Chapter 6. Frequentist statistics

6.1.
50% is the best misclassification rate that any method can achieve.
100% is the estimated misclassification rate of the same method using LOOCV.

6.2.
a. Let $\boldsymbol{\eta} = (m_0, \tau_0^2)$.

$$
\begin{aligned}
&\hat{\boldsymbol{\eta}} = \text{argmax}_{\boldsymbol{\eta}} \int p(\mathcal{D}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta} \\
&\Rightarrow p(y_i|\theta_i) = \frac{1}{\sqrt{2\pi \cdot 500}} e^{-\frac{(y_i-\theta_i)^2}{2\cdot 500}}, \\
&p(\theta_i|\boldsymbol{\eta}) = \frac{1}{\sqrt{2\pi \cdot \tau_0^2}} e^{-\frac{(\theta_i - m_0)^2}{2\cdot\tau_0^2}}. \\
&p(y_i|\theta_i)p(\theta_i|\boldsymbol{\eta}) = \frac{1}{2\pi\sqrt{500\tau_0^2}} e^{-\frac{\tau_0^2(y_i-\theta_i)^2 + 500(\theta_i - m_0)^2}{2\cdot 500\cdot\tau_0^2}} \\
&= \frac{1}{2\pi\sqrt{500\tau_0^2}} e^{-\frac{1}{2}(\frac{\tau_0^2+500}{500\tau_0^2}(\theta_i - \frac{\tau_0^2 y_i + 500 m_0}{\tau_0^2+500})^2 + \frac{500\tau_0^2(m_0-y_i)^2}{(\tau_0^2+500)^2})} \\
&\Rightarrow p(y_i|\boldsymbol{\eta}) = \frac{1}{\sqrt{2\pi(500+\tau_0^2)}} e^{-\frac{(m_0-y_i)^2}{2(\tau_0^2+500)}} \\
&\Rightarrow p(\mathcal{D}|\boldsymbol{\eta}) = \frac{1}{(\sqrt{2\pi(500+\tau_0^2)})^n} e^{-\frac{\sum_i y_i^2 - 2m_0\sum_i y_i + nm_0^2}{2(\tau_0^2+500)}} \\
&\Rightarrow \log p(\mathcal{D}|\boldsymbol{\eta}) = -\frac{n}{2}\log(2\pi(\tau_0^2+500)) - \frac{\sum_i y_i^2 - 2m_0\sum_i y_i + nm_0^2}{2(\tau_0^2+500)}.
\end{aligned}
\tag{1}
$$

$$
\begin{aligned}
&\frac{\partial \log p(\mathcal{D}|\boldsymbol{\eta})}{\partial \tau_0^2} = -\frac{n}{2(\tau_0^2+500)} + \frac{\sum_i y_i^2 - 2m_0\sum_i y_i + nm_0^2}{2(\tau_0^2+500)^2} = 0 \\
&\Leftrightarrow n(\tau_0^2+500) = \sum_i y_i^2 - 2m_0\sum_i y_i + nm_0^2. \\
&\frac{\partial \log p(\mathcal{D}|\boldsymbol{\eta})}{\partial m_0} = -\frac{-2\sum_i y_i + 2nm_0}{2(\tau_0^2+500)} = 0 \Leftrightarrow m_0 = \frac{\sum_i y_i}{n} \\
&\Rightarrow \hat{m}_0 = \frac{\sum_i y_i}{n}, \hat{\tau}_0^2 + 500 = \frac{\sum_i y_i^2}{n} - \frac{(\sum_i y_i)^2}{n^2} \\
&\Rightarrow \hat{m}_0 \approx 1527.5, \hat{\tau}_0^2 \approx 1378.58.
\end{aligned} \tag{2}
$$

b.

$$\mathbb{E}(\theta_1|y, m_0, \tau_0^2) = \frac{500}{500+\hat{\tau}_0^2}\hat{m}_0 + \frac{\hat{\tau}_0^2}{500+\hat{\tau}_0^2}x_1 \approx 1510.99$$

$$\mathrm{Var}(\theta_1|y, m_0, \tau_0^2) = \frac{\hat{\tau}_0^2}{500+\hat{\tau}_0^2} \cdot 500 \approx 366.92$$

c. $l = F^{-1}(0.025) = 1473.45, u = F^{-1}(0.975) = 1548.43$.
Both values are reasonable.

d. Smaller σ^2 yields bigger $\hat{\tau}_0^2$.

6.3.

$$
\begin{aligned}
&\mathbb{E}[\hat{\sigma}^2_{MLE}] = \mathbb{E}[\frac{1}{N}\sum_i (x_i - \hat{\mu})^2] = \frac{1}{N}\mathbb{E}[\sum_i x_i^2 - 2N\hat{\mu}^2 + N\hat{\mu}^2] \\
&= \frac{1}{N}\mathbb{E}[\sum_i x_i^2 - N\hat{\mu}^2] = \frac{1}{N}\mathbb{E}[\sum_i x_i^2] - \mathbb{E}[\hat{\mu}^2] \\
&= \mathbb{E}[x^2] - \mathbb{E}[\hat{\mu}^2] = \sigma^2 + \mathbb{E}[x]^2 - \mathrm{Var}[\hat{\mu}] - \mathbb{E}[\hat{\mu}]^2 \\
&= \sigma^2 - \mathrm{Var}[\hat{\mu}] = \sigma^2 - \mathrm{Var}[\frac{1}{N}\sum_i x_i] \\
&= \sigma^2 - \frac{1}{N^2}\mathrm{Var}[\sum_i x_i] = \sigma^2 - \frac{1}{N^2}N\sigma^2 = \frac{N-1}{N}\sigma^2 \neq \sigma^2.
\end{aligned} \tag{3}
$$

6.4.

$$
\begin{aligned}
&\hat{\sigma}^2_{MLE} = \frac{1}{N}\sum_i (X_i - \mu)^2 \\
&\mathbb{E}[\hat{\sigma}^2_{MLE}] = \mathbb{E}[x^2] - \mathbb{E}[\mu^2] = \mathbb{E}[x^2] - \mu^2 \\
&= \sigma^2 + \mathbb{E}[x]^2 - \mu^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2.
\end{aligned}
\tag{4}
$$

This is an unbiased estimator.