Exercise 7

Ex: 1)

a) ~~P~~Given $\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} x_i$

$$\text{Bias}[\hat{\mu}] = E[\hat{\mu} - \mu]$$

$$= E[\hat{\mu}] - \mu$$

$$= E\left[\frac{1}{N} \sum_{i=1}^{N} x_i\right] - \mu$$

We know, $\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$

$$= E[\mu] - \mu$$

$$= 0$$

$$\text{var}(\hat{\mu}) = E\left[(\hat{\mu} - E[\hat{\mu}])^2\right]$$

$$= E\left[\left(\frac{1}{N} \sum_{i=1}^{N} x_i - E\left[\left(\frac{1}{N} \sum_{i=1}^{N} x_i\right]\right)\right)^2\right]$$

$$= \frac{1}{N^2} E\left[\left(\sum_{i=1}^{N} x_i - E\left[\sum_{i=1}^{N} x_i\right]\right)^2\right]$$

$$= \frac{1}{N^2} \text{var}\left[\sum_{i=1}^{N} x_i\right]$$

$$= \frac{1}{N^2} \times N\sigma^2 = \frac{\sigma^2}{N}$$

$$\text{Error}(\hat{\mu}) = \text{Bias}(\hat{\mu})^2 + \text{var}(\hat{\mu})$$

$$= 0 + \sigma^2/N = \sigma^2/N$$

b). Given $\hat{\mu} = 0$

$$\text{Bias}(\hat{\mu}) = E[\hat{\mu} - \mu] = E[0 - \mu]$$

$$= E[-\mu] = -\mu$$

$$\text{var}(\hat{\mu}) = E\left[(\hat{\mu} - E[\hat{\mu}])^2\right]$$

$$= E\left[(0 - E[0])^2\right] = 0$$

$$\text{Error}(\hat{\mu}) = \text{Bias}(\hat{\mu})^2 + \text{var}(\hat{\mu})$$

$$= (-\mu)^2 + 0$$

$$= \mu^2$$

Ex 2.

(a) $\text{Bias}(\hat{f}(x)) = E[\hat{f}(x) - f(x)] = E[\hat{f}(x)] - f(x)$

$\text{Var}(\hat{f}(x)) = E[(\hat{f}(x) - E[\hat{f}(x)]^2]$

$= E[\hat{f}^2(x)] - 2E[\hat{f}(x)] \cdot E[\hat{f}(x)] + E[\hat{f}(x)]^2$

$= E[\hat{f}^2(x)] - E[\hat{f}(x)]^2$

$\text{Error}(\hat{f}(x)) = E[(\hat{f}(x) - f(x))^2]$

$= E[\hat{f}^2(x) - 2\hat{f}(x)f(x) + f^2(x)]$

$= (E[\hat{f}^2(x)] - E[\hat{f}(x)]^2) + (E[\hat{f}(x)]^2 - 2E[\hat{f}(x)]f(x) + f^2(x))$

$= \text{Var}(\hat{f}(x)) + \text{Bias}(\hat{f}(x))^2$

Ex 3

(a) use Lagrange Multiplier:

$$L = E[\sum_{i=1}^{c} R_i \log \frac{R_i}{\hat{p}_i}] - \lambda(\sum_{i=1}^{c} R_i - 1)$$

$$\frac{\partial L}{\partial R_i} = 1 - \log R_i + E[\log \hat{p}_i] - \lambda = 0$$

$$\frac{\partial L}{\partial \lambda} = \sum_{i=1}^{c} R_i - 1 = 0$$

$$R_i = \exp(1 - \lambda + E[\log \hat{p}_i])$$

$$\sum_{i=1}^{c} \exp((1 - \lambda + E[\log \hat{p}_i])) = 1$$

$$R_i = \frac{\exp(E[\log \hat{p}_i])}{\Sigma_{j=1}^{c} \exp(E[\log \hat{p}_j])}$$

(b) $\text{Error}(\hat{p}) = E[D_{KL}(P \| \hat{P})] = E[\sum_{i=1}^{c} p_i \log p_i - E[\sum_{i=1}^{c} p_i \log \hat{p}_i] = \sum_{i=1}^{c} p_i \log p_i - \sum_{i=1}^{c} p_i E[\log \hat{p}_i]$

$\text{Bias}(\hat{p}) = D_{KL}(P \| R) = \sum_{i=1}^{c} p_i \log p_i - \sum_{i=1}^{c} p_i \log R_i$

$\text{Var}(\hat{p}) = E[D_{KL}(R \| \hat{p})] = E[\sum_{i=1} (R_i \log R_i - R_i \log \hat{p}_i)]$

$$R_i = \frac{\exp(E(\log \hat{p}_i])}{\Sigma_{j=1}^{c} \exp(E[\log \hat{p}_j])}$$

$$Bias(\hat{P}) + Var(\hat{P}) = \sum_{i=1}^{C} P_i \log P_i - \sum_{i=1}^{C} P_i \log R_i + E[\sum_{i=1}^{C} (R_i \log R_i - R_i \log \hat{P}_i)]$$

$$= \sum_{i=1}^{C} P_i \log P_i - E[\sum_{i=1}^{C} (P_i \log R_i - R_i \log R_i + R_i \log \hat{P}_i)]$$

$$= \sum_{i=1}^{C} P_i \log P_i - \sum_{i=1}^{C} E[(P_i \log R_i - R_i \log R_i + R_i \log \hat{P}_i)]$$

$$= \sum_{i=1}^{C} P_i \log P_i - \sum_{i=1}^{C} (P_i E[\log \frac{\exp(E[\log \hat{P}_i])}{\sum_{j=1}^{C} \exp(E[\log \hat{P}_i])} - E[R_i \log R_i - R_i \log \hat{P}_i])$$

$$= \sum_{i=1}^{C} P_i \log P_i - \sum_{i=1}^{C} \left(P_i E[\log \hat{P}_i] - P_i E[\log \sum_{j=1}^{C} \exp(E[\log \hat{P}_i])] - E[R_i \log R_i - R_i \log \hat{P}_i]\right)$$

$$= \sum_{i=1}^{C} P_i \log P_i - \sum_{i=1}^{C} P_i E[\log \hat{P}_i] + \sum_{i=1}^{C} \left(P_i E[\log \sum_{j=1}^{C} \exp(E[\log \hat{P}_i])] + E[R_i \log R_i - R_i \log \hat{P}_i]\right)$$

$$\sum_{i=1}^{C} \left(P_i E[\log \sum_{j=1}^{C} \exp(E[\log \hat{P}_i])] + E[R_i \log R_i - R_i \log \hat{P}_i]\right)$$

$$= \sum_{i=1}^{C} \left(P_i E[\log \sum_{j=1}^{C} \exp(E[\log \hat{P}_i])] + E[R_i E[\log \hat{P}_i]] - E[R_i \log \sum_{j=1}^{C} \exp E[\log \hat{P}_i] - R_i \log \hat{P}_i]\right)$$

$$= \sum_{i=1}^{C} \left(P_i E[\log \sum_{j=1}^{C} \exp(E[\log \hat{P}_i])] - E[R_i \log \sum_{j=1}^{C} \exp E[\log \hat{P}_i]]\right)$$

$$= \sum_{i=1}^{C} E[P_i \log \sum_{j=1}^{C} \exp(E[\log \hat{P}_i]) - R_i \log \sum_{j=1}^{C} \exp E[\log \hat{P}_i]]$$

$$= 0$$

$$Bias(\hat{P}) + Var(\hat{P}) = \sum_{i=1}^{C} P_i \log P_i - \sum_{i=1}^{C} P_i E[\log \hat{P}_i] = Error(\hat{P})$$