①. $\theta = p$ since N is given

$$p(x_i|\theta) = p(x_i|p) = \binom{N}{x_i} p^{x_i}(1-p)^{N-x_i}$$

$$p(D|p) = \prod_{i=1}^{n} p(x_i|p)$$

$$L(\theta) = \ln p(D|p)$$

$$= \sum_{i=1}^{n}\left(\ln\binom{N}{x_i} + \ln(p^{x_i}) + \ln[(1-p)^{N-x_i}]\right)$$

since $\ln\binom{N}{x_i}$ has no relation with p,

$$\cancel{\nabla_\theta L(\theta) = \sum_{i=1}^{n} x_i \ln}$$

$$\nabla_\theta L(\theta) = \sum_{i=1}^{n}\frac{x_i}{p} - \sum_{i=1}^{n}\frac{N-x_i}{1-p}$$

To Find MLE means $\nabla_\theta L(\theta) = 0$

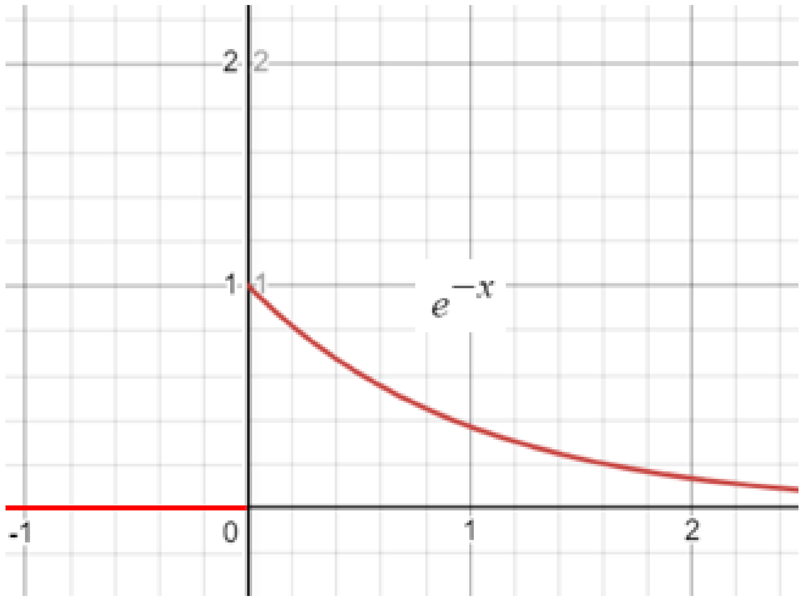
$$\sum_{i=1}^{n}\frac{x_i}{p} - \sum_{i=1}^{n}\frac{N-x_i}{1-p} = 0$$

$$\frac{\sum_{i=1}^{n} x_i}{p} = \frac{nN - \sum_{i=1}^{n} x_i}{1-p}$$

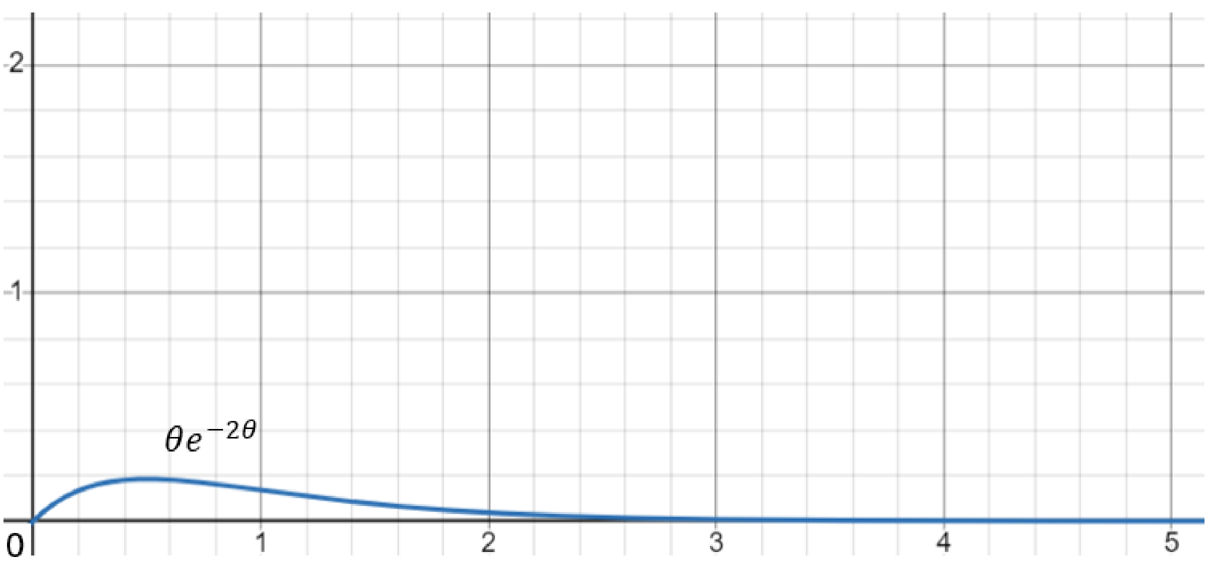
$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{nN}$$

The MLE of p is $\dfrac{\sum_{i=1}^{n} x_i}{nN}$

Q2.1



Q2.2



②. $p(x_i|\theta) = \begin{cases} \theta e^{-x\theta}, & \text{if } \cancel{x} \; x \geq 0 \\ 0, & \text{otherwise} \end{cases}$

$p(D|\theta) = \prod_{i=1}^{n} p(x_i|\theta)$

$L(\theta) = \ln p(D|\theta) = \sum_{i=1}^{n} \ln(\theta e^{-\theta x})$

$$= n \ln\theta - \theta \sum_{i=1}^{n} x_i$$

$$\nabla_\theta L(\theta) = \frac{n}{\theta} - \sum_{i=1}^{n} x_i$$

To find MLE for θ means $\nabla_\theta L(\theta) = 0$

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

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n} x_i}$$

~~The MLE~~ The MLE of θ is $\frac{n}{\sum_{i=1}^{n} x_i}$

Q3

3). To find the μ and Σ that leads to the smallest KLD means derivate of μ and Σ is zero.

$$D_{KL}(p_1(x), p_2(x)) = \int p_1(x) \ln \frac{p_1(x)}{p_2(x)} dx$$

$$= \int p_1(x) \ln p_1(x) dx - \int p_1(x) \ln p_2(x) dx$$

since $p_1(x) \ln p_1(x)$ has no relation with μ and Σ, it will be zero in derivation.

$$\cancel{\int p_1(x) \ln p_2(x)} \quad p_2(X) = \frac{1}{2\pi^{k/2} |\Sigma|^{\frac{1}{2}}} EXP[(x-\mu)^T \Sigma^{-1} (x-\mu)]$$

$$p_1(X) \ln p_2(X) = p_1(x)\left[\ln 1 - \ln\left(2\pi^{\frac{k}{2}} |\Sigma|^{\frac{1}{2}}\right) - \frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right]$$

$$= p_1(x)\left(-\frac{k}{2} \ln 2\pi - \frac{1}{2} \ln|\Sigma| - \frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

Find μ:

$$\frac{\partial}{\partial \mu} D_{KL} = -\frac{\partial}{\partial \mu}\left[\int p_1(x)\left(-\frac{k}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right) dx\right]$$

$$\cancel{= -\int \frac{\partial}{\partial \mu} p_1(x)}$$

$$= -\int p_1(x) \frac{\partial}{\partial \mu}\left(-\frac{k}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right) dx$$

$$\frac{\partial}{\partial \mu}\left(-\frac{k}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

$$= \frac{\partial}{\partial \mu}\left(-\frac{1}{2}(x^T\Sigma^{-1}x - 2x^T\Sigma^{-1}\mu + \mu^T\Sigma^{-1}\mu)\right)$$

$$= \frac{\partial}{\partial \mu}\left(x^T\Sigma^{-1}\mu - \frac{1}{2}\mu^T\Sigma\mu \right)$$

$$= x^T\Sigma^{-1} - \Sigma^{-1}\mu = \Sigma^{-1}(x-\mu)$$

Make $\frac{\partial}{\partial \mu} D_{KL} = 0$:

$$-\int p_1(x)\,\Sigma^{-1}(x-\mu)\,dx = 0$$

$$\int p_1(x)\,x\,dx - \int p(x)\,\mu\,dx = 0$$

$$E_1(X) - \mu = 0, \quad \mu = E_1(X)$$

Find Σ:

$$\frac{\partial}{\partial \Sigma}\left(-\frac{k}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

$$= -\frac{1}{2}\frac{1}{|\Sigma|} - \frac{1}{2}\left(-\Sigma^{-1}(x-\mu)(x-\mu)^T\Sigma^{-1}\right)$$

$$\frac{\partial}{\partial \Sigma} D_{KL} = -\int p_1(x)\left(-\frac{1}{2}\right)\left(\Sigma^{-1} + (-\Sigma^{-1}(x-\mu)(x-\mu)^T\Sigma^{-1})\right)dx$$

Make $\frac{\partial}{\partial \Sigma} D_{KL} = 0$.

$$-\int p_1(x)\left(-\frac{1}{2}\right)\left(\Sigma^{-1} - \Sigma^{-1}(x-\mu)(x-\mu)^T\Sigma^{-1}\right)dx = 0$$

$$\int p_1(x)\Sigma^{-1}dx - \int p_1(x)\Sigma^{-1}(x-\mu)(x-\mu)^T\Sigma^{-1}dx = 0$$

$$\Sigma^{-1} - \Sigma^{-1}E_1[(x-\mu)(x-\mu)^T\Sigma^{-1} = 0$$

$$\Sigma\Sigma^{-1} = \Sigma\Sigma^{-1}E_1[(x-\mu)(x-\mu)^T]\Sigma^{-1}$$

$$\Sigma = E_1[(x-\mu)(x-\mu)^T]$$

(4). $p(\theta | D^4) = \dfrac{p(x_4|\theta)\, p(\theta | D^3)}{\int p(x_4|\theta)\, p(\theta | D^3)\, d\theta}$, for $8 \le \theta \le 10$

otherwise $p(\theta | D^4) = 0$

$p(x_4|\theta)\, p(\theta|D^3) = \prod_{i=1}^{4} p(x_i|\theta)\, p(\theta) = \frac{1}{\theta^4} p(\theta)$

$\int_8^{10} p(x_4|\theta)\, p(\theta|D^3)\, d\theta = \int_8^{10} \frac{1}{\theta^4} p(\theta)\, d\theta$

since $p(\theta) = U(0, 10) = \frac{1}{10}$,

$p(\theta|D^4) = \dfrac{\frac{1}{\theta^4}}{\int_8^{10} \frac{1}{\theta^4} d\theta} = \dfrac{3147.54}{\theta^4}$, for $8 \le \theta \le 10$

$p(x|\theta) = \begin{cases} \frac{1}{\theta}, & \text{for } 0 \le x \le \theta \\ 0, & \text{otherwise} \end{cases}$

So $p(x|D^4) = \int_{-\infty}^{\infty} p(x|\theta)\, p(\theta|D^4)\, d\theta$

$= \begin{cases} \int_8^{10} \frac{1}{\theta} \times \frac{3147.54}{\theta^4} d\theta = 0.113, & \text{for } 0 \le x \le 8 \\ \int_x^{10} \frac{1}{\theta} \times \frac{3147.54}{\theta^4} d\theta = -\frac{3147.54}{4}\left(\frac{1}{10^4} - \frac{1}{x^4}\right) & \text{for } 8 \le x \le 10 \\ 0, & \text{otherwise} \end{cases}$

$$p(x|D^4) = \begin{cases} 0.113, & \text{for } 0 \le x \le 8 \\ \dfrac{786.89}{x^4} - \dfrac{786.89}{10^4}, & \text{for } 8 \le x \le 10 \\ 0, & \text{otherwise} \end{cases}$$

Q5.

Since all the feature vectors fall onto the same straight line in the 2-D space, the first principal component e1 will have the same direction as this line. e1 will be parallel to this straight line. And the second principal component e2 is orthogonal to e1. The second largest eigen value λ2 represent the variance under e2 direction, and since all the feature vectors fall onto a straight line that is orthogonal to e2, there is no variance in e2 direction. So λ2 is zero.

Q6.

(1)

A: There are three states. Each state has three transition probabilities, so the number of transition probabilities in total is 9.

B: There are five observation values. Each state has five emission probabilities. So, the number of emission probabilities in total is 15.

Pi: Since there have three states, there will have three initial states.

In total the parameters for this HMM is 9 + 15 + 3 = 27

(2)

First choose one from the three initial states as the starting. This choice is determined by the initial parameter Pi.

Then as the time changes from t to t+1, the state will also change based on the state at time t. And the transition probabilities distribution depends on the matrix A.

At the same time when the state changes, the model will also emit an observation based on the state. The emission probabilities distribution depends on matrix B.

Every time the model changes its state, there will be an observation. Finally, we will get an observation sequence.