1 Problem 1

a)

$$D = (x_1, x_2, x_3, x_4, x_5, x_6, x_7) = (head, head, tail, tail, head, head, head)$$

$$P(D|\theta) = \prod_{i=1}^{7} P(x_i|\theta) = \theta^5(1-\theta)^2 = \theta^5 + \theta^7 - 2\theta^6$$

Therefore, we could say that likelihood function $P(D|\theta)$depends on the parameter θ.

b)

$$l(\theta) = \ln(P(D|\theta)) = 5\ln\theta + 2\ln(1-\theta)$$

$$\frac{\partial l(\theta)}{\partial\theta} = \frac{5}{\theta} - \frac{2}{1-\theta} \stackrel{!}{=} 0$$

$$\hat{\theta} = \frac{5}{7}$$

Check the second derivative, plug $\hat{\theta}$ in the below formular,

$$\frac{\partial l(\theta)}{\partial\theta^2} = -\frac{5}{\theta^2} - \frac{2}{(1-\theta)^2}$$

$$\frac{\partial}{\partial\theta^2} l(\frac{5}{7}) < 0$$

$$P(x_8 = head, x_9 = head|\hat{\theta}) = \hat{\theta}^2 = \frac{25}{49}$$

c)

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} = \frac{P(D|\theta)P(\theta)}{\int_0^1 P(D|\theta)P(\theta)d\theta}$$

$$\int_0^1 P(D|\theta)P(\theta)d\theta = \int_0^1 \theta^5(1-\theta)^2 = \int_0^1 \theta^5 + \theta^7 - 2\theta^6 d\theta$$

$$= \frac{1}{6}\theta^6 + \frac{1}{8}\theta^8 - \frac{2}{7}\theta^7 \Big|_0^1 = \frac{1}{168}$$

$$P(\theta|D) = \begin{cases} (\theta^5 + \theta^7 - 2\theta^6) \times 168, & \text{if } 0 \le \theta \le 1 \\ 0, & \text{else} \end{cases}$$

$$\int P(x_8 = head, x_9 = head|\theta)p(\theta|D)d\theta = \int_0^1 168\theta^2(\theta^5 + \theta^7 - 2\theta^6)d\theta$$

$$= 168(\frac{1}{8}\theta^8 + \frac{1}{10}\theta^{10} - \frac{2}{9}\theta^7)\Big|_0^1 = \frac{7}{15}$$

2 Problem 2

a) Information of question on below,

$$\frac{1}{\sigma_n^2} = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}$$

$$\sigma_n^2 = \frac{\sigma^2\sigma_0^2}{n\sigma_0^2 + \sigma^2} \leq \frac{\sigma^2\sigma_0^2}{n\sigma_0^2} = \frac{\sigma^2}{n}$$

$$\sigma_n^2 = \frac{\sigma^2\sigma_0^2}{n\sigma_0^2 + \sigma^2} \leq \frac{\sigma^2\sigma_0^2}{\sigma_0^2} = \sigma_0^2$$

$$\sigma_n^2 \leq \frac{\sigma^2}{n} \qquad \text{and} \qquad \sigma_n^2 \leq \sigma_0^2$$

$$\sigma_n^2 \leq min\left(\frac{\sigma^2}{n}, \sigma_0^2\right)$$

b)

$$\frac{\mu_n}{\sigma_n^2} = \frac{n}{\sigma^2}\hat{\mu}_n + \frac{\mu_0}{\sigma_0^2}$$

$$\mu_n = \left(\frac{n}{\sigma^2}\hat{\mu}_n + \frac{\mu_0}{\sigma_0^2}\right) \times \sigma_n^2$$

$$\mu_n = \left(\frac{n}{\sigma^2}\hat{\mu}_n + \frac{\mu_0}{\sigma_0^2}\right) \times \frac{\sigma^2\sigma_0^2}{n\sigma_0^2 + \sigma^2}$$

$$\mu_n = \hat{\mu}_n \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} + \mu_0 \frac{\sigma^2}{n\sigma_0^2 + \sigma^2}$$

$$\mu_n = \hat{\mu}_n \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} + \mu_0 \frac{\sigma^2 + n\sigma_0^2 - n\sigma_0^2}{n\sigma_0^2 + \sigma^2}$$

$$\mu_n = \hat{\mu}_n \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} + \mu_0 \left(\frac{\sigma^2 + n\sigma_0^2}{n\sigma_0^2 + \sigma^2} - \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2}\right)$$

$$\mu_n = \hat{\mu}_n\lambda + \mu_0(1-\lambda) \qquad \text{where} \qquad \lambda = \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} \in (1, 0)$$

We can interpret the result as μ_n is a weighted sum of the sample mean $\hat{\mu}_n$ and the mean of the prior distribution μ_0. We can consider two cases.
In the first case, we assume $\hat{\mu}_n \geq \mu_0$, then:

$$\mu_n = \hat{\mu}_n\lambda + \mu_0(1-\lambda) \geq \mu_0\lambda + \mu_0(1-\lambda) = \mu_0 = min(\hat{\mu}_n, \mu_0)$$

$$\mu_n = \hat{\mu}_n\lambda + \mu_0(1-\lambda) \leq \hat{\mu}_n\lambda + \hat{\mu}_n(1-\lambda) = \hat{\mu}_n = max(\hat{\mu}_n, \mu_0)$$

In the second case, we assume that $\hat{\mu}_n \leq \mu_0$, then:

$$\mu_n = \hat{\mu}_n\lambda + \mu_0(1-\lambda) \leq \mu_0\lambda + \mu_0(1-\lambda) = \mu_0 = max(\hat{\mu}_n, \mu_0)$$

$$\mu_n = \hat{\mu}_n\lambda + \mu_0(1-\lambda) \geq \hat{\mu}_n\lambda + \hat{\mu}_n(1-\lambda) = \hat{\mu}_n = min(\hat{\mu}_n, \mu_0)$$

Proved!