Exercise 3

Ex. 1

a) $P(x|\theta) = \begin{cases} \theta & \text{if } x = \text{head} \\ 1-\theta & \text{if } x = \text{tail} \end{cases}$

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

For $D = \{x_1 \ldots x_7\}$:

$$= \theta^5 \cdot (1-\theta)^2 = \theta^5(1-2\theta+\theta^2)$$
$$= \theta^7 - 2\theta^6 + \theta^5$$

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

b) $= 5\ln\theta + 2\ln(1-\theta)$

$$\cancel{\nabla_l(\theta) = \frac{5\theta^4}{\theta^5} - \frac{2}{1-\theta} = \frac{5\theta^4(1-\theta) - 2\theta^5}{(1-\theta)}}$$

$$\nabla_l(\theta) = \frac{5}{\theta} - \frac{2}{(1-\theta)} = \frac{5(1-\theta) - 2\theta}{\theta(1-\theta)} = \frac{-7\theta+5}{(1-\theta)\theta}$$

- Maximum Likelihood solution: $\hat{\theta}$

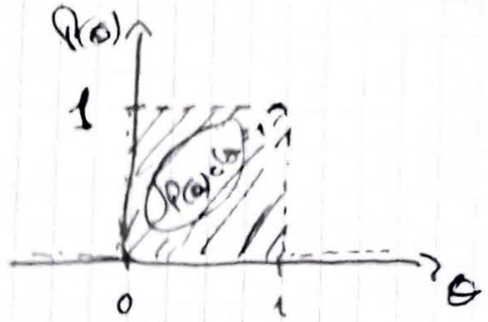
$$\nabla_l(\theta) = 0 \Rightarrow \frac{-7\theta+5}{(1-\theta)\theta} = 0 \Rightarrow -7\theta + 5 = 0$$

$$\theta = 5/7$$

- $P(x_8 = \text{head}, x_9 = \text{head} | \theta) =$

iid $= P(x_8 = \text{head}|\theta) \times P(x_9 = \text{head}|\theta) = \hat{\theta}^2 = \frac{25}{49} = 0{,}5102$

c) $P(\theta) = \begin{cases} 1 & \text{if } 0 \leq \theta \leq 1 \\ 0 & \text{else} \end{cases}$



• Posterior distribution:

$$P(\theta|D) = \frac{P(D|\theta) \cdot p(\theta)}{\int p(D|\theta) \cdot p(\theta) \, d\theta}$$

$$P(D|\theta) = \theta^5 \cdot (1-\theta)^2$$

• In range $0 \leq \theta \leq 1$

Note: in all of other $P(\theta|D) = 0$

$$P(\theta|D) = \frac{\theta^5 \cdot (1-\theta)^2}{\int_0^1 \theta^5 \cdot (1-\theta)^2 \, d\theta} = \frac{\theta^5 \cdot (1-\theta)^2}{\int_0^1 \theta^7 - 2\theta^6 + \theta^5 \, d\theta} =$$

$$= \frac{\theta^5 \cdot (1-\theta)^2}{\frac{1}{8} - \frac{2}{7} + \frac{1}{6}} = \frac{\theta^5 \cdot (1-\theta)^2}{\frac{1}{168}} = 168(\theta^5 \cdot (1-\theta)^2)$$

• Predictive distribution: $p(x|\omega_j, D_j) = \int P(x|\theta) P(\theta|D) \, d\theta$

$$\int P(x_8 = \text{head}) \times P(x_9 = \text{head}) \times 168(\theta^7 - 2\theta^6 + \theta^5) \, d\theta =$$

$$= \int_0^1 \theta^2 \cdot 168 \cdot (\theta^7 - 2\theta^6 + \theta^5) = 168 \int_0^1 \theta^9 - 2\theta^8 + \theta^7 \, d\theta =$$

$$= 168 \times \left[\frac{1}{10}\theta^{10} - \frac{2}{9}\theta^9 + \frac{1}{8}\theta^8\right]_0^1 = 168 \cdot \left(\frac{1}{10} - \frac{2}{9} + \frac{1}{8}\right) =$$

$$= 168 \times \left(\frac{1}{360}\right) = \frac{7}{15} = 0{,}466\overline{6}$$

Ex.2

a) $P(x|\mu) \sim N(\mu, \sigma^2)$

where σ^2 is known

μ is unknown, with $P(\mu) \sim N(\mu_0, \sigma_0^2)$

Posterior Prob. Distr.:

$$P(\mu|D) \sim N(\mu_n, \sigma_n^2)$$

where:

$$\frac{1}{\sigma_n^2} = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2} \quad \Big| \quad \frac{\mu_n}{\sigma_n^2} = \frac{n}{\sigma_n^2}\hat{\mu}_n + \frac{\mu_0}{\sigma_0^2} \quad \Big| \quad \hat{\mu}_n = \frac{1}{n}\sum_{k=1}^{n} x_k$$

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

Assuming $\frac{\sigma^2}{n} \leq \sigma_0^2$:

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

$$n\sigma_0^2 \cdot \cancel{\sigma^2} + \sigma^4 \geq n\cancel{\sigma^2} \cdot \sigma_0^2$$

$$\cancel{n\sigma_0^2} + \sigma^2 \geq \cancel{n\sigma_0^2}$$

$$\boxed{\sigma^2 \geq 0} \quad V$$

Assuming $\sigma_0^2 \leq \frac{\sigma^2}{n}$:

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

$$\boxed{1 \geq \frac{\sigma^2}{n\sigma_0^2 + \sigma^2}} \quad V$$

—— // ——

Both assumptions are true for all $n > 0$, $\sigma^2 > 0$ and $\sigma_0^2 > 0$, which leads to:

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

b) ① ASSUMING $\hat{\mu}_n \leq \mu_0 \Rightarrow \hat{\mu}_n \leq \mu_n \leq \mu_0$

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

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

· TO PROVE THE LEFT TERM:

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

NOTE: $\sigma_n^2 = \frac{\sigma^2 \cdot \sigma_0^2}{n\sigma_0^2 + \sigma^2}$

$$\hat{\mu}_n \leq \frac{\cancel{\sigma^2} \cdot \sigma_0^2 \cdot n}{n \cdot \sigma_0^2 \cancel{\sigma^2} + \sigma^2 \cdot \cancel{\sigma^2}} \cdot \hat{\mu}_n + \frac{\sigma^2 \cdot \cancel{\sigma_0^2}}{n \cdot \cancel{\sigma_0^2} \cdot \sigma_0^2 + \sigma^2 \cancel{\sigma_0^2}} \cdot \mu_0$$

$$\hat{\mu}_n \leq \frac{\sigma_0^2 \cdot n}{n \cdot \sigma_0^2 + \sigma^2} \hat{\mu}_n + \frac{\sigma^2}{n \cdot \sigma_0^2 + \sigma^2} \cdot \mu_0$$

$$(\cancel{n \sigma_0^2} + \sigma^2) \cdot \hat{\mu}_n \leq \cancel{\sigma_0^2 \cdot n \cdot \hat{\mu}_n} + \sigma^2 \cdot \mu_0$$

$$\cancel{\sigma^2} \cdot \hat{\mu}_n \leq \cancel{\sigma^2} \cdot \mu_0 \Rightarrow \hat{\mu}_n \leq \mu_0 \quad ✓$$

· TO PROVE THE RIGHT TERM:

$$\mu_0 \geq \frac{\sigma_0^2 \cdot n}{n \cdot \sigma_0^2 + \sigma^2} \cdot \hat{\mu}_n + \frac{\sigma^2}{n \cdot \sigma_0^2 + \sigma^2} \cdot \mu_0$$

$$(n \cdot \sigma_0^2 + \cancel{\sigma^2}) \mu_0 \geq n \cdot \sigma_0^2 \cdot \hat{\mu} + \cancel{\sigma^2 \cdot \mu_0}$$

$$\cancel{n \cdot \sigma_0^2} \cdot \mu_0 \geq \cancel{n \cdot \sigma_0^2} \cdot \hat{\mu} \Rightarrow \mu_0 \geq \hat{\mu} \quad ✓$$

② ASSUMING $\mu_0 \le \hat{\mu}_n \Rightarrow \mu_0 \le \mu_n \le \hat{\mu}_n$

· TO PROVE LEFT TERM:

$$\mu_0 \le \frac{\sigma_0^2 \cdot n}{n \cdot \sigma_0^2 + \sigma^2} \hat{\mu}_n + \frac{\sigma^2}{n \cdot \sigma_0^2 + \sigma^2} \cdot \mu_0$$

$$(n \cdot \cancel{\sigma_0^2} + \cancel{\sigma^2}) \cdot \mu_0 \le \cancel{\sigma_0^2} \cdot n \cdot \hat{\mu}_n + \cancel{\sigma^2 \cdot \mu_0}$$

$$\mu_0 \le \hat{\mu}_n \quad ✓$$

· TO PROVE RIGHT TERM:

(...) ✓