#Day 5 - Oct 7, 2024.

Chain Rule (Product Rule for Probabilities)

Showing example of 3 variables

$$p(a, b, c) = p(a \mid b, c)\, p(b \mid c) \cdot p(c)$$

Generally the eqn - 5.1 of Chain Rule -

$$p(x^1 \ldots x^n) = p(x^1) \prod_{i=2}^{n} p(x^i \mid x^1 \ldots\ldots x^{i-1})$$

$P(A|B) \rightarrow$ probability of A given B

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

$\therefore P(A \cap B) \Rightarrow$ Both events A and B occuring Probability

$\therefore P(B) \Rightarrow$ probability that event B occurs and must be greater than 0.

Suppose A is numbers or even numbers, $B \geq 3$. if event B has occured $P(A/B)$ what is 'p' of event A ?

Here,

$A = \{2, 4, 6\}$

$B = \{3, 4, 5, 6\}$

$\therefore (A \cap B) = 2$

So, the conditional probability $P(A/B)$ is :

$$P(A/B) \Rightarrow \frac{P(A \cap B)}{P(B)} \Rightarrow \frac{2}{3} > 0$$

$P(A|B, C)$ = Probability of A Occuring given that both B and C occured

$$P(A|B, C) \Rightarrow \frac{P(A \cap B \cap C)}{P(B \cap C)}$$

$\therefore$ $P(A \cap B \cap C)$ is the probability that all events occur.

$\therefore$ $P(B \cap C) \Rightarrow$ probability of both event B and C occur together (must be $P(B \cap C) > 0$)

Example:-

A = Card is a King

B ⇒ Card is a red

C ⇒ the card is from the hearts suit

Lets find $\frac{P(A \cap B \cap C)}{P(B \cap C)} \Rightarrow P(A/B \mid C)$ finding the probability that the card is a King, given that it is both red and from the hearts suit.

- $A \cap B \cap C \Rightarrow 1/52$ (a King which is red and hearts suits)

$P(B \cap C) \Rightarrow$ Probability of red card from the hearts suit is $\frac{13}{52}$ (since there are 13 hearts)

Thus,

$$P(A/B, C) \Rightarrow \frac{P(A \cap B \cap C)}{P(B \cap C)} \Rightarrow \frac{1/52}{13/52}$$

$$\therefore P(A|B, C) \Rightarrow 1/13 > 0$$

$\therefore$ Probability of drawing Red Hearts Suit King is $1/13$

Q. What is the probability you will reach College on time ?

OT ⇒ On time is a binary variable

OT ⇒ 0 means not on time

OT ⇒ 1 means on time

Same for TJ ⇒ Traffic Jam is also a binary variable

0 ⇒ NO TJ

1 ⇒ Yes TJ

- On a regular day there is 70% chance, that there will be jam.

$P(TJ = 1) \Rightarrow 0.7$

$P(TJ = 0) \Rightarrow 0.3$

$\therefore P(OT \mid TJ) \Rightarrow \dfrac{P(OT \cap TJ)}{P(TJ)}$

$\therefore OT \Rightarrow ?$

Traffic Jam (TJ = 1)

20% you'll reach on time

$$P(OT = 1 \mid TJ = 1) = 0.2$$

So,

$$P(OT = 0 \mid TJ = 1) = 0.8 \quad = 1 - P(OT = 1 \mid TJ = 1)$$

No Traffic Jam (TJ = 0) 90% reaching on time

$$P(OT = 0 \mid TJ = 0) = 0.1$$

$$P(OT = 1 \mid TJ = 0) = 0.9$$

P (TJ, OT) table (uses Chain Rule)

	TJ => 0	TJ => 1
OT => 0	0.3 * 0.1 => 0.03	0.7 * 0.8 => 0.56
OT => 1	0.3 * 0.9 => 0.27	0.7 * 0.2 => 0.14

∴ TJ and Reaching College on time => 0.14.

∴ TJ and Not Reaching College on time => 0.56

Chain Rule => Decompose the Joint Probability into Single elements of probability see the eqn 5.1

Expectation in Stats

In DL we must minimize the loss function's Expectation with some respect to parameters in ANN.

In dominate term, Expectation is the average.

Definition: Expected value of a function $g: \mathbb{R} \rightarrow \mathbb{R}$ of a univariate continuous random variable $X \sim p(x)$ is given by

$$E_X[g(x)] = \int_X g(x) \cdot p(x) \cdot dx$$

Correspondingly, the expected value of a function g of discrete random variable $X \sim p(x)$ is given by

$$\mathbb{E}_X[g(x)] = \sum_{x \in \mathcal{X}} g(x) \cdot p(x)$$

Where $\mathcal{X}$ is a set of possible outcomes (the target space) of the random variable X.

- Expected Random Variable's value is its average value

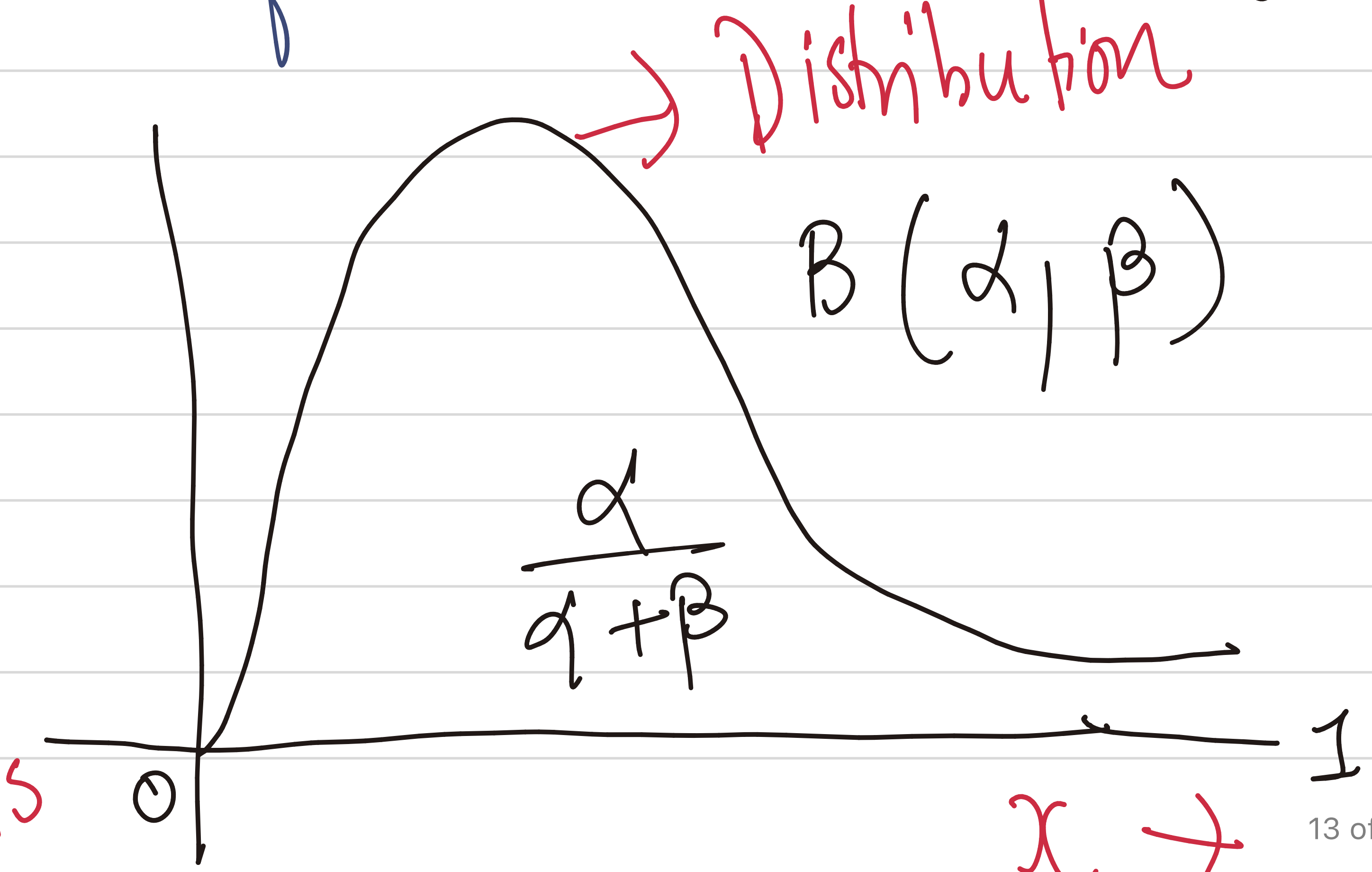
$$E_x[x] = \int x \, p(x) \cdot dx$$

(Expectation of x is mean of x

Geometric Intuition: Centre of mas

Beta distribution with

mean $\Rightarrow \dfrac{\alpha}{\alpha+\beta}$



Finding Center of mass at x-axis

Expectation is linear

Let $f(x) = ag(x) + bh(x) \qquad x \in \mathbb{R}^D$

$$E_x[f(x)] \Rightarrow \int f(x) \cdot px \cdot dx$$

$$\Rightarrow \int [ag(x) + bh(x)] \cdot px \cdot dx$$

$$\Rightarrow a \int g(x) \cdot px \cdot dx + b \int h(x) \cdot px \cdot dx$$

$$\Rightarrow a\, E_x[g(x)] + b\, E_x[h(x)]$$

Linear with combination of other two functions $g(x)$ and $h(x)$.

E_x =) Expectation

$f(x)$ =) function

g and h are the density

$g(x)$ and $h(x)$ are the density function

$E_x[f(x)]$ is equivalent to $E_x\, g(x)$ and $E_x\, h(x)$

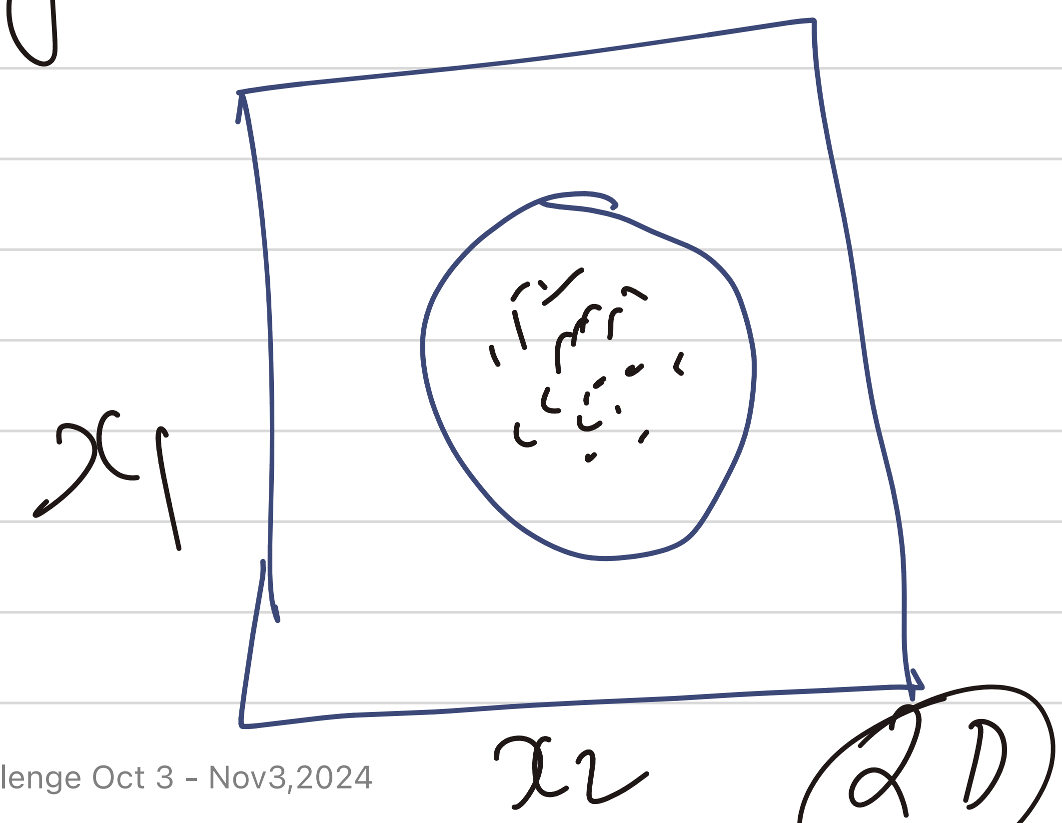
Independence

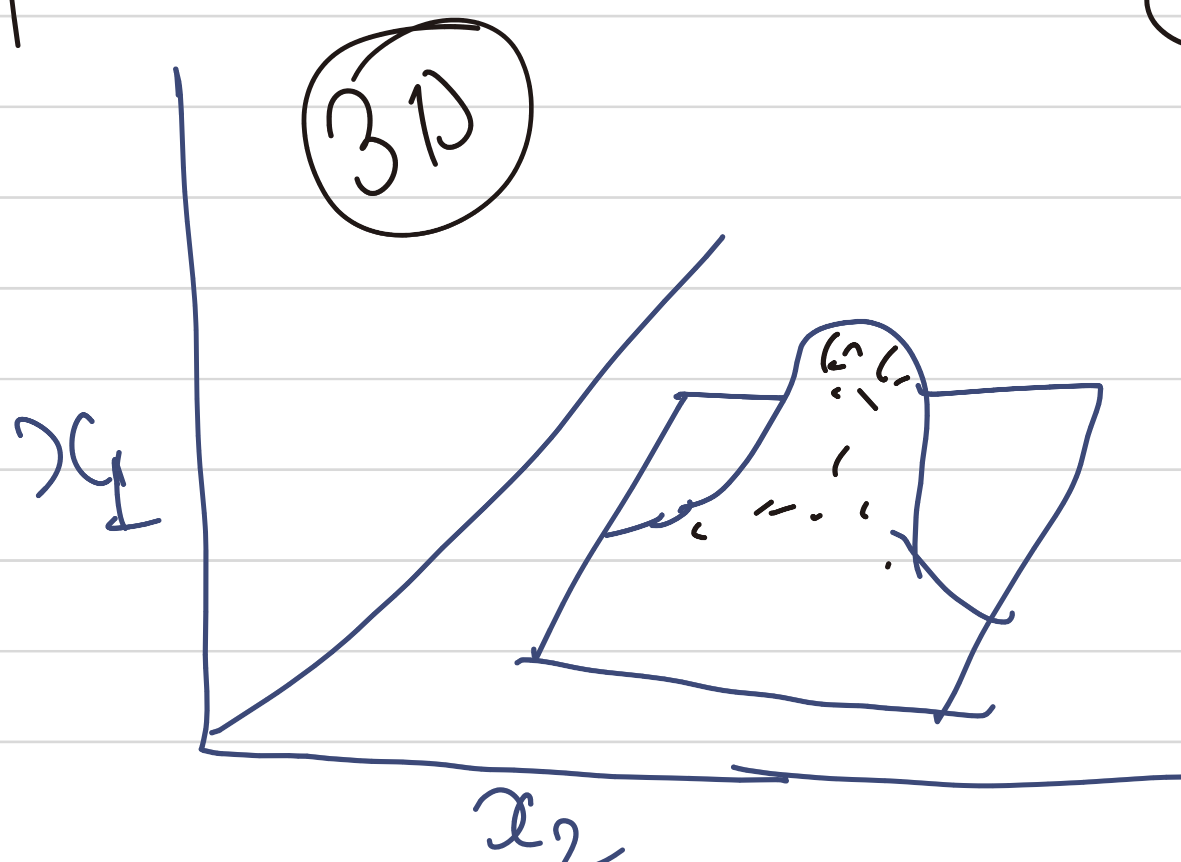
When two events X and Y are independent, we cannot infer anything about X when we know Y.

- Two random variables that are not related at all.

for an example: rolling a dice and drawing a card.

- Even if two events are related physically, they can still be independent statistically.





- Two random variables X and Y are independent if

$$p(X=x, Y=y) \Rightarrow p(X=x)\, p(Y=y)$$

$$p(x,y) \Rightarrow p(x) \cdot p(y)$$

It implies Conditional is same as marginal

$$p(x,y) = p(y|x)\, p(x)$$

$$\Rightarrow p(x)\,p(y)$$

$$\Rightarrow P(y|x) = p(y)$$

So, Independence means product JPD and MP.

Variance: Spread of a things or data set. Let's dive into the property of variance.

- $V(X+Y) \Rightarrow V(X) + V(Y)$ if X and Y are independent

$$V(X+Y) \Rightarrow E\left[\left(X+Y - E[X+Y]\right)^2\right]$$

$$\Rightarrow E\left[\left(X - E[X] + Y - E[Y]\right)^2\right]$$

$$\Rightarrow E\left[\left(X - E[X]\right)^2 + \left(Y - E[Y]\right)^2 + 2\left(X - E[X]\right)\left(Y - E[Y]\right)\right]$$

$$\Rightarrow E\left[(X - E[X])^2\right] + E(Y - E[Y])^2 +$$

$$2E\left[(X - E[X])(Y - E[Y])\right]$$

$$\Rightarrow V(X) + V(Y) + 2E[X - E(X)]\,E[Y - E[Y]]$$

→ X and Y are Independent ; term vanishes.

So are $X - E[x]$ and $Y - E[Y]$;

Example : Variance of Bernoulli

- By Definition:

$$V_X[x] = E_X[x^2] - \left(E[x]\right)^2$$

- for a Bernoulli distribution,

$$E[x] = \phi$$

$$E_X[x^2] \Rightarrow 1^2 \cdot p(X=1) + 0^2 \cdot p(X=0)$$

$$\Rightarrow \phi$$

Variance is given by

$$V_X[x] \Rightarrow \phi - \phi^2$$

$$\Rightarrow \phi(1-\phi)$$

Covariance :

Definition (Covariance Univariate):
The covariance between two univariate random variables $X, Y \in \mathbb{R}$ is given by the expected product of their deviations from their respective means i.e -

$$\mathrm{Cov}_{X,Y}[X,Y] := E_{X,Y}\left[(X - E_X[x])(y - E_Y[y])\right].$$

or equivalenty

$$\mathrm{Cov}[x,y] = E[xy] - E[x] \cdot E[Y]$$

Definition: (Covariance Multivariate):

If we consider two multivariate random variables X and Y with states $x \in \mathbb{R}^D$ and $y \in \mathbb{R}^E$ respectively, the Covariance between X and Y is defined as:

$$\mathrm{Cov}[x, y] = \mathbb{E}[xy^T] - \mathbb{E}[x]\,\mathbb{E}[y]^T$$

$$\Rightarrow \mathrm{Cov}[y, x]^T \in \mathbb{R}^{D \times E}$$

Co-variance Matrix :

the variance of a random variable X with states $x \in \mathbb{R}^D$ and a mean vector $\mu \in \mathbb{R}^D$ is defined as:

$$V_x[x] \Rightarrow \mathrm{Cov}_X[x, x]$$

$$\Rightarrow E_x\left[(x-\mu)(x-\mu)^T\right]$$

$$\Rightarrow \mathbb{E}_X\left[x x^T\right] - \mathbb{E}_X[x]\,\mathbb{E}_X[x]^T$$

$$\Rightarrow \begin{bmatrix} \mathrm{Cov}[x_1, x_2] & \mathrm{Cov}[x_1, x_2] & \cdots & \mathrm{Cov}[x_1, x_D] \\ \mathrm{Cov}[x_2, x_1] & \mathrm{Cov}[x_2, x_2] & \cdots & \mathrm{Cov}[x_2, x_D] \\ \vdots & \vdots & & \vdots \\ \mathrm{Cov}[x_D, x_1] & \cdots & \cdots & \mathrm{Cov}[x_D, x_D] \end{bmatrix}$$

Empirical Mean and Covariance

We have Samples $x_1, x_2 \cdots x_n$ from some distribution

- Empirical Mean

$$\bar{x} := \frac{1}{N} \sum_{n=1}^{N} x_n$$

- Empirical Covariance:

$$\Sigma := \frac{1}{N} \sum_{n=1}^{N} (x_n - \bar{x})(x_n - \bar{x})^T$$

Note: No need to know distributions to compute empirical mean and Σ (variance).

Co-variance: How x_1 is related to y_1 or x_2

Variance: Spreadness and variety of x_1 and x_2

Example of Co-variance means how two variables related and varied

→ Price of icecream and rise in temperature

Summary

1) Conditional Probability finds the probability of A given B event has already occured so

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

For 3 variables

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

2) Product Rule/Chain Rule Compute the Probability of multiple events or Joint Probability, decompose JP into a Sequence and Compute Combined Probability

So, $P(A \cap B) = P(A/B) \cdot P(B)$

In Generalization terms using Chain Rule to get Combined (P)

$$P(A_1 \cap A_2 \cap \cdots\cdots A_n) = P(A_1) \cdot P\left(\frac{A_2}{A_1}\right) \cdot P\left(\frac{A_3}{A_1 \cap A_2}\right) \cdot$$

$$P\left(\frac{A_n}{A_1 \cap A_2 \cap \cdots A_{n-1}}\right)$$

Expected Value (Expectation) $E[x]$

→ average or mean when random variable takes for trials

→ E_x for both discrete and continuous so for discrete we use Σ whereas for continuous we use $\int$ both using $f(x)$ or $g(x)$ representing PDF or weighted x for discrete.

Notations used by Expectations

- $E[x]$ → Expected value of X
- $E[X|Y]$ → Expected Conditional Probability of X given Y
- $\sum_i x_i P(x_i)$ → Expectation for discrete random variables

- $\int x f(x) \cdot dx \rightarrow$ Expectation for Continuous Random Variables

Variance:

How much random variable deviates from the mean for a random variable X

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

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

Co-variance:

two random variable changes together or what measures of degree changes together so for X and Y

$$\text{Cov}(X, Y) = E[XY] - E[X] \cdot E[Y]$$

if $\text{Cov}(X, Y) = 0$ No relationship

Independence:

X and Y do not each other knowing

X doesnot help Y

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

Empirical Mean:- Average value of observed datasets

Empirical Co-variance:

Estimation of true Co-variance based on the observed data, helps to find the un-biased estimate of Covariance.

Bernoulli Co-variance:

Shows result in 0 or 1 and take discrete random variable 0 or 1, useful in logistic Regression

$p_X = 0.6$ (Rainy)
$p_Y = 0.7$ (Sunny)

$Cov(X,Y) = p(X=1, Y=1) - p_X \cdot p_Y$

$\Rightarrow 0.42 - (0.6 \times 0.7)$

$\Rightarrow 0$ [No-Relation]