EXERCISE SHEET 7

Ex. 1

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

$$\text{Bias}(\hat{\mu}) = \mathbb{E}[\hat{\Theta} - \Theta] = \mathbb{E}[\hat{\mu} - \mu] = \mathbb{E}[\hat{\mu}] - \mu =$$
$$= \mathbb{E}\left[\frac{1}{N} \sum_{i=1}^{N} x_i\right] - \mu = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}[x_i] - \mu =$$
$$= \frac{1}{N} \cdot N \cdot \mu - \mu = 0$$

$$\text{VAR}(\hat{\mu}) = \mathbb{E}[(\hat{\Theta} - \mathbb{E}[\hat{\Theta}])^2] = \mathbb{E}[\hat{\Theta}^2] - \mathbb{E}[\hat{\Theta}]^2 =$$
$$= \mathbb{E}\left[\left(\frac{1}{N} \sum_{i=1}^{N} x_i\right)^2\right] - \mathbb{E}\left[\frac{1}{N} \sum_{i=1}^{N} x_i\right]^2 =$$
$$= \frac{1}{N^2} \mathbb{E}\left[\left(\sum_{i=1}^{N} x_i\right)^2\right] - \mu^2 =$$
$$= \frac{1}{N^2} \mathbb{E}\left[\sum_{i=1}^{N} x_i^2 + \sum_{i \neq j} x_i x_j\right] - \mu^2 =$$
$$= \frac{1}{N^2} \cdot \sum_{i=1}^{N} \left(\mathbb{E}[x_i^2] + \sum_{j=1}^{N-1} \mathbb{E}[x_j]^2\right) - \mu^2 =$$

NOTE: $\sigma^2 = \mathbb{E}[x^2] - \mu^2$ *

$$\overset{*}{=} \frac{1}{N^2} \cdot \sum_{i=1}^{N} \left(\sigma^2 + \mu^2 + (N-1) \cdot \mu^2\right) - \mu^2 =$$
$$= \frac{1}{N^2} \cdot (N \cdot \sigma^2 + N^2 \mu^2) - \mu^2 = \frac{1}{N} \sigma^2$$

$$\text{MSE}(\hat{\mu}) = \text{Bias}(\hat{\mu})^2 + \text{VAR}(\hat{\mu}) = \frac{\sigma^2}{N}$$

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

$$\text{Bias}(\hat{\mu}) = \mathbb{E}[0] - \mu = -\mu$$
$$\text{VAR}(\hat{\mu}) = \mathbb{E}[(0 - \mathbb{E}[0])^2] = 0$$
$$\text{MSE}(\hat{\mu}) = \text{Bias}(\hat{\mu})^2 + \text{Var}(\hat{\mu}) = \mu^2$$

Ex. 2

a)

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

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

$$= \underbrace{\mathbb{E}[\hat{f}(x)^2] \underbrace{- 2(\mathbb{E}[\hat{f}(x)])^2 + (\mathbb{E}[\hat{f}(x)])^2 + (\mathbb{E}[\hat{f}(x)])^2}_{=0}}_{A} \underbrace{- 2 f(x)\mathbb{E}[\hat{f}(x)] + f^2(x)}_{B}$$

A

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

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

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

$$= \text{Var}(\hat{f}(x))$$

B

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

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

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

Hence:

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

Ex. 3

a) $\mathbb{E}[D_{KL}(R \| \hat{P})] = \mathbb{E}[\sum_i^C R_i \log \frac{R_i}{\hat{P}_i}]$

$$\min_R \mathbb{E}[D_{KL}(R\|\hat{P})]$$

$$\text{S.T. } \sum_i^C R_i = 1$$

$$\mathcal{L}(R, l) = \mathbb{E}[\sum_i^C R_i \log \frac{R_i}{\hat{P}_i}] + l(\sum_i^C R_i - 1) =$$

$$= \mathbb{E}[\sum_i^C R_i \cdot \log R_i] - \mathbb{E}[\sum_i^C R_i \cdot \log \hat{P}_i] + l(\sum_i^C R_i - 1) =$$

$$= \sum_i^C R_i \cdot \log R_i - \sum_i^C R_i \cdot \mathbb{E}[\log \hat{P}_i] + l(\sum_i^C R_i - 1)$$

$$\frac{\partial \mathcal{L}}{\partial R_i} = \log(R_i) + 1 - \mathbb{E}[\log \hat{P}_i] + l = 0$$

$$\Rightarrow \log(R_i) = \mathbb{E}[\log \hat{P}_i] - 1 - l$$

$$R_i = \exp(\mathbb{E}[\log \hat{P}_i] - 1 - l)$$

$$\frac{\partial \mathcal{L}}{\partial l} = \sum_i^C R_i - 1 = 0$$

$$\Rightarrow \sum_i^C R_i = 1 \Leftrightarrow \sum_i^C \exp(\mathbb{E}[\log \hat{P}_i] - 1 - l) = 1$$

$$\Leftrightarrow \sum_i^C \mathbb{E}[\log \hat{P}_i] - 1 - l = 0$$

$$\Leftrightarrow l = \sum_i^C \mathbb{E}[\log \hat{P}_i] - 1$$

Hence:

$$R_i = \exp(\mathbb{E}[\log \hat{P}_i] \cancel{-1} - \sum_j^C \mathbb{E}[\log \hat{P}_j] \cancel{+1}) =$$

$$= \exp(\mathbb{E}[\log \hat{P}_i] - \sum_j^C \mathbb{E}[\log \hat{P}_j]) =$$

$$= \frac{\exp \mathbb{E}[\log \hat{P}_i]}{\sum_j^C \exp \mathbb{E}[\log \hat{P}_j]}$$

b)

$$\underbrace{D_{KL}(P\|R)}_{=\text{BIAS}(\hat{P})} + \underbrace{\mathbb{E}[D_{KL}(R\|\hat{P})]}_{=\text{VAR}(\hat{P})} \overset{?}{=} \underbrace{\mathbb{E}[D_{KL}(P\|\hat{P})]}_{=\text{ERROR}(\hat{P})}$$

$$\sum_i P_i \cdot (\log P_i - \log R_i) + \sum_i R_i (\log R_i - \mathbb{E}[\log \hat{P}_i]) =$$

* NOTE:

$$\left| \log R_i = \mathbb{E}[\log \hat{P}_i] - \log \sum_j \exp \mathbb{E}[\log \hat{P}_i] \right|$$

$$= \sum_i P_i \log P_i - \sum_i P_i (\mathbb{E}[\log \hat{P}_i] - \log \sum_j \exp \mathbb{E}[\log \hat{P}_i])$$

$$+ \sum_i R_i (\cancel{\mathbb{E}[\log \hat{P}_i]} - \log \sum_j \exp \mathbb{E}[\log \hat{P}_i] - \cancel{\mathbb{E}[\log \hat{P}_i]}) =$$

$$= \sum_i P_i \log P_i - \sum_i P_i \mathbb{E}[\log \hat{P}_i] + \cancel{\log \sum_j \exp \mathbb{E}[\log \hat{P}_i]} \cdot \underbrace{(\sum_i P_i - \sum_i R_i)}_{=0}$$

$$= \sum_i P_i \log P_i - \sum_i P_i \mathbb{E}[\log \hat{P}_i] =$$

$$= \sum_i P_i \frac{\log P_i}{\mathbb{E}[\log \hat{P}_i]} = \mathbb{E}[D_{KL}(P\|\hat{P})]$$