Solution 3 - Basics of Estimation Theory

1.

$$
\begin{aligned}
\text{NLL} &= \sum_{i=1}^{m} \left[\frac{n}{2} \ln(2\pi) + \frac{n}{2} \ln \sigma^2 + \frac{1}{2\sigma^2} (x_i - \mu)^\top (x_i - \mu) \right] \\
0 = \nabla_\mu \text{NLL} &= \sum_{i=1}^{m} -\frac{1}{\sigma^2} (x_i - \mu) \\
0 &= \sum_{i=1}^{m} x_i - \sum_{i=1}^{m} \mu = \left(\sum_{i=1}^{m} x_i \right) - m\mu \\
\implies \mu &= \frac{1}{m} \sum_{i=1}^{m} x_i \\
0 = \frac{\partial \text{NLL}}{\partial \sigma^2} &= \sum_{i=1}^{m} \frac{n}{2\sigma^2} - \frac{1}{2(\sigma^2)^2} (x_i - \mu)^\top (x_i - \mu) \\
\frac{mn\sigma^2}{2} &= \frac{1}{2} \sum_{i=1}^{m} (x_i - \mu)^\top (x_i - \mu) \\
\implies \sigma^2 &= \frac{1}{mn} \sum_{i=1}^{m} (x_i - \mu)^\top (x_i - \mu)
\end{aligned}
$$

2. In publicly available solution manual.

   Correction for typos in the solution:

$$
\begin{aligned}
E_{X_1, X_2, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)}[\hat{\sigma}^2(X_1, X_2, \ldots, X_n)] &= E[\frac{1}{n} \sum_{i=1}^{n} (X_i - \frac{\sum_{j=1}^{n} X_j}{n})^2] \\
&= \frac{1}{n} \sum_{i=1}^{n} E[(X_i - \frac{\sum_{j=1}^{n} X_j}{n})^2] \\
&= \frac{1}{n} \sum_{i=1}^{n} E[(X_i - \frac{\sum_{j=1}^{n} X_j}{n})(X_i - \frac{\sum_{j=1}^{n} X_j}{n})] \\
&= \frac{1}{n} \sum_{i=1}^{n} E[X_i^2 - \frac{2}{n} X_i \sum_{j=1}^{n} X_j + \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} X_j X_k] \\
&= \frac{1}{n} \sum_{i=1}^{n} E[X_i^2] - \frac{2}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} E[X_i X_j] + \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} E[X_j X_k] \\
&= \frac{1}{n} \sum_{i=1}^{n} E[X_i^2] - \frac{2}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} E[X_i X_j] + \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} E[X_j X_k] \\
&= \frac{1}{n} \sum_{i=1}^{n} E[X_i^2] - \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} E[X_i X_j]
\end{aligned}
$$

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n}E[X_iX_j] &= \sum_{j=1}^{n}\sum_{k=1}^{n}E[X_jX_k] \\ &= nE[X^2]+n(n-1)E[X]E[X] \\ &= n(\sigma^2+\mu^2)+(n^2-n)\mu^2 \\ &= n\sigma^2+n^2\mu^2\end{aligned}$$

$$\begin{aligned}E_{X_1,X_2,\dots,X_n\sim\mathcal{N}(\mu,\sigma^2)}[\hat{\sigma}^2(X_1,X_2,\dots,X_n)] &= \frac{1}{n}\sum_{i=1}^{n}(\sigma^2+\mu^2)-\frac{1}{n^2}(n\sigma^2+n^2\mu^2) \\ &= \sigma^2+\mu^2-(\frac{1}{n}\sigma^2+\mu^2) \\ &= \frac{n-1}{n}\sigma^2\neq\sigma^2\end{aligned}$$

3.

$$\begin{aligned}\text{NLL} &= \sum_i r-n_i\ln r+\ln n_i! \\ 0 &= \frac{\partial\text{NLL}}{\partial r}=\sum_i 1-\frac{n_i}{r} \\ m &= \frac{1}{r}\sum_i n_i \\ r &= \frac{1}{m}\sum_i n_i\end{aligned}$$

4.

$$\begin{aligned}p(y_i;\theta) &= \theta e^{-\theta y_i},\ y_i\geq 0 \\ w(\theta) &= \alpha e^{-\alpha\theta},\ \theta\geq 0 \\ p(\mathcal{D};\theta) &= \prod_{i=1}^{m}p(y_i;\theta)\end{aligned}$$

$$\ln p(\mathcal{D};\theta)=\sum_{i=1}^{m}\ln p(y_i;\theta)=\sum_{i=1}^{m}(\ln\theta-\theta y_i)=m\ln\theta-\theta\sum_{i=1}^{m}y_i$$

Posterior:

$$w(\theta\mid\mathcal{D})=\frac{p(\mathcal{D};\theta)w(\theta)}{\int p(\mathcal{D};\theta)w(\theta)\,d\theta},\ \theta\geq 0.$$

$$\begin{aligned}\hat{\theta}_{\text{MAP}} &\triangleq \arg\max_\theta w(\theta\mid\mathcal{D}) \\ &= \arg\max_\theta[\log p(\mathcal{D};\theta)+\log w(\theta)]\end{aligned}$$

$$\begin{aligned}0=\frac{d}{d\theta}[\log p(\mathcal{D}|\theta)+\log w(\theta)] &= \frac{d}{d\theta}\left(m\ln\theta-\theta\sum_{i=1}^{m}y_i+\ln\alpha-\alpha\theta\right) \\ &= \frac{m}{\theta}-\sum_{i=1}^{m}y_i-\alpha\end{aligned}$$

Now,

$$\frac{d^2}{d\theta^2}[\log p(\mathcal{D}|\theta) + \log w(\theta)] = -\frac{m}{\theta^2} < 0$$

Therefore,

$$\hat{\theta}_{\text{MAP}} = \frac{m}{\alpha + \sum_{i=1}^{m} y_i}$$

5. LSE criteria:

$$\begin{aligned}
f &= \sum_{i=1}^{n} [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 \\
0 &= \frac{\partial f}{\partial \hat{\beta}_0} = \sum_{i=1}^{n} 2(y_i - \beta_0 - \beta_1 x_i)(-1) \\
0 &= \frac{\partial f}{\partial \hat{\beta}_1} = \sum_{i=1}^{n} 2(y_i - \beta_0 - \beta_1 x_i)(-x_i) \\
\implies \hat{\beta}_0 &= \frac{\sum_{i=1}^{n} y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i}{n} \\
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}
\end{aligned}$$

Details on calculation of $\hat{\beta}_0, \hat{\beta}_1$:

$$\begin{aligned}
0 = \frac{\partial f}{\partial \hat{\beta}_0} &= \sum_{i=1}^{n} 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-1) \\
&= \sum_{i=1}^{n} y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} x_i \\
\Longrightarrow \hat{\beta}_0 &= \frac{1}{n}\left(\sum_{i=1}^{n} y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i\right) \\
&= \bar{y} - \hat{\beta}_1 \bar{x}
\end{aligned}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

$$\begin{aligned}
0 = \frac{\partial f}{\partial \hat{\beta}_1} &= \sum_{i=1}^{n} 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-x_i) \\
&= \sum_{i=1}^{n} y_i x_i - \hat{\beta}_0 \sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 \\
&= n\overline{xy} - (\bar{y} - \hat{\beta}_1 \bar{x})(n\bar{x}) - \hat{\beta}_1 n\overline{x^2} \\
&= \overline{xy} - \bar{x}\bar{y} + \hat{\beta}_1 (\bar{x})^2 - \hat{\beta}_1 \overline{x^2} \\
&= \hat{\beta}_1 \left[(\bar{x})^2 - \overline{x^2}\right] + \overline{xy} - \bar{x}\bar{y} \\
\Longrightarrow \hat{\beta}_1 &= \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2}
\end{aligned}$$

where $\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i, \overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$.
Note that $\dfrac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2}$ can be the final result of $\hat{\beta}_1$. We can also get exactly the same result as the solution by the definitions of variance and covariance.

6.

$$\begin{aligned}
P_Y(y) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{k=1}^{n}(y_k - \mu s_k)^2\right] \\
\hat{\mu} &= \arg\max_{\mu} P_Y(y) \\
\frac{\partial}{\partial \mu} &\left(\frac{1}{2}\sum_{k=1}^{n}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sum_{k=1}^{n}(y_k - \mu s_k)^2\right)\Bigg|_{\mu=\hat{\mu}}^{=0} \\
\Longrightarrow &\frac{1}{\sigma^2}\sum_{k=1}^{n} s_k(y_k - \hat{\mu} s_k) = 0 \\
\Longrightarrow &\hat{\mu} = \frac{\sum_{k=1}^{n} s_k y_k}{\sum_{k=1}^{n} s_k^2}
\end{aligned}$$

7.

$$
\begin{aligned}
-\ln(p(\mathcal{D} \mid \theta)p(\theta)) &= \frac{1}{2}\ln 2\pi + \frac{1}{2}\ln \tau^2 + \frac{1}{2\tau^2}(\mu-\eta)^2 + \sum_{i=1}^{m}\left[\frac{1}{2}\ln 2\pi + \frac{1}{2}\ln \sigma^2 + \frac{1}{2\sigma^2}(x_i-\mu)^2\right] \\
0 &= \frac{\partial \ln(p(\mathcal{D} \mid \theta)p(\theta))}{\partial \mu} \\
&= \frac{\mu-\eta}{\tau^2} + \sum_{i=1}^{m}\frac{\mu-x_i}{\sigma^2} \\
&= \left(\frac{1}{\tau^2}+\frac{m}{\sigma^2}\right)\mu - \left(\frac{\eta}{\tau^2}+\frac{1}{\sigma^2}\sum_{i=1}^{m}x_i\right) \\
\hat{\mu} &= \frac{\left(\frac{\eta}{\tau^2}+\frac{1}{\sigma^2}\sum_{i=1}^{m}x_i\right)}{\left(\frac{1}{\tau^2}+\frac{m}{\sigma^2}\right)} = \frac{\eta\sigma^2+\tau^2\sum_{i=1}^{m}x_i}{\sigma^2+m\tau^2}
\end{aligned}
$$