Assignment-1

CMPUT-560

Submitted by :

- Afia Anjum
- ccid - afia
- Student Id - 1595188

Question-1- Answer

given that X is a random variable and

outcome space, $\Omega = \{a, b, c\}$.

$p(a) = 0.1$, $p(b) = 0.2$ and $p(c) = 0.7$.

Also given, $f(a) = 10$

$f(b) = 5$

$f(c) = 10/7$.

(a) Here, we can see that the random variable X can possess some discrete values that are given in the question.

so, $E[f(X)] = \sum_{x \in X} f(x) \cdot p(x)$

$= f(a) \cdot p(a) + f(b) \cdot p(b) + f(c) \cdot p(b)$

$= 10 \times 0.1 + 5 \times 0.2 + (10/7) \times 0.7$.

$= 3$ (Ans).

(b) Similarly we can define the following function as:

$E[1/p(X)] = \sum_{x \in X} \left(1/p(x)\right) \cdot p(x)$.

$= \left(1/p(a)\right) \cdot p(a) + \left(1/p(b)\right) \cdot p(b) + \left(\frac{1}{p(c)}\right) p(c)$

$= \left(\frac{1}{0.1}\right) \times 0.1 + \left(\frac{1}{0.2}\right) \times 0.2 + \left(1/0.7\right) \times 0.7$

$= 1 + 1 + 1$

$= 3$ (Ans).

(c) ~~for~~ An arbitrary pmf p. means that the outcome space is not specified with certain values, rathere we can extend the outcome space to an arbitrary number of random variables.

Hence we can write,
for an arbitrary pmf,

$$E[1/p(X)] = \sum_{x \in X} \frac{1}{p(x)} \times p(x)$$

$$= \sum_{i=1}^{n} \frac{1}{P_i(x_i)} \times p(x_i)$$

$$= \sum_{i=1}^{n} 1$$

$$= n \quad \text{(Ans)}$$

Question 2- Answer

Here, we are given that $X_1, X_2, \ldots X_m$ are independent multivariate Guassian random variables.

$$X_i \sim N(\mu_i, \Sigma_i)$$

$\mu_i \in R^d$ (mean) and $\Sigma_i \in R^{d \times d}$ (co-variance).

$$X = a_1 X_1 + a_2 X_2 + \cdots + a_m X_m \;; \quad a_i \geq 0$$

and $\sum_{i=1}^{m} a_i = 1$.

(a) So here X is expressed as a combination of a scalar and the random variables $X_1, X_2 \ldots$ $\ldots X_d$.

Also, μ_i is a $d \times 1$ vector, Σ_i is a $d \times d$ vector and $\sum_{i=1}^{m} a_i = 1$ means that here a is a scalar.

Now, since, $X = a_1 X_1 + a_2 X_2 + \ldots + a_m X_m$ (given)

so, our $E[X] = E[a_1 X_1 + a_2 X_2 + \ldots + a_m X_m]$

$= E[a_1 X_1] + E[a_2 X_2] + \ldots + E[a_m X_m]$

$= a_1 E[X_1] + a_2 E[X_2] + \ldots + a_m E[X_m]$

(using diff properties mentioned in notes) — (1).

A d-dimensional multivariate Guassian distribution can be defined as :

$$p(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}[(x-\mu)^T \Sigma^{-1} (x-\mu)]}$$

where μ & Σ are the parameters and $-\infty < x < \infty$

Therefore to calculate $E[X_i]$ we have to write,

$$E[X_i] = \int_{-\infty}^{\infty} x_i \, p(x_i) \, dx.$$

$$= \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \, e^{-\frac{1}{2}\left[(x_i-\mu)^T \Sigma^{-1} (x_i-\mu)\right]} \right) \cdot x_i \cdot dx$$

let, $z = x - \mu$

$$\Rightarrow x = z + \mu.$$

$$\therefore E[X_i] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \, e^{-\frac{1}{2} z^T \Sigma^{-1} z} \cdot (z_i + \mu_i) \cdot dz$$

$$= \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \int_{-\infty}^{\infty} z_i \, e^{-\frac{1}{2} z^T \Sigma^{-1} z} \cdot dz$$

$$+ \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \int_{-\infty}^{\infty} \mu_i \, e^{-\frac{1}{2} z^T \Sigma^{-1} z} \cdot dz$$

Here, $\frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \int_{-\infty}^{\infty} z_i \, e^{-\frac{1}{2} z^T \Sigma^{-1} z} \cdot dz$ is an odd function. According to wikipedia, it's integral will be equal to zero.

$$\therefore \; E[X_i] = 0 + \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \int_{-\infty}^{\infty} \mu_i \, e^{-\frac{1}{2} z^T \Sigma^{-1} z} \cdot dz$$

$$= 0 + \mu_i \times 1$$

$\left[\because \int_{-\infty}^{\infty} \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} e^{-\frac{1}{2} z^T . \Sigma^{-1} z} . dz\right.$ is a ~~probability~~ probability mass function and integral of such pmf is equal to $\left.1.\right]$.

So, $E[X_i] = \mu_i$

Putting this value in equ^n. (1) we get,

$$E[X] = a_1 \mu_1 + a_2 \mu_2 + \dots + a_m \mu_m .$$

$$= \sum_{i=1}^{m} a_i \mu_i \quad \text{(Ans)}$$

As our μ_i is a ~~dxd~~ $d \times 1$ vector and our a_i is given to be a scalar. So, the dimension of ~~the~~ $E[X]$ will be the same as μ_i's dimension, which is d. (Ans)

(b) $cov[X] = ?$

as, $X = a_1 X_1 + a_2 X_2 + \cdots + a_m X_m$

Similarly we can write like the previous solution,

$$cov[X] = cov[a_1 X_1 + a_2 X_2 + \cdots + a_m X_m]$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{m} cov[a_i X_i, a_j X_j]$$

$$= \sum_{i=1}^{m} V[a_i X_i] + 2 \sum_{1 \leq i < j \leq m} cov[a_i X_i, a_j X_j]$$

Here, inside the covariance calculation, value of a does not mean anything, so we can ignore that,

$$\therefore cov[X] = \sum_{i=1}^{m} V[a_i X_i] + 2 \sum_{1 \leq i < j \leq m} cov[X_i, X_j]$$

$$= \sum_{i=1}^{m} V[a_i X_i] + 2 \times 0 \quad \left[\text{Since } X_i \text{ \& } X_j \text{ are independent, so } cov[X_i, X_j] = 0\right]$$

$$= \sum_{i=1}^{m} V[a_i X_i]$$

$$= \sum_{i=1}^{m} a_i^2 V[X_i]$$

$$= \left(\sum_{i=1}^{m} a_i^2\right) V[X_1 + X_2 + \cdots X_m]$$

$$= \left(\sum_{i=1}^{m} a_i^2\right) \Big\{ V[X_1] + V[X_2] + \cdots + V[X_m] + 2\,\mathrm{Cov}[X_1, X_2] + 2\,\mathrm{Cov}[X_2, X_3] + \cdots + 2\,\mathrm{cov}[X_{m-1}, X_m] \Big\}$$

$$= \left(\sum_{i=1}^{m} a_i^2\right) \Big\{ V[X_1] + V[X_2] + \cdots + V[X_m] \Big\}$$

[$\because$ $\mathrm{Cov}[X_i, X_j] = 0$ when X_i & X_j are independent]

$$= \left(\sum_{i=1}^{m} a_i^2\right) \Big\{ \mathrm{cov}(X_1, X_1) + \mathrm{cov}(X_2, X_2) + \cdots + \mathrm{cov}(X_m, X_m)$$

$$= \sum_{i=1}^{m} a_i^2 \cdot \mathrm{cov}(X_i, X_i)$$

$$= \sum_{i=1}^{m} a_i^2 \Sigma_i \qquad [\because \mathrm{Cov}[X_i, X_i] = \Sigma_i]$$

(Ans)

Now, since a_i is a scalar quantity and Σ_i is a $d \times d$ vector, so the dimension of covariance, $\mathrm{cov}[X]$ is $d \times d$. (Ans)

Now if X_1 and X_2 are not independent, so $cov[X_1, X_2]$ is ~~ne~~ a non-zero and according to question, $cov[X_1, X_2] = \Lambda$ where $\Lambda \in R^{d \times d}$

So, $cov[X] = \sum_{i=1}^{m} V[a_i X_i] + 2 \underbrace{\sum_{1 \le i < j \le m} Cov[X_i, X_j]}$

in this part, $cov[X_1, X_2]$ will be equal to Λ and not zero like previo

Hence, the overall value of $cov[X]$ will be equal to $cov[X] = \sum_{i=1}^{m} a_i^2 \Sigma_i + 2 [cov[a_1 X_1, a_2 X_2]]$

$= \sum_{i=1}^{m} a_i^2 \Sigma_i + 2 a_1 a_2 \Lambda$

(Ans)

Here the solution is not complete, as we do not know the value of covariance between other random variables.

Question - 3 - Answer

(a) Keeping dim = 1 and varying σ and number of samples we ~~ea~~ have to observe the output of the code.

Dim = 1, σ = 1.0

no. of samples = 10, mean = 0.0111549
no. of samples = 100, mean = − 0.01839
no. of samples = 1000, mean = − 0.07585

Again when,
Dim = 1, σ = 10.0

no. of samples = 10, mean = −2.278
no. of samples = 100, mean = 0.5579
no. of samples = 1000, mean = 0.0783

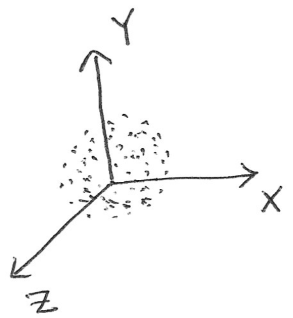
As the script which is provided, is a script written for guassian/normal distribution. In such distributions mean and mode values are placed in the same place. Also, in our script, the value of mean is given 0. As we run the code with different number of samples and different value of sigma (standard deviation), we can observe that with the increasing number of samples, value goes closer to zero (the actual mean provided in the script). Also when the standard deviation (σ) was small, the value of sample mean was concentrated tightly around the mean (μ). But, as the standard deviation (σ) increased, the sample mean is less tightly concentrated than it was before.

(Ans)

(b)

Given dim = 3 and given co-variance matrix,

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$



which can also be written as,

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix}$$

Here, $\sigma_1^2 = 1$, $\sigma_2^2 = 1$ and $\sigma_3^2 = 1$

while σ_{12}, σ_{13}, σ_{23} all are equal to zero, which suggests that the three random variables X, Y & Z all are independent. Also, putting the value of dim = 3 in the script, we see that the plot is scattered in all the directions.

(Ans)

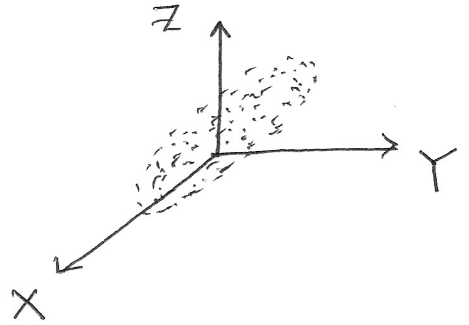
(c) Here dim = 3 and given co-variance matrix,

$$\Sigma = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix}$$

Here, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$ and also, $\sigma_{13} = \sigma_{31} = 1$ while $\sigma_{21} = \sigma_{12} = \sigma_{23} = \sigma_{32} = 0$, which suggests that random variables Y and Z and random variables X and Y are independent, while random variables X and Z are dependent, i.e. they have a co-relation.

By running the script, we can see that they are positively co-related, also, $\sigma_{13} = \sigma_{31} = 1 > 0$ suggest the same thing.

When the co-variance matrix is updated with the aforementioned value, the plots are seen to be scattered along the X-Z plane. and also, the value of μ becomes -0.000299.



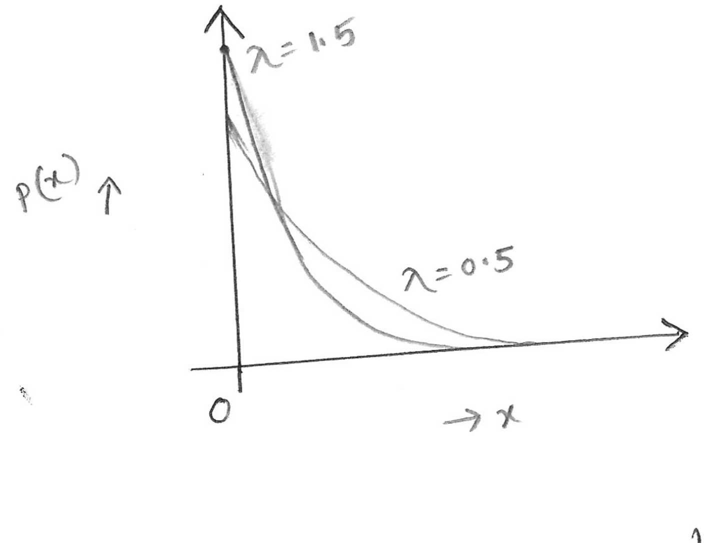
(Ans)

Question 4 – Answer

(a) Here the most likely value for λ can mean two different things. It mainly means the mode of the distribution.

But in other words, in general thinking, the most likely value is the value that one expects to happen on an average.

Since this is an exponential distribution, and the range of λ is given from 0 to ∞, so we can say that the mode of a continuous random variable in an exponential distribution is 0.



Mode = 0.

(Ans)

On the other hand, if the most likely value corresponds to the average value, then we have to deduce the expected value.

given $p(\lambda) = \theta e^{-\theta\lambda}$.

so, $E[\lambda] = \int_0^\infty \lambda p(\lambda)\, d\lambda$

$= \int_0^\infty \lambda \theta e^{-\theta\lambda}\, d\lambda$

$= \int_0^\infty y e^{-y} \frac{1}{\theta} \cdot dy$

$= \frac{1}{\theta} \int_0^\infty y e^{-y}\, dy$.

$= \frac{1}{\theta} \left[\left[-e^{-y} y \right]_0^\infty + \int_0^\infty e^{-y}\, dy \right]$

$= \frac{1}{\theta} \left\{ \left[-ye^{-y} \right]_0^\infty + \left[-e^{-y} \right]_0^\infty \right\}$

$= \frac{1}{\theta} \left[-ye^{-y} - e^{-y} \right]_0^\infty$

$= \frac{1}{\theta} \left\{ \lim_{x \to \infty} \left[-xe^{-x} - e^{-x} \right] - \left[-0e^{-0} - e^{-0} \right] \right\}$

$= \frac{1}{\theta} \left\{ [0-0] - [0-1] \right\}$

$= 1/\theta = 1/\tfrac{1}{2} = 2$ (Ans)

Let $y = \lambda\theta$.

$\frac{dy}{d\lambda} = 0$

$\Rightarrow d\lambda = \frac{1}{\theta} dy$

Limit,

$\lambda = 0, y = 0$

$\lambda = \infty, y = \infty$

Again,

$\int f dg = fg \Big|_0^\infty - \int_0^\infty g df$

$f = y$ | $g = -e^{-y}$

$df = dy$ | $dg = e^{-y}$

(b) given, $p(\lambda) = \theta e^{-\theta\lambda}$

we have to determine $\lambda_{MLE} = ?$

our $\lambda_{MLE} = \operatorname{argmax} p(D|\lambda)$

This distribution is followed by no. of accidents occuring daily is Poisson distribution,

so, $p(D|\lambda) = \sum_{i=1}^{n} p(D|\lambda)$

$= \sum_{i=1}^{n} p(x_i|\lambda)$

$= \sum_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$ —— (1)

Without summation part, taking ln on both sides of (1),

$\ln p(D|\lambda) = \ln\left(\frac{e^{-\lambda}\lambda^{x}}{x!}\right)$

$= \ln e^{-\lambda} + \ln \lambda^{x} - \ln x!$

$\Rightarrow \ln p(D|\lambda) = -\lambda + x\ln\lambda - \ln x!$

Putting back summation,

$$\ln p(D|\lambda) = \sum_{i=1}^{n} -\lambda + \sum_{i=1}^{n} x_i \ln\lambda - \sum_{i=1}^{n} \ln x!$$

$$= -n\lambda + \sum_{i=1}^{n} x_i \ln\lambda - \sum_{i=1}^{n} \ln x! \quad \text{—}(2)$$

Taking derivative on both sides w.r.to λ in eqⁿ (2)

$$\frac{d}{d\lambda}\{\ln p(D|\lambda)\} = -n + \sum_{i=1}^{n} x_i \frac{1}{\lambda} \quad \text{—}(3)$$

For finding stationary point,

$$0 = -n + \sum_{i=1}^{n} x_i \frac{1}{\lambda}$$

$$\Rightarrow n = \sum_{i=1}^{n} \frac{x_i}{\lambda}$$

$$\Rightarrow \lambda = \sum_{i=1}^{n} \frac{x_i}{n}$$

$$\therefore \lambda_{MLE} = \sum_{i=1}^{n} \frac{x_i}{n}$$

So putting value, $\lambda_{MLE} = \frac{79}{9} = 8.78$ (Ans)

(c) Here, given, $p(\lambda) = \theta e^{-\theta\lambda}$

and $\lambda_{MAP} = \arg\max \; p(D|\lambda)\, p(\lambda)$ — (1)

From the previous solution we got,

$$\frac{d}{d\lambda}\{\ln p(D|\lambda)\} = -n + \sum_{i=1}^{n} x_i \frac{1}{\lambda} \quad \text{— (2)}$$

Now, $P(\lambda) = \theta e^{-\theta\lambda}$

$= \frac{1}{2} e^{-\lambda/2}$ — (3)

Taking ln on both sides,

$$\ln p(\lambda) = \ln\theta + \ln e^{-\theta\lambda}$$

$$= \ln\theta - \theta\lambda.$$

$$= \ln(1/2) - \lambda/2$$

$$= \ln(2^{-1}) - \lambda/2$$

$$= -\ln 2 - \lambda/2$$

~~Taking~~ to determine the stationary point, we need to take it's derivative with r. to λ and equate with zero,

$$\frac{d}{d\lambda} \ln p(\lambda) = \frac{d}{d\lambda}\left(-\lambda/_2\right)$$

$$= -1/_2 \quad \text{———} (4)$$

Using equⁿ : (1), (2) & (4) we can write,

$$-n + \sum_{i=1}^{n} x_i \frac{1}{\lambda} - 1/_2 = 0.$$

$$\Rightarrow \sum_{i=1}^{n} x_i \frac{1}{\lambda} = n + 1/_2.$$

$$\Rightarrow \quad \lambda_{MAP} = \frac{\sum_{i=1}^{n} x_i}{n + 1/_2}$$

So putting value we get,

$$\lambda_{MAP} = \frac{79}{9 + 1/_2}$$

$$= \frac{79}{9.5}$$

$$= 8.315 \quad \text{(Ans)}$$

(d) As we are said that our no. of accidents will have poisson distribution, so we can write, $p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$ where λ is the mean.

If we want to predict or determine the number of accidents for tomorrow, we have to determine the Expected value of the distribution.

$$\text{so, } E(X) = \sum_{x \in X} x \cdot p(x)$$

$$= \sum_{x \in X} x \cdot \frac{\lambda^x e^{-\lambda}}{x!}$$

$$= \sum_{x \in X} \frac{x \lambda^x e^{-\lambda}}{x(x-1)!}$$

$$= \sum_{x \in X} \frac{(\lambda)(\lambda)^{x-1} e^{-\lambda}}{(x-1)!}$$

$$= \lambda \sum_{x \in X} \frac{\lambda^{x-1} e^{-\lambda}}{(x-1)!}$$

$$= \lambda .$$

So; if we need to estimate no. of accidents for tomorrow, we need $E[X] = ?$

As we calculated already, $E[X] = \lambda$

So, when we use the λ_{MLE} then, expected value for the number of accidents tomorrow $= 8.77$

And when we use the λ_{MAP} then the expected value for the number of accidents tomorrow becomes $= 8.315$

They both predict the number of accidents, where MLE does not use the prior information but MAP does.

(Ans)

(e) As we know that, if we have large amount of data, we do not need to know the prior knowledge. But if we have very small number of data, then the knowledge of prior ~~is~~ becomes important and the MAP estimates, in that case the most correct value, as close as possible. So, if we have a good prior, then MLE estimate would be more reliable, given that we are provided with limited number of data. Ans)

(f) Due to safety measures, number of accidents occurring daily would decrease, which suggests decrease in value of λ.

From solution 4 (a), we ~~see~~ observed that, when $\theta = 1/2$, our expected value or expected mean was, $E[\lambda] = 2$.

Again putting $\theta = 1/4$, our $E[\lambda] = 4$; this is something we do not want.

Our main goal here is to decrease the λ, so in that case, we have to increase the value of θ.

For $\theta = 2$, our $E[\lambda] = 1/2$.

~~So~~ This is because when θ falls, then λ rises, which causes our probability density function for exponential distribution to rise. and when θ increases, our λ decreases which cause a sharp decrease in the plot.

(Ans)

Question 5 - Answer

(a) In order to predict whether the table is free in our favourite restaurant, we need to know about two conditions, whether it was sunny or a not sunny day.

We have,

$p(\text{sunny}) = 1$ or, $P(S) = 1$

$p(\text{not sunny}) = 0$. or, $P(NS) = 0$.

$p(\text{free}) = 1$. or, $P(F) = 1$

$P(\text{not free}) = 0$ .or $P(NF) = 0$.

For Bernouli distribution, we can write,

Now, Free, $F = \begin{cases} 0 & \text{not free} \\ 1 & \text{free} \end{cases}$

and Sunny, $S = \begin{cases} 0 & \text{not sunny} \\ 1 & \text{sunny} \end{cases}$

$$P(S) = \begin{cases} \alpha & \text{if } S = 1 \\ 1-\alpha & \text{if } S = 0 \end{cases} \qquad \alpha \in (0,1).$$

$$p(S;\alpha) = \alpha^{S}(1-\alpha)^{1-S}$$

$$p(F \mid S=1) = \begin{cases} a & \text{if } F = 1 \\ 1-a & \text{if } F = 0 \end{cases}$$

$$p(F \mid S=0) = \begin{cases} b & \text{if } F = 1 \\ 1-b & \text{if } F = 0 \end{cases}$$

y \ x	Sunny	not sunny
free	a	b
not free	$1-a$	$1-b$

So MLE of $P(D \mid \lambda) = \prod_{i=1}^{n} P(F_i, S_i)$.

$$= \prod_{i=1}^{n} P(F_i, S_i)\, P(S_i).$$

$$= \left[\alpha^{\#\text{sunny}} (1-\alpha)^{\#\text{not sunny}}\right] \times$$

$$\left[a^{\#(F,S)} (1-a)^{(NF,S)}\right] \times$$

$$\left[b^{\#(F,NS)} (1-b)^{\#(NF,NS)}\right.$$

(Ans)

(b) Having data for the last 10 days

and having MLE solutions from the 5(a) part.

We can make the predictions as,

$$P(F=1|S) = a_{MLE}$$ if $a > 0.5$ then

the table would be free.

Again, $P(F=0|S) = b_{MLE}$, if $b > 0.5$ then

the table would be free.

the table would be free. the MLE

This prediction is made looking at the MLE

that is formulated for part 5(a).

(c) Adding another information to the previous

5(a) part, now we have another three outcomes

So the Maximum likelihood problem would change

in the following way,

$$\text{Free}, F = \begin{cases} 0 & \text{not free} \\ 1 & \text{free} \end{cases}$$

$$\text{Sunny}, S = \begin{cases} 0 & \text{not sunny} \\ 1 & \text{Sunny} \end{cases}$$

x	$P(x)$
sunny	a
not sunny	1−a

z	P(z)
morn	b
even	c
afternoon	1−b−c

y \ xz	sunny			not sunny		
	Morn	even.	afternoon	Morn	even.	afternoon
free	d	e	f	g	h	i
not free	1−d	1−e	1−f	1−g	1−h	1−i

This is how the probability would change adding ~~this~~ the time of day in the information. We can still use Bernouli for ~~this~~ representing and estimating the maximum likelihood.

(Ans)

(d) Here given $d \in \{1, 2, 4, 8, 16, 32, 64, 128, 256\}$

Using np.identity (d), the covariance matrix is generated and using np.zeros (d), the mean matri is generated for d-dimensions.

We have used the np.random.multivariate_normal from the numpy package and used it to generate data samples from this distribution. We have used the l_2 distance (Euclidean distance) to generate the average distance from each of the samples to the origin.

dimension	l_2 distance
d = 1	-1.17×10^{-1}
d = 2	7.31×10^{-1}
d = 4	1.792
d = 8	2.651
d = 16	3.957
d = 32	5.542
d = 64	7.68
d = 128	11.215
d = 256	15.497

We can observe from the output that with the increase in dimension, average distance between each sample and the origin also increases. It suggests that the ratio of the sample data availability and the dimension is very low, which indicates that most of the spaces between these dimensions are empty.

Using k-means clustering, we can not expect good performance because it does not work well in high dimension. The reason is, with the increase in dimensions, distance between a centre of a cluster and a sample also increases. Any distance metric is not expected to perform good in this condition. If the growth of data ~~is~~ remains directly proportional to the dimensions, then the situation will change.

(e)

Given,

Length of side of cube1 = a1 = 1
Length of side of cube 2 = a2 = $1-\epsilon$ $(0<\epsilon<1)$

We know that, if the dimension of a hypercube is d, ~~then~~ and it's one side is a, then it's volume becomes:

$$a^d$$

So, $\dfrac{\text{vol}^m \text{ of 2nd hypercube}}{\text{Vol}^m \text{ of 1st hypercube}}$

$$= \frac{(1-\epsilon)^d}{1^d}$$

$$= (1-\epsilon)^d \quad \text{[Ratio of the volumes]}$$

As the value of d increases, the ratio will change as following:

$0<\epsilon<1$

ϵ	dimension	Ratio
0.635	1	0.365
0.536	2	0.215
0.272	3	0.385
0.238	128	7.76×10^{-16}
0.373	256	1.26×10^{-52}

Here value of ϵ is generated by using random function importing random modu between the values 0 & 1.

So from the above table, we can observe that change in ratio increasing the dimensions. When dimension d would increase, the volume of the first hypercube would always be the Same. But the second hypercube volume would vary and the difference in volume between these two hypercubes would be bigger and bigger. It also suggests that the distance between these cubes will also get increased. We had similar observation in part 5(d). In between these distance Sample points would be scattered but most of the spaces between them would be empty because ϵ ranges from 0 to 1.