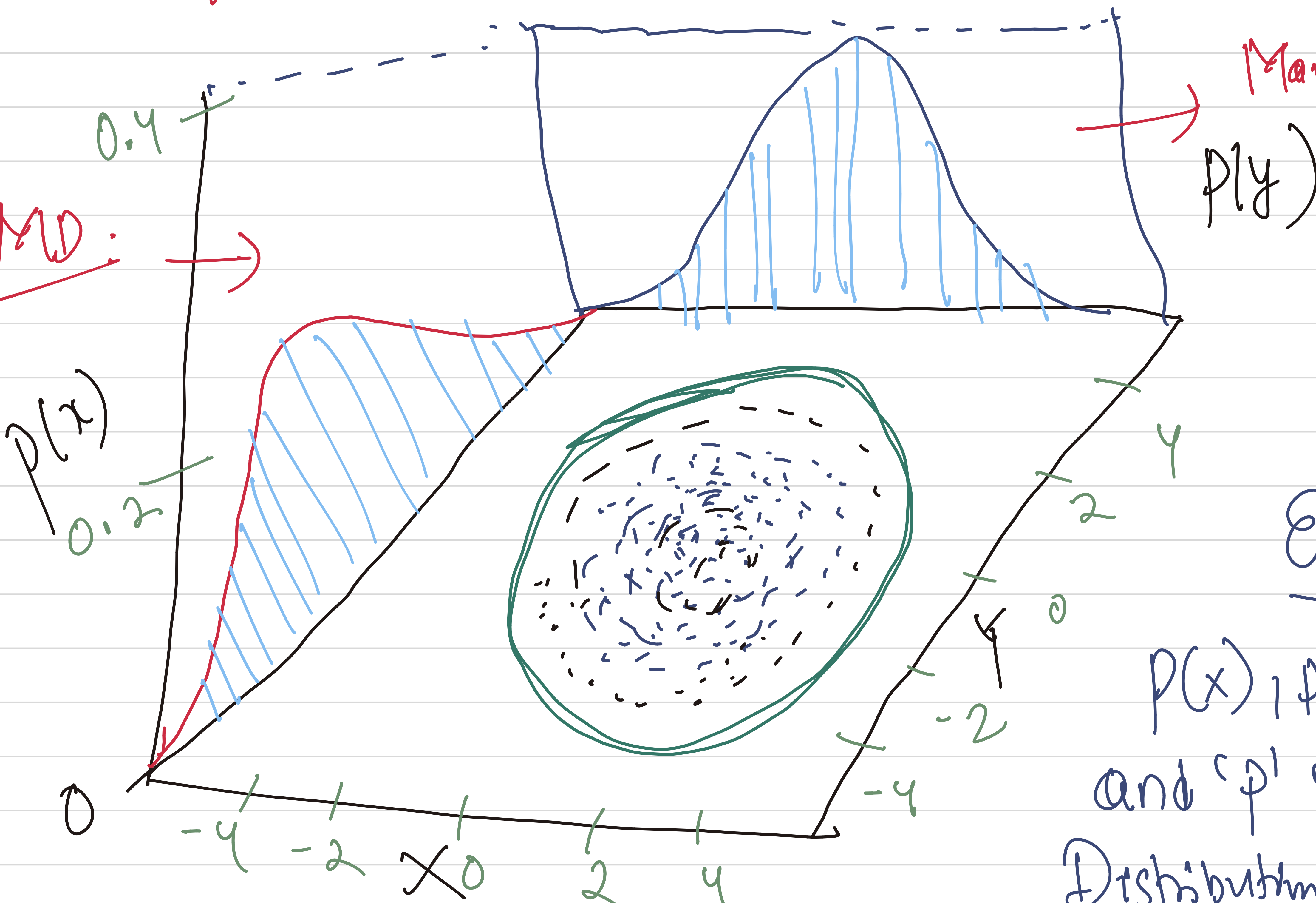
where
D is the number of dimensions

$|\Sigma| \Rightarrow$ determinant of covariance matrix Σ

$\mu \Rightarrow$ mean, $(x-\mu)^T \rightarrow$ transpose of vector $(x-\mu)$, x = variable

Figure of PDF Distribution (Multi-variate Gaussian)

MD :



Marginal Distribution

Except '1'

$P(x), P(Y) \Rightarrow MP$

and 'P' is JPD

Distributions are Gaussian

Conditional Probability

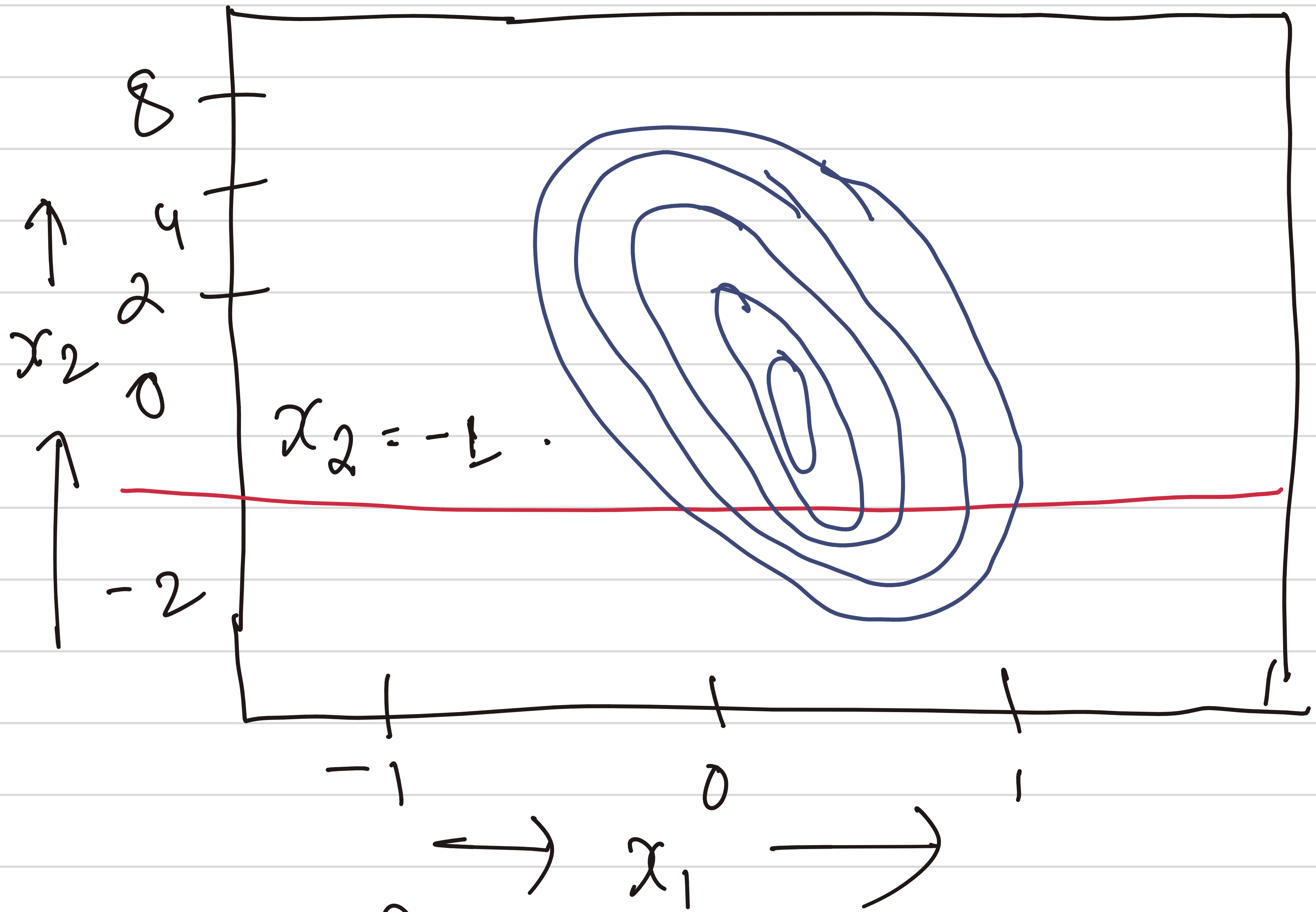
→ Fix the value of one variable
→ Compute the probability of the other variable

But first let's understand distribution (describe the 'p' occuring within a range) and Normal distribution has mean (μ) at peak, σ measures dispersion, and it is bell-shaped, Continuous PD, is Symmetry, μ is tends to be normalized, have two Parameters

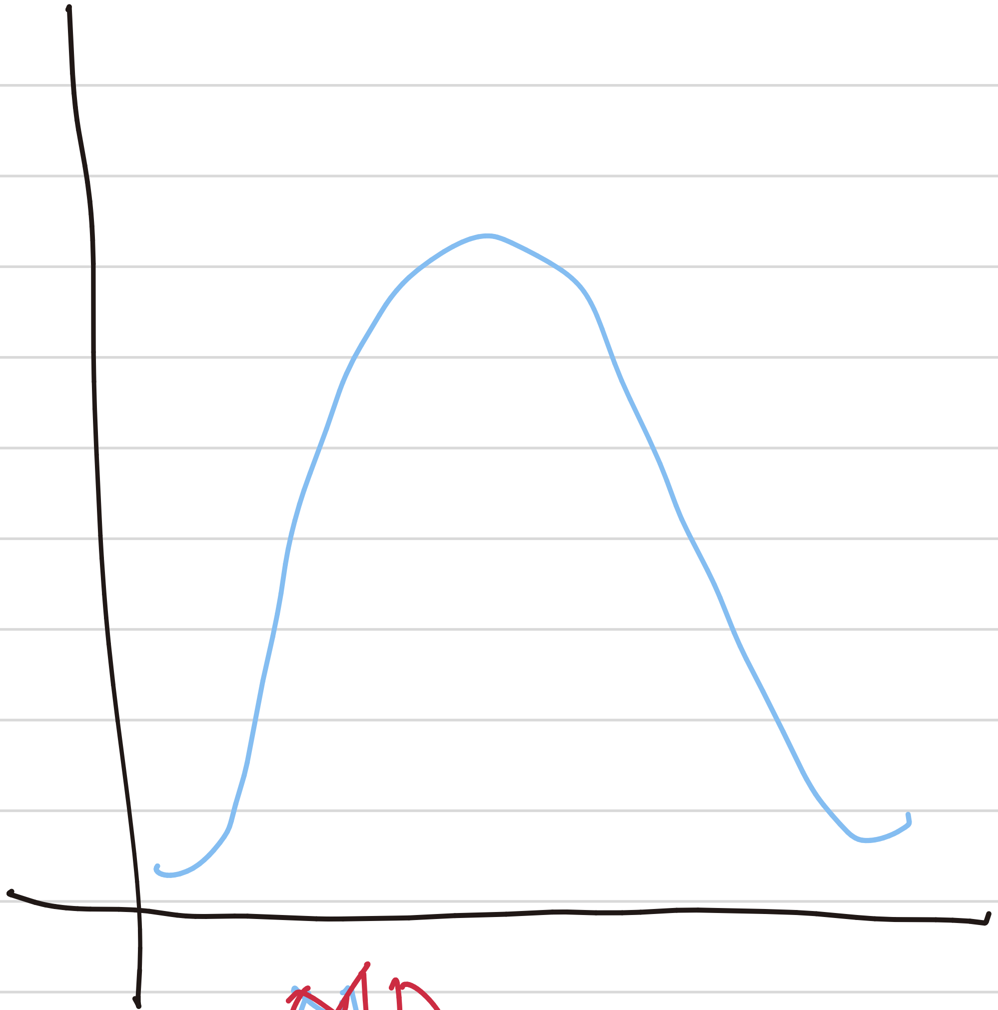
Figures of Conditional Probability

$$P(Y=y \mid X=x)$$

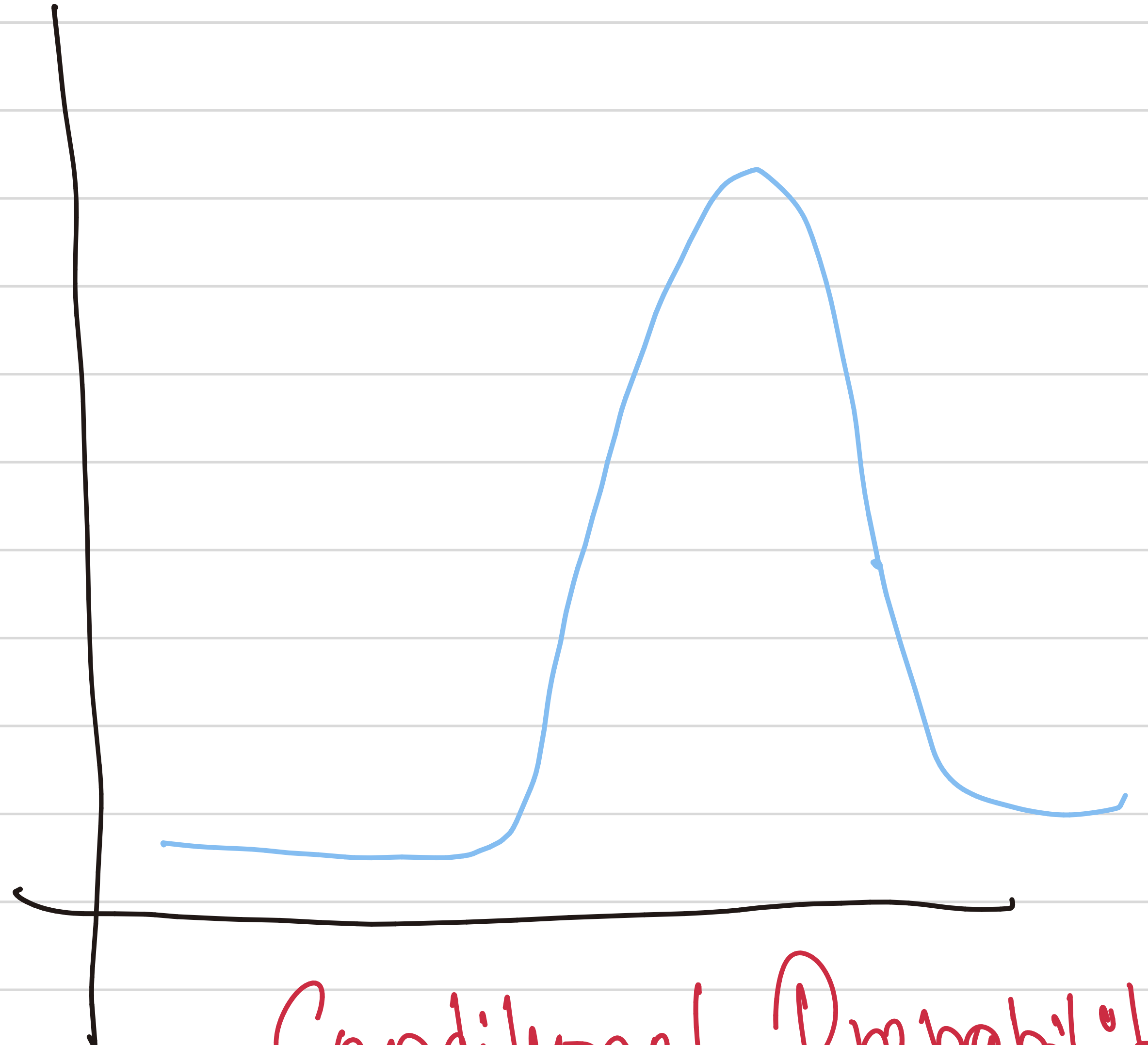
$$\Rightarrow \frac{P(Y=y, X=x)}{P(X=x)}$$



d) Bivariate Gaussian.



MD



Conditional Probability

#. So for a given fix (x) or (y) we slice the Probability as $P(Y=y \mid X=x) \Rightarrow \dfrac{P(Y=y, X=x)}{P(X=x)}$