Q_2

a) $E[X] = \sum_{i=1}^{m} a_i X_i P(x_i)$

Since $X_i \sim N(\mu_i, \Sigma_i)$, Gaussian Distribution

$$P(X_i) = \frac{1}{\sqrt{2\pi \Sigma_i^2}} e^{-\frac{1}{2\Sigma_i^2}(X_i - \mu_i)^2}$$

$$E[X] = \sum_{i=1}^{m} a_i X_i \frac{1}{\sqrt{2\pi \Sigma_i^2}} e^{-\frac{1}{2\Sigma_i^2}(X_i - \mu_i)^2}$$

$$= \sum_{i=1}^{m} a_i \mu_i \quad \text{, since } E[X_i] = \mu_i \text{ for Gaussian Distribution}$$

Dimension of $E[X]$ is $d \times 1$.

b) ~~Cov[X]~~ $\Sigma \in \mathbb{R}^{d \times d}$ has (i,j) entry

$$\Sigma_{ij} = Cov[X_i, X_j]$$

$$Cov[X] = \sum_i \sum_j Cov[a_i X_i, a_j x_j]$$

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

$Cov[X_i, X_j] = 0$, for independent multivariate radom variables

$$Cov[X] = \sum_{i=1}^{m} V[a_i X_i] + 0$$

$$= \sum_{i=1}^{m} a_i^2 \Sigma_i$$

Dimension of $Cov[X]$ is $d \times d$.

Next Page.

If the variables X_1 and X_2 are not independent:

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

Result would be $\sum_{i=1}^{m} a_i^2 \Sigma_i \Lambda$ with dimension $d \times d$.

Q3

a) (dim = 1, σ = 1.0)

samples:	10	100	1000
means:	-0.38…	0.089…	0.002…
	-0.58…	-0.076…	-0.016…

(dim = 1, σ = 10)

Samples:	3.572…	1.009…	0.161…
means:	-1.331…	-1.549…	0.122…

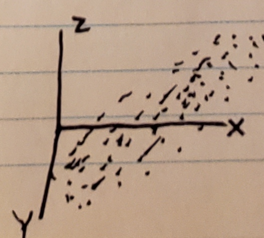
We can notice that as sample size goes up, we get closer to the actual mean. As ~~…~~ σ goes up, we moves away from the actual mean (zero).

b) when sigma = [[1,0,0], [0,1,0], [0,0,1]]

The Multivariable Gaussian distribution have ~~zero~~ no correlations between all 3 ~~dimensions~~ variables, so the plot shows a scatter view within the 3D space.

c) when Sigma = [[1,0,1], [0,1,0], [1,0,1]]

The multivariable Gaussian distribution have correlation between X and Z variables. so the plot shows scatter view of Y along the (X = Z plane).



Assignment 1

Q1

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

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

$$= 3$$

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

$$= \frac{0.1}{0.1} + \frac{0.2}{0.2} + \frac{0.7}{0.7}$$

$$= 1 + 1 + 1$$

$$= 3$$

c) For arbitary Pmf P:

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

$$= \sum_{x \in X} \frac{\cancel{P(x)}}{\cancel{P(x)}}$$

$$= 3$$

Q5

a) We can assume, is Sunny {0,1} and is free {0,1} have certain correlations. Thus build a model using maxium likelihood estimates to estimates the likelihood that when one event is (1), what is the likelihood of other event?

for D{(is sunny, is free)}:

$$P(T|S) = X^{t}(1-X)^{1-t} \cdot Y^{s}(1-Y)^{1-s}$$

b) is sunny =(1)

$$P(T|S=1) = ?$$

$$= X^{t}(1-X)^{1-t}\ Y^{s}(1-Y)^{1-s}$$

$$s=1$$

$$= X^{t}(1-X)^{1-t}\,y$$

we can use data set to ~~conve~~ predict x, y,

thus

$$P(T=1|S=1) = X \cdot Y$$

$$P(T=0|S=1) = X \cdot (1-Y)$$

c) Build distributionof Time {0, 1, 2} morning, afternoon evening, and find corelation when Time = Timen with distribution

Q4

a) $P(\lambda) = \frac{1}{2} e^{-\frac{\lambda}{2}}$

$\ln(P(\lambda)) = \cancel{\lambda} \ln(\frac{1}{2} e^{-\frac{\lambda}{2}})$

$\ln(P(\lambda))\cancel{\ } = -\frac{\lambda}{2} \cdot \ln(\frac{1}{2})$

$\lambda = -1.386$?

b) $\sum_{i=1}^{9} x_i = \cancel{\ } 79$

$$\lambda ML = \arg\max_{\lambda} \{P(D|\lambda)\}$$

$$P(D|\lambda) = \prod_{i=1}^{n} P(x_i|\lambda) = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\prod_{i=1}^{n} \cancel{\ } x_i!} \quad \cdots\cdots (a)$$

$$ll(\lambda) = \log((a)) = \log(\lambda)\sum_{i=1}^{n} x_i - n\lambda - \sum_{i=1}^{n} \log(x_i!) \quad \cdots (b)$$

$$\frac{\partial ll}{\partial \lambda} = \frac{\partial}{\partial \lambda}[(b)] = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n \quad \cdots (c)$$

when $(c) = 0$, $\lambda = \frac{\sum_{i=1}^{n} x_i}{n}$

check Global: $\frac{\partial^2 ll}{\partial \lambda^2} = \frac{\partial}{\partial \lambda}[(c)] = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2} < 0 \Rightarrow$ global maximum

$$\lambda ML = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{79}{9}$$

c) $\lambda_{MAP} = \arg\max_{\lambda} \{P(D|\lambda)P(\lambda)\}$

$$P(D|\lambda)p(\lambda) = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!} \cdot \theta e^{-\theta\lambda} = \frac{\lambda^{\sum_{i=1}^{n} x_i} \theta e^{-\lambda(n+\theta)}}{\prod_{i=1}^{n} x_i!} \quad \ldots (a)$$

$$\log(P(D|\lambda)p(\lambda)) = \log[(a)]$$

(b)

$$= \log(\lambda)\sum_{i=1}^{n} x_i + \log(\theta) - \lambda(n+\theta) - \sum_{i=1}^{n}\log(x_i!)$$

$$\frac{\partial}{\partial\lambda}[\log(p(D|\lambda)p(\lambda))] = \frac{\partial}{\partial\lambda}[(b)] = \frac{\sum_{i=1}^{n} x_i}{\lambda} - (n+\theta)$$

for $\frac{\partial}{\partial\lambda}[\log(P(D|\lambda)p(\lambda))] = 0$, $\lambda = \frac{\sum_{i=1}^{n} x_i}{n+\theta}$

check if global maximum: $\frac{\partial^2}{\partial\lambda^2}[\log(P(D|\lambda)P(\lambda))] < 0$

$$\lambda_{MAP} = \frac{\sum_{i=1}^{n} x_i}{n+\theta} = \frac{79}{9+\frac{1}{2}} = \frac{158}{19}$$

d) We could use the $\lambda_{ML} = \frac{79}{9}$ as the mean of number of accidents will happen tomorrow, also considering variance. However $\lambda_{MAP} = \frac{158}{19}$ would be a better prediction since we have previous knowledge. Both λ_{ML} and λ_{MAP} would be pretty similiar.
(prior)

e) The prior help our estimated distribution take consideration of previous knowledge. It is really helpful ~~when~~ we just started and did not view many data, it helps get the shape of distribution better.

f). To better reflect the new belief, we would like to increase the parameter θ, larger $\theta \Rightarrow$ reduced accident.