CMPUT 466 A1
Yuhan Ye 1463504

Q1 :

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

$= 10 \cdot 0.1 + 5 \cdot 0.2 + \frac{10}{7} \cdot 0.7 = 3$

b) $E[1/p(x)] = \sum_{x \in X} 1/p(x) \cdot p(x) = \sum_{x \in X} 1 = 3$

c) For an arbitrary pmf p. $E[1/p(x)] =$ the number of elements in the outcome space

Q2 :

a) $E[X] = E[a_1 X_1 + a_2 X_2 + \cdots a_m X_m]$

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

$= a_1 \mu_1 + a_2 \mu_2 + \cdots a_m \mu_m$

$= \sum_{i=1}^{m} a_i \mu_i$

the dimension of $E[X]$ is d

b) $Cov[X] = Cov(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{m} a_j X_j) = \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j Cov(X_i, X_j)$

Since $X_1, X_2 \cdots X_m$ are independent,

$$Cov(X_i, X_j) = \begin{cases} 0, & i \neq j \\ \Sigma_i, & i = j \end{cases}$$

$Cov[X] = a_1^2 \Sigma_1 + a_2^2 \Sigma_2 + \cdots a_m^2 \Sigma_m$

$= \sum_{i=1}^{m} a_i^2 \Sigma_i$ ~~with dimension $d \times d$~~ with dimension d

if X_1, X_2 are dependent, $Cov[X_1, X_2] \neq 0$

$Cov[X_1, X_2] = E[(X_1 - \mu_1)(X_2 - \mu_2)]$

since for X_i, $Cov[X_i] = \Sigma_i \in R^{d \times d}$

$Cov[X_1, X_2] = \Lambda$ for $\Lambda \in R^{d \times d}$

Q3:

a) dim = 1

σ \ sample #	10	100	1000
1	-0.151	-0.004	-0.074
10	-3.518	-1.378	0.131

We can find that sample mean when $\sigma = 1$ is closer to 0 than sample mean when $\sigma = 10$

b) Since the covariance matrix is an identity matrix, X, Y, Z are independent

c) samples lie on the plane $X = Z$

Q4:

a) $\lambda = 0$ when there is no data

$p(\lambda)$ decreases monotonically as λ increase, $\lambda \in [0, +\infty)$

$p(0) \Rightarrow \lambda = 0$ has maximum probability

b) $p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$

$\lambda_{MLE} = \underset{\lambda \in [0,+\infty)}{\operatorname{argmax}} \; p(D|\lambda)$

$$p(D|\lambda) = p(\{x_i\}_{i=1}^{n} | \lambda) = \frac{\lambda^{\sum_{i=1}^{n} x_i} \cdot e^{-n\lambda}}{\prod_{i=1}^{n} x_i!}$$

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

$$\frac{\partial ll(D|\lambda)}{\partial \lambda} = \frac{1}{\lambda}\sum_{i=1}^{n} x_i - n = 0$$

$$\lambda_{MLE} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{79}{9}$$

c) $\lambda_{MAP} = \underset{\lambda \in [0, +\infty)}{\operatorname{argmax}} \; p(\lambda | D) = \underset{\lambda \in [0,+\infty)}{\operatorname{argmax}} \; p(D|\lambda) p(\lambda)$

$\theta = \frac{1}{2}$

$$p(\lambda | D) = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!} \cdot \frac{1}{2} e^{-\frac{1}{2}\lambda}$$

$$ll(D|\lambda) = \ln p(\lambda|D) = \ln\lambda \sum_{i=1}^{n} x_i - n\lambda \sum_{i=1}^{n} \ln(x_i!) + \ln(\tfrac{1}{2}) - \tfrac{1}{2}\lambda$$

$$\frac{\partial ll(D|\lambda)}{\partial \lambda} = \frac{1}{\lambda} \sum_{i=1}^{n} x_i - n - \frac{1}{2} = 0$$

$$\lambda_{MAP} = \frac{1}{n+\frac{1}{2}} \sum_{i=1}^{n} x_i = \frac{158}{19}$$

d) For MLE, we now know that $\lambda_{MLE} = \frac{79}{7}$, the number of accidents follow the Poisson Distribution with mean $\lambda_{MLE} = 79/7$.

For MAP, it will follow the Poisson Distribution with mean $\lambda_{MAP} = 158/19$

They can predict probability of each number of accidents that may happen tomorrow. $p(X=1)$, $p(X=2)$ ....

e) The prior is used to incorporate prior knowledge and helps to model the estimate distribution with mean λ_{MAP} since $p(\lambda|D) \propto p(D|\lambda)\, p(\lambda)$

f) $\lambda_{MAP} = \frac{1}{n+\theta} \sum_{i=1}^{n} x_i$. θ should increase to reflect this new belief. It can also be seen from plots

Q5:

a) $P(y_i = 1 \mid x_i, \lambda) = f(x_i, \lambda)$

Parameter : $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, $y_i \in \{0, 1\}$

n is number of samples

$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, $x_i \in \{0, 1\}$

$\lambda \in \mathbb{R}$

$y_i = \begin{cases} 0 & \text{table not free} \\ 1 & \text{table free} \end{cases}$ $\quad x_i = \begin{cases} 0 & \text{not sunny} \\ 1 & \text{sunny} \end{cases}$

f is model function (or distribution) with parameters λ, X and output y, in this case may be Bernoulli Distribution with binary input X and parameter λ.

Determine the maximum likelihood estimate of λ

b) Now we have λ_{MLE}, if it is sunny today,

$P(y_i = 1 \mid 1, \lambda_{MLE}) = f(1, \lambda_{MLE})$

c) We could expand X to $\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{bmatrix}$

$x_{i1} \in \{0, 1\} = \begin{cases} 0 & \text{not sunny} \\ 1 & \text{sunny} \end{cases}$

$x_{i2} \in \{0, 1, 2\} = \begin{cases} 0 & \text{morning} \\ 1 & \text{afternoon} \\ 2 & \text{evening} \end{cases}$

and expand λ to $\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$

f is then a model function or distribution with new parameters λ and X (vector) and output y

$P(y_i = 1 \mid [x_{i1}, x_{i2}], \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}) = f(x_i, \lambda)$

Determine the maximum likelihood estimate of vector λ

Hilroy

Bonus.

d)

d	avg. distance
1	0.792
2	1.247
4	1.873
8	2.742
16	3.918
32	5.618
64	7.951
128	11.279

As d increases, average distance increases. We can infer that for k-means clustering, it won't work or it has bad performance for a high dimensional space since the volume of the space increases at a great rate relative to increase of dimensions. Average distance to means (centroid) could be too large.

e) the ratio is $(1-\varepsilon)^d$, $(0<\varepsilon<1)$

As d gets large, the ratio becomes small

And $\lim_{d \to \infty} (1-\varepsilon)^d = 0$, the volume of d-dimensional hypercube with side length $1-\varepsilon$ has nearly 0 volume.

For high d, if side length, here distance to the origin is less than 1, samples will be clustered in a very small volume. While the distance is larger than 1, samples will be distributed in a large volume. For random generated multivariate Gaussian samples in such large volume, average distance will increase as d increases since $P(-0.5<x<0.5)$ for a Gaussian Normal is only 38.3%