1 Problem 1

a)

$$\begin{aligned} Bias(\hat{\mu}) &= E\left(\hat{\mu} - \mu\right) = E\left(\frac{1}{N}\sum_{i=1}^{N} X_i - \mu\right) \\ &= E\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) - E\left(\mu\right) = \frac{1}{N}E\left(\sum_{i=1}^{N} X_i\right) - \mu \\ &= \frac{1}{N}\sum_{i=1}^{N} E\left(X_i\right) - \mu \overset{\text{i.i.d}}{=} \frac{N}{N}\mu - \mu \\ &= 0 \end{aligned} \tag{1}$$

Since the X_is are identically distributed, they have the same, true, mean μ.

$$\begin{aligned} Var(\hat{\mu}) &= Var(\frac{1}{N}\sum_{i=1}^{N} X_i) = \frac{1}{N^2}Var\left(\sum_{i=1}^{N} X_i\right) \\ &\overset{\text{i.i.d}}{=} \frac{1}{N^2}\sum_{i=1}^{N} Var(X_i) = \frac{N}{N^2}\sigma^2 \\ &= \frac{\sigma^2}{N} \end{aligned} \tag{2}$$

Note that $Cov(X_i, X_j) = 0 \; \forall j \neq i$ because of independence.

$$\begin{aligned} MSE(\hat{\mu}) &= \left(Bias(\hat{\mu})\right)^2 + Var(\hat{\mu}) \\ &= 0^2 + \frac{\sigma^2}{N} \\ &= \frac{\sigma^2}{N} \end{aligned} \tag{3}$$

b)

$$\begin{aligned} Bias(0) &= E\left(0 - \mu\right) \\ &= -\mu \end{aligned} \tag{4}$$

$$Var(0) = 0 \tag{5}$$

Since it is a constant and therefore not random.

$$\begin{aligned} MSE(0) &= (Bias(0))^2 + Var(0) \\ &= (-\mu)^2 + 0 \\ &= \mu^2 \end{aligned} \tag{6}$$

2 Problem 2

a)

$$\begin{aligned} Error\left[\hat{f}(x)\right] &= E\left[(\hat{f}(x) - f(x))^2\right] \\ &= E\left[\left(\hat{f}(x)\right)^2 - 2\hat{f}(x)f(x) + (f(x))^2\right] \\ &= E\left[(\hat{f}(x))^2\right] - 2f(x)E\left[\hat{f}(x)\right] + (f(x))^2 \\ &= E\left[\left(\hat{f}(x)\right)^2\right] - 2\left(E\left[\left(\hat{f}(x)\right)^2\right]\right)^2 + \left(E\left[\left(\hat{f}(x)\right)^2\right]\right)^2 \\ &\quad + \left(E[\left(\hat{f}(x)\right)^2]\right)^2 - 2f(x)E\left[\hat{f}(x)\right] + (f(x))^2 \\ &= E\left[\left(\hat{f}(x)\right)^2\right] - 2E\left[\hat{f}(x)E\left[\hat{f}(x)\right]\right] + E\left[E\left[\hat{f}(x)\right]\right]^2 \\ &\quad + \left(E\left[\hat{f}(x)\right] - E\left[f(x)\right]\right)^2 \\ &= E\left[\left(\hat{f}(x)\right)^2 - 2\hat{f}(x)E\left[\left(\hat{f}(x)\right)^2\right] + E\left[\left(\hat{f}(x)\right)^2\right]\right] \\ &\quad + \left(\underbrace{E\left[\hat{f}(x)\right] - f(x)}_{Bias}\right)^2 \\ &= \underbrace{E\left[\left(\hat{f}(x) - E\left[\hat{f}(x)\right]\right)^2\right]}_{Var} + Bias^2 \\ &= Var + Bias^2 \end{aligned} \tag{7}$$

3 Problem 3

a)

$$\begin{aligned}
\underset{R}{min} A &\overset{!}{=} \underset{R}{min} E\left[D_{KL}(R||\hat{P})\right] \\
&= \underset{R}{min} E\left[\sum_{i=1}^{N} R_i \log\left(\frac{R_i}{\hat{P}_i}\right)\right] \\
&= \underset{R}{min} \sum_{i=1}^{N}\left(R_i \log R_i - R_i E\left[\log \hat{P}_i\right]\right)
\end{aligned} \tag{8}$$

$$\begin{aligned}
\frac{\partial A}{\partial R_i} &= \log R_i + R_i \cdot \frac{1}{R_i} - E\left[\log \hat{P}_i\right] \\
&= \log R_i + 1 - E\left[\log \hat{P}_i\right] \overset{!}{=} 0 \\
&\Leftrightarrow \log R_i = E\left[\log \hat{P}_i\right] - 1 \\
&\Leftrightarrow R_i = \frac{exp(E\left[\log \hat{P}_i\right])}{exp(1)} \\
&\Leftrightarrow R_i = \frac{exp(E\left[\log \hat{P}_i\right])}{exp(\sum_{j=1}^{C} \hat{P}_i)}
\end{aligned} \tag{9}$$

b)

$$
\begin{aligned}
Error\left(\hat{P}\right) &= E\left[D_{KL}(P||\hat{P})\right] \\
&= E\left[\sum_{i=1}^{N} P_i \log\left(\frac{P_i}{\hat{P}_i}\right)\right] \\
&= E\left[\sum_{i=1}^{N} P_i \log\left(\frac{P_i R_i}{\hat{P}_i R_i}\right)\right] \\
&= E\left[\sum_{i=1}^{N} P_i \left(\log P_i - \log R_i + \log \hat{P}_i - \log R_i\right)\right] \\
&= E\left[\sum_{i=1}^{N} P_i \log\left(\frac{P_i}{R_i}\right) + \sum_{i=1}^{N} P_i \log\left(\frac{R_i}{\hat{P}_i}\right)\right] \\
&= \sum_{i=1}^{N} P_i \log\left(\frac{P_i}{R_i}\right) + E\left[\sum_{i=1}^{N} P_i \log\left(\frac{R_i}{\hat{P}_i}\right)\right] \\
&= D_{KL}(P||R) + E\left[D_{KL}(R||\hat{P})\right] \\
&= Bias\left(\hat{P}\right) + Var\left(\hat{P}\right)
\end{aligned} \tag{10}
$$