# MA 413 - Statistics for Data Science

# Solutions to Exercise Sheet 11

1. We note that for an individual observation $X_i$

$$\begin{aligned} P\left(X_i = GG\right) &= \frac{\theta^2}{(1+\theta)^2} \\ P\left(X_i = Gg \text{ or } X_i = gG\right) &= 2\frac{\theta}{(1+\theta)^2} \\ P\left(X_i = gg\right) &= \frac{1}{(1+\theta)^2} \end{aligned}$$

Thus the likelihood is multinomial in the three classes and we observe:

$$\begin{aligned} l(\theta) &= \frac{\theta^{2a}}{(1+\theta)^{2a}} 2^b \frac{\theta^b}{(1+\theta)^{2b}} \frac{1}{(1+\theta)^{2c}} \\ &\propto \frac{\theta^{2a+b}}{(1+\theta)^{2(a+b+c)}} \end{aligned}$$

Taking log of this quantity we observe:

$$\begin{aligned} L(\theta) &= \log(l(\theta)) \\ &= (2a+b)\log(\theta) - (2a+2b+2c)\log(1+\theta) + C \end{aligned}$$

where $C$ is some appropriate constant. We then find

$$\begin{aligned} \frac{d}{d\theta}L(\theta) &= \frac{2a+b}{\theta} - \frac{2a+2b+2c}{1+\theta} \\ \frac{d}{d\theta}L(\theta) &= 0 \\ \Leftrightarrow \frac{2a+b}{\widehat{\theta}} &= \frac{2a+2b+2c}{\widehat{\theta}+1} \\ \left.\frac{d^2}{d\theta^2}L(\theta)\right|_{\theta=\bar{\theta}} &= -\frac{2a+b}{\widehat{\theta}^2} + \frac{2a+2b+2c}{(1+\widehat{\theta})^2} \\ &= -\frac{(2a+b)(1+\widehat{\theta})^2 - 2(a+b+c)\hat{\theta}^2}{\hat{\theta}^2(1+\widehat{\theta})^2} \\ &= -\frac{(2a+b)(2a+2b+2c)^2 - 2(a+b+c)(2a+b)^2}{\hat{\theta}^2(1+\widehat{\theta})^2(2c+b)^2} \\ &< 0 \end{aligned}$$

hence this corresponds to a maximum.

2. Assume that either a person is a republican, or a democrat, and that the data set contains no apolitical elements. Denote the total number of Democrat Death penalty supporters $D$ and the total number of Republican Death penalty supporters $R$. Denote each individual sample point from the republicans as $X_{1,i}$ and each individual sample point from the democrats as $X_{2,i}$. Let $n_1$ be the sample size of the Republican sample and let $n_2$ be the sample size of the Democrat sample. Then

$$X_{k,i} \sim Ber\left(p_k\right)$$

   (a) We form the likelihood for both samples

$$\begin{aligned} l\left(p_k\right) &= \prod_{j=1}^{n_k} p_k^{x_{k,j}} \left(1-p_k\right)^{1-x_{k,j}} \\ &= p_k^{\sum_{j=1}^{n_k} x_{k,j}} \left(1-p_k\right)^{n_k - \sum_{j=1}^{n_k} x_{k,j}} \\ &= p_k^{n_k \bar{x}_{k,\cdot}} \left(1-p_k\right)^{n_k - n_k \bar{x}_{k,\cdot}}, \quad k = 1,2 \end{aligned}$$

We take $\log\left(l\left(p_k\right)\right)$ and minimise:

$$\begin{aligned}
L\left(p_k\right) &= n_k\bar{x}_{k,}\log\left(p_k\right) + \left(n_k - n_k\bar{x}_{k,}\right)\log\left(1-p_k\right) \\
\frac{\partial}{\partial p_k}L\left(p_k\right) &= \frac{n_k\bar{x}_{k,}}{p_k} - \frac{n_k - n_k\bar{x}_{k,}}{1-p_k} \\
&= 0, \quad \text{if ,} \quad p_k = \widehat{p}_k \\
&\iff \frac{n_k\bar{x}_{k,}}{\widehat{p}_k} = \frac{n_k - n_k\bar{x}_{k,}}{1-\hat{p}_k} \\
n_k\bar{x}_{k,\cdot}\left(1-\widehat{p}_k\right) &= n_k\left(1-\bar{x}_{k,}\right)\widehat{p}_k \\
\implies \hat{p}_k = \bar{x}_{k,} \\
\frac{\partial^2}{\partial p_k^2}L\left(p_k\right) &= -\frac{n_k\bar{x}_{k,}}{p_k^2} - \frac{n_k - n_k\bar{x}_{k,}}{\left(1-p_k\right)^2} \\
&< 0 \implies \max
\end{aligned}$$

(b) We know that the variance of a Bernoulli is $\sigma_k^2 = p_k\left(1-p_k\right)$ We hence form the estimate

$$\hat{\sigma}_k^2 = \widehat{p}_k\left(1-\widehat{p}_k\right) = \bar{x}_{k,-}\left(1-\bar{x}_{k,-}\right)$$

Consider a plot of the variance: see Figure 1. Clearly this is not a monotonic function, and thus we cannot use the invariance property to justify this procedure. Formally

$$\begin{aligned}
\frac{\partial}{\partial p_k}\sigma_k^2 &= 1 - 2p_k \\
\frac{\partial}{\partial p_k}\sigma_k^2 &< 0, \quad \text{if} \quad p_k > \frac{1}{2} \\
\frac{\partial}{\partial p_k}\sigma_k^2 &> 0, \quad \text{if} \quad p_k < \frac{1}{2}
\end{aligned}$$

Again, this is not a monotone function as the derivative of the function changes sign, within the permitted range of the parameter.

(c) The null hypothesis corresponds to

$$\begin{aligned}
H_0 &: \quad p_1 - p_2 = 0 \\
H_0 &: \quad p_1 - p_2 > 0
\end{aligned}$$

(d) A pivotal quantity should be based on the data, only depend on the data, and the parameter of interest and its distribution should be known. We note that $\hat{p}_k = \frac{1}{n_k}\sum X_{k,j}$, where each $X_{k,j}$ is independent. We note that

$$\mathrm{E}\left[X_{k,j}\right] = p_k, \quad \operatorname{var}\left[X_{k,j}\right] = p_k\left(1-p_k\right)$$

By using a CLT we note that

$$\begin{aligned}
Z_{k,n_k} &= \frac{\sum X_{k,j} - n_k p_k}{\sqrt{n_k p_k\left(1-p_k\right)}} \\
&\to Z_k
\end{aligned}$$

where $Z_k \sim N(0,1)$ Note that

$$\begin{aligned}
\hat{p}_k &\to \frac{1}{n_k}\left[n_k p_k + \sqrt{n_k p_k\left(1-p_k\right)}Z_k\right] \\
&= p_k + \sqrt{\frac{p_k\left(1-p_k\right)}{n_k}}Z_k
\end{aligned}$$

where the random samples are independent across $k$. Thus it transpires that

$$\begin{aligned}
\widehat{p}_1 - \widehat{p}_2 - \left(p_1 - p_2\right) &\to p_1 + \sqrt{\frac{p_1\left(1-p_1\right)}{n_1}}Z_1 - p_2 - \sqrt{\frac{p_2\left(1-p_2\right)}{n_2}}Z_2 - \left(p_1 - p_2\right) \\
&= \sqrt{\frac{p_1\left(1-p_1\right)}{n_1}}Z_1 - \sqrt{\frac{p_2\left(1-p_2\right)}{n_2}}Z_2 \\
&= T_1
\end{aligned}$$

say. We may note that from class it follows that:

$$
\begin{aligned}
T_1 &\stackrel{d}{=} N\left(0-0, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right) \\
&= N\left(0, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)
\end{aligned}
$$

So conditional on fixed values of $\sigma_1^2$ and $\sigma_2^2, T_1$ is a pivotal quantity.

(e) Clearly the critical region is going to take the form $T_1 > c_\alpha$. Thus we construct the critical value by:

$$
\begin{aligned}
1-\alpha &= P\left(T_1 \le c_\alpha \mid H_0\right) \\
&= P\left(T_1 / \sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)} \le c_\alpha / \sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)} \mid H_0\right) \\
&= \Phi\left(\frac{c_\alpha}{\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}}\right)
\end{aligned}
$$

Denoting the $100\gamma$ percentile by $z_\gamma$ we obtain that

$$
\frac{c_\alpha}{\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}} = z_{1-\alpha}
$$

Thus

$$
c_\alpha = z_{1-\alpha}\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}
$$

We obtain that

$$
\bar{x}_{1,\cdot} = \frac{46}{200} = 0.23, \quad \bar{x}_{2,\cdot} = \frac{34}{200} = 0.17
$$

Thus

$$
\widehat{\sigma}_1^2 = 0.23 \times (1-0.23) = 0.1771, \quad \widehat{\sigma}_2^2 = 0.23 \times (1-0.23) = 0.1411
$$

and we can note that as $z_{0.95} = 1.6449$,

$$
\begin{aligned}
c_\alpha &= 1.6449\sqrt{\left(\frac{0.1771}{200} + \frac{0.1411}{200}\right)} \\
&= 0.0656
\end{aligned}
$$

We observe $t_1 = 0.06$ and do not reject the null hypothesis of no difference between the two samples.

3. We make the assumption of normality in the data. To perform the test we use the two-sample test, which of course necessitates a test for equivalence of variance. We denote the means and variances of the two closing prices by $\mu_1, \mu_2, \sigma_1$ and $\sigma_2$ respectively. If the variances are equal then

$$
H_0 : \sigma_1^2/\sigma_2^2 = 1
$$

The pivotal quantity for testing equivalence of variance is:

$$
V = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1, n_2-1}
$$

We define the $100\gamma$ th percentile of the $F(d_1, d_2)$ distribution as $f_\gamma(d_1, d_2)$, and then note that:

$$
\begin{aligned}
1-\alpha &= P\left(f_{\alpha/2}(n_1-1, n_2-1) < \frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2} < f_{1-\alpha/2}(n_1-1, n_2-1) \mid H_0\right) \\
&= \left(f_{\alpha/2}(n_1-1, n_2-1) < \frac{S_1^2}{S_2^2} < f_{1-\alpha/2}(n_1-1, n_2-1) \mid H_0\right)
\end{aligned}
$$

Thus we observe a value of $V = 1.54/1.69 = 0.91$, note that $n_1 = n_2 = 15$ and thus observe at $\alpha = 0.02$ that $f_{0.01}(14, 14) = 0.27$ whilst $f_{0.99}(14, 14) = 3.6975$, and so we cannot reject the null hypothesis of equivalence of variance.

(a) We shall construct a test for equivalence of means. This will be based on a pivotal quantity constructed from an appropriately scaled difference of sample means. If $\sigma_1^2 = \sigma_2^2$ then we use a pooled estimator of the common variance $\sigma^2 = \sigma_1^2 = \sigma_2^2$ given by

$$S_p^2 = \frac{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}{n_1 + n_2 - 2}$$

$$\frac{(n_1 + n_2 - 2)\, S_p^2}{\sigma^2} = \frac{(n_1 - 1)\, S_1^2}{\sigma^2} + \frac{(n_2 - 1)\, S_2^2}{\sigma^2}$$

$$\sim \chi^2_{n_1+n_2-2}$$

Now as $\bar{X}$ and $\bar{Y}$ are independent of $S_p^2$ we find that

$$\begin{aligned} T &= \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma\sqrt{n_1^{-1} + n_2^{-1}}} \frac{1}{\sqrt{S_p^2/\sigma^2}} \\ &= \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_p\sqrt{n_1^{-1} + n_2^{-1}}} \end{aligned}$$

We have as null and alternate hypotheses:

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 \\ H_1 &: \mu_1 \neq \mu_2 \end{aligned}$$

Note that under $H_0$ we find that:

$$T \sim t\,(n_1 + n_2 - 2)$$

Hence we pick as critical region $\{(x_1, \ldots, x_n) : |t| > c_\alpha\}$

$$\begin{aligned} \alpha &= P\,(T < -c_\alpha \quad \text{or} \quad T > c_\alpha) \\ &= F_{n_1+n_2-2}\,(-c_\alpha) + 1 - F_{n_1+n_2-2}\,(c_\alpha) \\ &= 2 - 2F_{n_1+n_2-2}\,(c_\alpha) \\ F_{n_1+n_2-2}\,(c_\alpha) &= 1 - \alpha/2 \\ c_\alpha &= t_{1-\alpha/2}\,(n_1 + n_2 - 2) \end{aligned}$$

where $t_\gamma\,(n_1 + n_2 - 2)$ is the $100\gamma$ th percentile of the $t\,(n_1 + n_2 - 2)$ distribution. We note that $t_{0.99}(28) = 2.46$. We find that $s_p^2 = (14 \times 1.54 + 14 \times 1.69)/28 = 1.615$. Of course the quantity $T$ under $H_0$ takes the value $t = (40.33 - 42.54)/(\sqrt{1.615(1/15 + 1/15)}) = -4.76$, and so we find that we reject the null hypothesis.

(b) We have as null and alternate hypotheses:

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 \\ H_1 &: \mu_1 < \mu_2 \end{aligned}$$

Note that under $H_0$ we find that:

$$T \sim t\,(n_1 + n_2 - 2)$$

Hence we pick as critical region $\{(x_1, \ldots, x_n) : t < c'_\alpha\}$

$$\begin{aligned} \alpha &= P\,(T < c'_\alpha) \\ &= F_{n_1+n_2-2}\,(c'_\alpha) \\ &= F_{n_1+n_2-2}\,(c'_\alpha) \\ c'_\alpha &= t_\alpha\,(n_1 + n_2 - 2) \end{aligned}$$

where $t_\gamma\,(n_1 + n_2 - 2)$ is the $100\gamma$ th percentile of the $t\,(n_1 + n_2 - 2)$ distribution. We note that $t_{0.02}(28) = -2.1539$. Of course the quantity $T$ under $H_0$ takes the value $t = (40.33 - 42.54)/(\sqrt{1.615(1/15 + 1/15)}) =$ -4.7625 and so we find that of course we still reject the null hypothesis.

4. This is a simple test based on the mean. We would like to test

$$\begin{aligned} H_0 &: \quad \mu = 15 \\ H_1 &: \mu > 15 \end{aligned}$$

We make the assumption of normality of the data.

(a) We shall base this test on the sample mean. We note that

$$T=\frac{\bar{X}-\mu}{S\sqrt{1/n}} \sim t_{n-1}$$

where $n=36$ Obviously the critical region takes the form $\{(X_1,\ldots,X_n): T>c_\alpha\}$. We determine the value of $c_\alpha$ by

$$\alpha=P\left(T>c_\alpha\right)=1-F_{35}\left(c_\alpha\right)$$

Hence $c_\alpha$ is $t_{1-\alpha}(35)$. When $\alpha=0.05$ this corresponds to $t_{0.95}(35)=1.6896$. Of course from the data we find that

$$t=\frac{17-15}{3\sqrt{1/35}}=3.9441>1.6896$$

Hence there is evidence to reject the null corresponding to the vice-president's claim.

(b) Denote the cdf of a $t(n-1)$ by $F_{n-1}(\cdot)$. Do not confuse this with the notation of the empirical cdf

$$\alpha=P\left(T>c_\alpha\right)=1-F_{n-1}\left(c_\alpha\right)$$

and so $c_\alpha=t_{1-\alpha}(n-1)$ We now consider $\beta$. Assume that $S^2 \approx \sigma^2$.

$$\begin{aligned}
\beta &= P\left(T<c_\alpha \mid H_1\right) \\
&= P\left(\frac{\bar{X}-\mu_0}{S\sqrt{1/n}}<c_\alpha \mid H_1\right) \\
&= P\left(\frac{\bar{X}-\mu_1}{S\sqrt{1/n}}<c_\alpha+\frac{\mu_0-\mu_1}{S\sqrt{1/n}}\right) \\
&= F_{n-1}\left(c_\alpha+\frac{\mu_0-\mu_1}{S\sqrt{1/n}}\right)
\end{aligned}$$

To have

$$\beta<F_{n-1}\left(c_\alpha+\frac{\mu_0-\mu_1}{S\sqrt{1/n}}\right)$$

we need

$$t_\beta(n-1)<t_{1-\alpha}(n-1)+\frac{\mu_0-\mu_1}{\sigma\sqrt{1/n}}$$

Thus for any fixed value $\beta$ we take:

$$\begin{aligned}
\left(t_\beta(n-1)-t_{1-\alpha}(n-1)\right) &< \frac{\left(\mu_0-\mu_1\right)\sqrt{n}}{\sigma} \\
n &> \sigma^2\left(t_\beta(n-1)-t_{1-\alpha}(n-1)\right)^2/\left(\mu_0-\mu_1\right)^2
\end{aligned}$$

and for $\beta=0.01$ and $\mu_0=\mu_1-1$ we have in fact:

$$n>9\left(t_{1-\beta}(n-1)+t_{1-\alpha}(n-1)\right)^2$$

Let us calculate this for some values of $n$, and note that $n=145$ will do. For a plot of the relationship. see Figure 2 .

5. (a) The likelihood is given by (if you have done the problem for a Bernoulli set $m=1$ ):

$$l(\theta)=\prod_{i=1}^{n}\binom{m}{x_i}\theta^{x_i}(1-\theta)^{m-x_i}$$

so that

$$l(\theta) \propto \theta^{\sum x_i}(1-\theta)^{nm-\sum x_i}$$

(b) We then multiply this by the prior:

$$p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

and so get the joint distribution as

$$f_{\Theta,\underline{X}}(\theta,\underline{x}) = l(\theta)p(\theta) \propto \theta^{\alpha+\sum x_i-1}(1-\theta)^{\beta+nm-\sum x_i-1}$$

and so we can note that $\Theta \mid \underline{X}$ is Beta $(\alpha+\sum x_i, \beta+nm-\sum x_i) = \text{Beta}(\alpha^*, \beta^*)$

(c) Hence the posterior mean is just

$$\tilde{\theta}_{\text{mean}} = \frac{\alpha^*}{\alpha^*+\beta^*} = \frac{\alpha+\sum x_i}{\beta+nm+\alpha}$$

(d) To find the max we take logs:

$$\log\left(p(\theta \mid \underline{x}) = C + (\alpha^*-1)\log(\theta) + (\beta^*-1)\log(1-\theta)\right.$$

Then we find

$$\frac{\partial \log(p(\theta \mid \underline{x})}{\partial\theta} = \frac{\alpha^*-1}{\theta} - \frac{\beta^*-1}{1-\theta}$$

For this to be equal to zero we have $\tilde{\theta}_{\text{mode}} = \frac{\alpha^*-1}{\alpha^*+\beta^*-2}$ which is not the same as the mean estimate.

(e) The incomplete beta function is defined via

$$I_x(\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^x t^{\alpha-1}(1-t)^{\beta-1}dt$$

So we find that

$$I_{\tilde{\theta}_{\text{median}}}(\alpha^*,\beta^*) = \frac{1}{2}$$

numerically defines the required estimate.

(f) This question mainly means setting up the credible region, which can be done in an infinite number of ways. Decide on $\alpha_1$ and $\alpha_2$ such that $\alpha_1+\alpha_2 = 0.05$ (standard practise would be to take $\alpha_1 = \alpha_2 = 0.025$ ). Then numerically find the percentiles such that

$$F_{\Theta|X}(t_1 \mid x) = I_{t_1}(\alpha^*,\beta^*) = \alpha_1$$

and

$$F_{\Theta|X}(t_2 \mid x) = I_{t_2}(\alpha^*,\beta^*) = 1-\alpha_2$$

(see figure 5, where the shaded regions have area $\alpha_1$ and $\alpha_2$ respectibely). The 95% credible region is given by

$$(t_1, t_2)$$

6. (a) The likelihood is given by:

$$l(\theta) = \prod_{i=1}^{n} \theta e^{-x_i\theta}$$

so that

$$l(\theta) \propto \theta^n e^{-\sum x_i\theta}$$

(b) We then multiply this by the prior:

$$p(\theta) \propto \theta^{\alpha-1}e^{-\beta\theta}$$

and so get the joint distribution as

$$f_{\Theta,\underline{X}}(\theta,\underline{x}) = l(\theta)p(\theta) \propto \theta^{\alpha+n-1}e^{-(\beta+\sum x_i)\theta}$$

and so we can note that $\Theta \mid \underline{X}$ is Gamma $(\alpha+n, \beta+\sum x_i) = \text{Gamma}(\alpha^*,\beta^*)$

(c) Hence the posterior mean is just

$$\tilde{\theta}_{\text{mean}} = \frac{\alpha^*}{\beta^*} = \frac{\alpha+n}{\beta+\sum x_i}$$

(d) To find the max we take logs:

$$\log\left(p(\theta \mid \underline{x}\right) = C + (\alpha^* - 1)\log(\theta) - \beta^*\theta$$

Then we find

$$\frac{\partial \log(p(\theta \mid \underline{x})}{\partial \theta} = \frac{\alpha^* - 1}{\theta} - \beta^*$$

For this to be equal to zero we have $\tilde{\theta}_{\text{mode}} = \frac{\alpha^*-1}{\beta^*}$ which is not the same as the mean estimate.

(e) The incomplete Gamma function is defined via

$$\Gamma_x(\alpha) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t} dt$$

So we find that

$$\Gamma_{\tilde{\theta}_{median}\beta^*}(\alpha^*) = \frac{1}{2}$$

numerically defines the required estimate.

(f) This question mainly means setting up the credible region, which can be done in an infinite number of ways. Decide on $\alpha_1$ and $\alpha_2$ such that $\alpha_1 + \alpha_2 = 0.05$ (standard practise would be to take $\alpha_1 = \alpha_2 = 0.025$ ). Then numerically find the percentiles such that

$$F_{\Theta|X}(t_1 \mid x) = \Gamma_{t_1\beta^*}(\alpha^*) = \alpha_1$$

and

$$F_{\Theta|X}(t_2 \mid x) = \Gamma_{t_2\beta^*}(\alpha^*) = 1 - \alpha_2$$

(see figure 6 , where the shaded regions have area $\alpha_1$ and $\alpha_2$ respectibely). The 95% credible region is given by

$$(t_1, t_2)$$

7. The computation goes as:

$$\begin{aligned}
\operatorname{Var}(\hat{Y}(x)) &= \operatorname{Var}\left(\hat{\beta}_0 + \hat{\beta}_1 x\right) \\
&= \operatorname{Var}\left(\hat{\beta}_0\right) + x^2 \operatorname{Var}\left(\hat{\beta}_1\right) + 2x \operatorname{Cov}\left(\hat{\beta}_0, \hat{\beta}_1\right)
\end{aligned}$$

Now we know that

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdots & \cdots \\ 1 & x_n \end{pmatrix}$$

So that it follows (as shown in class)

$$\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \frac{1}{nS_{XX}} \begin{pmatrix} \sum x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{pmatrix}$$

We can then use that

$$\operatorname{Var}(\underline{\beta}) = \sigma^2 \left(\mathbf{X}^T\mathbf{X}\right)^{-1}$$

Hence

$$\begin{aligned}
\operatorname{Var}(\hat{Y}(x)) &= \sigma^2 \frac{\sum x_i^2}{nS_{XX}} + x^2 \frac{\sigma^2}{S_{XX}} - 2x\bar{x}\frac{\sigma^2}{S_{XX}} \\
&= \frac{\sigma^2}{nS_{XX}} \left(\sum x_i^2 + nx^2 - 2n\bar{x}x\right) \\
&= \frac{\sigma^2}{nS_{XX}} \left(\sum_i (x_i - x)^2\right)
\end{aligned}$$

As usual we maximise by differentiation:

$$\frac{\partial \operatorname{Var}(\hat{Y}(x))}{\partial x} = \frac{\sigma^2}{nS_{XX}} \left(2\sum_i (x - x_i)\right)$$

If we find $x^*$ such that this is zero then

$$\frac{\sigma^2}{nS_{XX}}\left(2\sum_i\left(x^*-x_i\right)\right)=0\Rightarrow nx^*-n\bar{x}=0\Rightarrow x^*=\bar{x}$$

Substitute this back into the expression for $\hat{Y}(x)$

$$\hat{Y}\left(x^*\right)=\frac{\sigma^2}{nS_{XX}}\left(\sum_i\left(x_i-\bar{x}\right)^2\right)=\frac{\sigma^2}{n}$$

8. Initially we have

$$E_{Y|\beta}(Y\mid\beta)=\mathbf{X}\beta$$

then the model is augmented to

$$E_{Y|\beta,\gamma}(Y\mid\beta,\gamma)=\mathbf{X}\beta+\gamma\underline{x}$$

We can write this in matrix form as

$$E_{Y|\beta,\gamma}(Y\mid\beta,\gamma)=\left[\begin{array}{c|c}\mathbf{X} & \underline{x}\end{array}\right]\left(\begin{array}{c}\beta\\ \gamma\end{array}\right)=\tilde{\mathbf{X}}\tilde{\beta}$$

We then find

$$\begin{aligned}\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}&=\left[\begin{array}{c}\mathbf{X}^T\\ \hline \underline{x}^T\end{array}\right]\\ &=\left[\begin{array}{c|c}\mathbf{X}^T\mathbf{X} & \mathbf{X}^T\underline{x}\\ \hline \underline{x}^T\mathbf{X} & \underline{x}^T\underline{x}\end{array}\right]\end{aligned}$$

$$\tilde{\mathbf{X}}^TY=\left[\begin{array}{c}\mathbf{X}^TY\\ \hline \underline{x}^TY\end{array}\right]$$

The least square equations then imply

$$\begin{aligned}\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}\hat{\tilde{\beta}}&=\tilde{\mathbf{X}}^TY\\ \Rightarrow\left[\begin{array}{c|c}\mathbf{X}^T\mathbf{X} & \\ \mathbf{X}^T\underline{x} & \\ \underline{x}^T\mathbf{X} & \underline{x}^T\underline{x}\end{array}\right]\left(\begin{array}{c}\dot{\beta}_N\\ \hat{\gamma}\end{array}\right)&=\left[\begin{array}{c}\mathbf{X}^TY\\ \underline{x}^TY\end{array}\right]\end{aligned}$$

This in turn gives us two matrix equations:

$$\mathbf{X}^T\mathbf{X}\hat{\beta}_N+\mathbf{X}^T\underline{x}\hat{\gamma}=\mathbf{X}^TY$$

and

$$\underline{x}^T\mathbf{X}\hat{\beta}_N+\underline{x}^T\underline{x}\hat{\gamma}=\underline{x}^TY$$

We then find from equation (8) that

$$\begin{aligned}\hat{\beta}_N&=\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^TY-\hat{\gamma}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\underline{x}\\ &=\hat{\beta}-\hat{\gamma}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\underline{x}\end{aligned}$$

Substitute this estimate $\hat{\beta}_N$ in and we in fact recover

$$\begin{gathered}\underline{x}^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^TY-\underline{x}^T\mathbf{X}\hat{\gamma}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\underline{x}+\underline{x}^T\underline{x}\hat{\gamma}=\underline{x}^TY\\ \Rightarrow\hat{\gamma}\underline{x}^T(\mathbf{I}-\mathbf{P})\underline{x}=\underline{x}^T(\mathbf{I}-\mathbf{P})Y\\ \Rightarrow\hat{\gamma}=\frac{\underline{x}^T\mathbf{A}Y}{\underline{x}^T\mathbf{A}\underline{x}}\end{gathered}$$

9. The equation goes as:

$$
\begin{aligned}
&\left(\mathbf{X}^T\mathbf{X}+\underline{x}\underline{x}^T\right)\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\mathbf{I}_p-\frac{xx^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}}{1+\underline{x}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}}\right)\\
=&\left(\mathbf{I}_p+\underline{x}\underline{x}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right)\left(\mathbf{I}_p-\frac{xx^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}}{1+\underline{x}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}}\right)\\
=&\mathbf{I}_p+\frac{\underline{x}x^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}+\underline{x}x^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}}{1+\underline{x}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}}\\
&-\frac{xx^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}+\underline{x}x^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}}{1+\underline{x}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}}\\
=&\mathbf{I}_p
\end{aligned}
$$

QED Let us define the new X matrix as

$$
\mathbf{Z}=\begin{pmatrix} \mathbf{X} & \\ x_1 & \dots x_p \end{pmatrix}
$$

This means that

$$
\mathbf{Z}^T\mathbf{Z}=\mathbf{X}^T\mathbf{X}+\underline{x}\underline{x}^T
$$

and if the new vector of observations is

$$
\mathbf{U}=\begin{pmatrix} \mathbf{Y} \\ y_{n+1} \end{pmatrix}
$$

then the new least squares equations are:

$$
\mathbf{Z}^T\mathbf{Z}\tilde{\beta}=\mathbf{Z}^T\mathbf{U}
$$

which implies that

$$
\left(\mathbf{X}^T\mathbf{X}+\underline{x}x^T\right)\tilde{\beta}=\mathbf{X}^T\mathbf{Y}+\underline{x}y_{n+1}
$$

Using the inverse given above we first write:

$$
\begin{aligned}
\tilde{\beta}&=\left(\mathbf{X}^T\mathbf{X}+\underline{x}\underline{x}^T\right)^{-1}\left(\mathbf{X}^T\mathbf{Y}+\underline{x}y_{n+1}\right)\\
&=\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\mathbf{I}_p-\frac{xx^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}}{1+\underline{x}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}}\right)\left(\mathbf{X}^T\mathbf{Y}+\underline{x}y_{n+1}\right)\\
&=\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\mathbf{X}^T\mathbf{Y}+\underline{x}y_{n+1}\right.\\
&\quad\left.-\frac{xx^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y}+\underline{x}x^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}y_{n+1}}{1+\underline{x}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}}\right)\\
&=\hat{\beta}+\frac{\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}y_{n+1}+\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}y_{n+1}\underline{x}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}}{1+\underline{x}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}}\\
&\quad-\frac{\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}\underline{x}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y}+\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}\underline{x}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}y_{n+1}}{1+\underline{x}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}}\\
&=\hat{\beta}+\frac{\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}y_{n+1}-\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}x^T\hat{\beta}}{1+\underline{x}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}}\\
&=\hat{\beta}+\frac{\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\left(y_{n+1}-\underline{x}^T\hat{\beta}\right)\underline{x}}{1+\underline{x}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}}
\end{aligned}
$$

Hence

$$
\tilde{\beta}=\hat{\beta}+\frac{\left(Y_{n+1}-\underline{x}^T\hat{\beta}\right)\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}}{1+\underline{x}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}}
$$

where $\tilde{\beta}$ and $\hat{\beta}$ are the least squares estimates $\beta$, from the augmented and first given models respectively. We would expect $\tilde{\beta}=\hat{\beta}$ when $Y_{n+1}=\underline{x}^T\hat{\beta}$, as then the model is already without the data predicting the same value for the added observation.

10. Let

$$U = \begin{pmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_n \\ Z_1 \\ \cdots \\ Z_n \end{pmatrix}$$

also let

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & x_n & 0 & 0 \\ 0 & 0 & 1 & x_1 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 1 & x_n \end{pmatrix}$$

and finally let

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \alpha_0 \\ \alpha_1 \end{pmatrix}$$

Then we have

$$E_{U|\beta}(U \mid \beta) = \mathbf{X}\beta$$

and the least squares equations give

$$\hat{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T U$$

We have that

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} n & n\bar{x} & 0 & 0 \\ n\bar{x} & \sum x_i^2 & 0 & 0 \\ 0 & 0 & n & n\bar{x} \\ 0 & 0 & n\bar{x} & \sum x_i^2 \end{pmatrix}$$

$$\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \begin{pmatrix} \frac{\sum x_i^2}{nS_{xx}} & -\frac{n\bar{x}}{nS_{xx}} & 0 & 0 \\ -\frac{n\bar{x}}{nS_{xx}} & \frac{n}{nS_{xx}} & 0 & 0 \\ 0 & 0 & \frac{\sum x_i^2}{nS_{xx}} & -\frac{n\bar{x}}{nS_{xx}} \\ 0 & 0 & -\frac{n\bar{x}}{nS_{xx}} & \frac{n}{nS_{xx}} \end{pmatrix}$$

$$\mathbf{X}^T U = \begin{pmatrix} n\bar{y} \\ \sum x_i y_i \\ n\bar{z} \\ \sum x_i z_i \end{pmatrix}$$

and so the least squares equations give us Or just note that the estimation of $\beta_0, \beta_1$ and $\alpha_0, \alpha_1$ can de-facto be carried out seperately due to the structure of X. We then use that

$$\hat{\beta} \sim N\left(\begin{pmatrix} \beta_0 \\ \beta_1 \\ \alpha_0 \\ \alpha_1 \end{pmatrix}, \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right)$$

and is independent of

$$\hat{\sigma^2} = \frac{RSS}{n-4} \sim \frac{\sigma^2}{n-4}\chi^2_{n-4}$$

where the RSS are found from the entire model, which is simply the sum of the RSS from the two seperate models. Then in fact we have

$$\hat{\beta}_1 - \hat{\alpha}_1 \sim N\left(\beta_1 - \alpha_1, \frac{2\sigma^2}{S_{xx}}\right)$$

as

$$\operatorname{Var}\left(\hat{\beta}_1-\hat{\alpha}_1\right)=\operatorname{Var}\left(\hat{\beta}_1\right)+\operatorname{Var}\left(\hat{\alpha}_1\right)-2\operatorname{Cov}\left(\hat{\alpha}_1,\hat{\beta}_1\right)$$

$$\stackrel{(*)}{\Rightarrow}\frac{\frac{\hat{\beta}_1-\hat{\alpha}_1-(\beta_1-\alpha_1)}{\sqrt{\frac{2\sigma^2}{S_{xx}}}}}{\sqrt{\frac{\frac{(n-4)\hat{\sigma}^2}{\sigma^2}}{n-4}}}=\frac{\hat{\beta}_1-\hat{\alpha}_1-(\beta_1-\alpha_1)}{\sqrt{\frac{2\sigma^2}{S_{xx}}}}\frac{\sigma}{\sqrt{\hat{\sigma}^2}}=\frac{\hat{\beta}_1-\hat{\alpha}_1-(\beta_1-\alpha_1)}{\sqrt{\frac{2}{S_{xx}}}}\frac{1}{\sqrt{\hat{\sigma}^2}}\sim t_{n-4}$$

where (*) use the fact that $\frac{X}{\sqrt{\frac{Z}{s}}}\sim t_s$ where $X\sim N(0,1)$ and $Z\sim\chi^2_s$. Indeed, here

$$\frac{\hat{\beta}_1-\hat{\alpha}_1-(\beta_1-\alpha_1)}{\sqrt{\frac{2\sigma^2}{S_{xx}}}}\sim N(0,1)$$

and

$$\frac{(n-4)\hat{\sigma}^2}{\sigma^2}\sim\chi^2_{n-4}$$

So if we let $T\sim t_{n-4}$ and

$$P\left(T<t_{n-4,\gamma/2}\right)=1-\gamma/2$$

Then

$$P\left(-t_{n-4,\gamma/2}<\frac{\hat{\beta}_1-\hat{\alpha}_1-(\beta_1-\alpha_1)}{\sqrt{\frac{2\sigma^2}{S_{xx}}}}\le t_{n-4,\gamma/2}\right)=1-\gamma$$
$$\Leftrightarrow P\left(\hat{\beta}_1-\hat{\alpha}_1-\sqrt{\frac{2\sigma^2}{S_{xx}}}t_{n-4,\gamma/2}<\beta_1-\alpha_1\le\hat{\beta}_1-\hat{\alpha}_1+\sqrt{\frac{2\sigma^2}{S_{xx}}}t_{n-4,\gamma/2}\right)=1-\gamma$$

This then provides the required confidence interval.

*Computing $Var(\beta_1-\alpha_1)$ using Multivariate operations:*

First note that $\beta_1-\alpha_1=(0,1,0,-1)\begin{pmatrix}\beta_0\\ \beta_1\\ \alpha_0\\ \alpha_1\end{pmatrix}$

Therefore, we have

$$\operatorname{Var}(\beta_1-\alpha_1)=(0,1,0,-1)\sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}(0,1,0,-1)^T$$

Now as

$$\left(\mathbf{X}^T\mathbf{X}\right)^{-1}=\begin{pmatrix}\frac{\sum x_i^2}{nS_{xx}} & -\frac{n\bar{x}}{nS_{xx}} & 0 & 0\\ -\frac{n\bar{x}}{nS_{xx}} & \frac{n}{nS_{xx}} & 0 & 0\\ 0 & 0 & \frac{\sum x_i^2}{nS_{xx}} & -\frac{n\bar{x}}{nS_{xx}}\\ 0 & 0 & -\frac{n\bar{x}}{nS_{xx}} & \frac{n}{nS_{xx}}\end{pmatrix}$$

We have that

$$\begin{aligned}\operatorname{Var}(\beta_1-\alpha_1)&=(0,1,0,-1)\sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}(0,1,0,-1)^T\\ &=\sigma^2(-\frac{n\bar{x}}{nS_{xx}},-\frac{n}{nS_{xx}},-\frac{n\bar{x}}{nS_{xx}},-\frac{n}{nS_{xx}})(0,1,0,-1)^T\\ &=\frac{2\sigma^2}{S_{xx}}\end{aligned}$$

11.

$$
\begin{aligned}
\mathbf{X} &= \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T \\
E_{Y|\mu}(Y \mid \mu) &= \mu \\
\Rightarrow \mathbf{X} &= \begin{pmatrix} 1 \\ \cdots \\ 1 \end{pmatrix} \\
\Rightarrow \mathbf{X}^T\mathbf{X} &= n \\
\Rightarrow \left(\mathbf{X}^T\mathbf{X}\right)^{-1} &= n^{-1} \\
\Rightarrow \mathbf{P} &= \begin{pmatrix} 1 \\ \cdots \\ 1 \end{pmatrix} \tfrac{1}{n}(1\ldots 1) = \tfrac{1}{n}\mathbf{J}_n
\end{aligned}
$$

where $\mathbf{J}_n$ is the $n \times n$ matrix which consists only of 1 s.

12. Let $u_i = x_i^2$ and so we have simple linear regression in $u_i$

$$E(Y) = \mathbf{U}\underline{\beta}$$

where

$$\underline{\beta} = \begin{pmatrix} \mu \\ \beta \end{pmatrix}$$

and

$$\mathbf{U} = \begin{pmatrix} 1 & u_1 \\ \cdots & \cdots \\ 1 & u_n \end{pmatrix}$$

Then we find that

$$\left(\mathbf{U}^T\mathbf{U}\right) = \begin{pmatrix} n & n\bar{u} \\ n\bar{u} & \sum u_j^2 \end{pmatrix}$$

and

$$\left(\mathbf{U}^T\mathbf{U}\right)^{-1} = \frac{1}{nS_{UU}} \begin{pmatrix} \sum u_j^2 & -n\bar{u} \\ -n\bar{u} & n \end{pmatrix}$$

$$
\begin{aligned}
\mathbf{U}\left(\mathbf{U}^T\mathbf{U}\right)^{-1} &= \begin{pmatrix} 1 & u_1 \\ \cdots & \cdots \\ 1 & u_n \end{pmatrix} \frac{1}{nS_{UU}} \begin{pmatrix} \sum u_j^2 & -n\bar{u} \\ -n\bar{u} & n \end{pmatrix} \\
&= \frac{1}{nS_{UU}} \begin{pmatrix} \sum u_i^2 - n\bar{u}u_1 & -n\bar{u} + nu_1 \\ \sum u_i^2 - n\bar{u}u_2 & -n\bar{u} + nu_2 \\ \cdots & \cdots \\ \sum u_i^2 - n\bar{u}u_n & -n\bar{u} + nu_n \end{pmatrix}
\end{aligned}
$$

Finally we find that:

$$
\begin{aligned}
\mathbf{P} &= \mathbf{U}\left(\mathbf{U}^T\mathbf{U}\right)^{-1}\mathbf{U}^T \\
&= \frac{1}{nS_{UU}} \begin{pmatrix} \sum u_i^2 - n\bar{u}u_1 & -n\bar{u} + nu_1 \\ \sum u_i^2 - n\bar{u}u_2 & -n\bar{u} + nu_2 \\ \cdots & \cdots \\ \sum u_i^2 - n\bar{u}u_n & -n\bar{u} + nu_n \end{pmatrix} \begin{pmatrix} 1 & \ldots & 1 \\ u_1 & \ldots & u_n \end{pmatrix} \\
&= \frac{1}{nS_{UU}} \begin{pmatrix} \sum u_i^2 - n\bar{u}u_1 - n\bar{u}u_1 + nu_1^2 & \sum u_i^2 - n\bar{u}u_1 - n\bar{u}u_2 + nu_1u_2 & \ldots & \sum u_i^2 - n\bar{u}u_1 - n\bar{u}u_n + nu_1v \\ \cdots & \cdots & \cdots & \cdots \\ \sum u_i^2 - n\bar{u}u_1 - n\bar{u}u_n + nu_1u_n & \sum u_i^2 - n\bar{u}u_2 - n\bar{u}u_n + nu_2u_n & \ldots & \sum u_i^2 - n\bar{u}u_n - n\bar{u}u_n + nu_n^2 \end{pmatrix}. \\
&= \frac{1}{nS_{UU}} \begin{pmatrix} \sum\left(u_i - u_1\right)^2 & \sum\left(u_i - u_1\right)\left(u_i - u_2\right) & \ldots & \sum\left(u_i - u_1\right)\left(u_i - u_n\right) \\ \cdots & \cdots & \cdots & \cdots \\ \sum\left(u_i - u_1\right)\left(u_i - u_n\right) & \sum\left(u_i - u_n\right)\left(u_i - u_2\right) & \ldots & \sum\left(u_i - u_n\right)^2 \end{pmatrix}
\end{aligned}
$$

Let $u_i = x_i^2$ to get the result.

13. Rewrite as a simple linear regression model with SOA by

$$Z_i = \begin{cases} Y_i & \text{if} \quad i \leq n/2 \\ Y_i/\sqrt{2} & \text{if} \quad i > n/2 \end{cases}$$

$$
u_i = \begin{cases} \sqrt{x_i} & \text{if} \quad i \leq n/2 \\ \sqrt{x_i/2} & \text{if} \quad i > n/2 \end{cases}
$$

Then as usual

$$
\widehat{\beta_0} = \bar{z} - \bar{u}\widehat{\beta_1}
$$

and

$$
\widehat{\beta_1} = \frac{S_{ZU}}{S_{UU}}
$$

14. When writing out the solutions for this, I realised it is a tiny bit involved. I wouldn't worry too much if you can't do it.

 (a)

$$
\begin{gathered}
\mathbf{W} = \mathbf{U} + \epsilon \\
\epsilon_j \sim N\left(0, \sigma^2\right) \Rightarrow W_j \mid U_j = u_j \sim N\left(u_j, \sigma^2\right)
\end{gathered}
$$

 Hence

$$
f_{W_j \mid U_j}(w \mid u) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(w-u)^2\right) w \in \mathbb{R}
$$

 (b) As noted in class (on multiple occasions)

$$
f_{U_j \mid W_j}(u \mid w) = C f_{U_j, W_j}(u, w)
$$

 for some constant (wrt $u$ ) $C$. We then have

$$
\begin{aligned}
f_{U_j, W_j}(u, w) = & \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(w-u)^2\right)\left(\frac{\pi_j}{\sqrt{2\pi\tau_j^2}} \exp\left(-\frac{1}{2\tau_j^2}u^2\right) + \left(1-\pi_j\right)\delta(u)\right) \\
= & \frac{1}{\sqrt{2\pi\sigma^2}}\left(\frac{\pi_j}{\sqrt{2\pi\tau_j^2}} \exp\left(-\frac{\tau_j^2+\sigma^2}{\tau_j^2\sigma^2}\left(u^2 - 2\frac{2\tau_j^2}{\Omega_j^2}uw\right) - \frac{1}{2\sigma^2}w^2\right)\right. \\
& \left. + \exp\left(-\frac{1}{2\sigma^2}(u-w)^2\right)\left(1-\pi_j\right)\delta(u)\right)
\end{aligned}
$$

 Using the equation above and the proportionality relation we then find

$$
P\left(U_j = 0 \mid W_j = w\right) = \frac{C}{\sqrt{2\pi\sigma^2}}\left(1-\pi_j\right)\exp\left(-\frac{1}{2\sigma^2}w^2\right)
$$

 Let $\tilde{C} = \frac{C}{\sqrt{2\pi\sigma^2}}$ and $C^{\dagger} = \tilde{C}e^{-\frac{1}{2\sigma^2}w^2}$. Then from first principles we know that

$$
\begin{aligned}
F_{U_j \mid W_j}(u \mid w) = & \int_{-\infty}^{u} f_{U_j \mid W_j}(u \mid w) du \\
= & \tilde{C}\left(\pi_j \int_{-\infty}^{u} \frac{e^{-\frac{1}{2\sigma^2}w^2}}{\sqrt{2\pi\tau_j^2}} \exp\left(-\frac{\tau_j^2+\sigma^2}{\tau_j^2\sigma^2}\left(u^2 - 2\frac{2\tau_j^2}{\Omega_j^2}uw\right)\right) du + \left(1-\pi_j\right)\exp\left(-\frac{1}{2\sigma^2}w^2\right) I(w \geq 0)\right) \\
= & C^{\dagger}\left(\pi_j e^{\frac{\Omega_j^2}{\tau_j^2\sigma^2}\frac{\tau_j^2}{\Omega_j^4}w^2} \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi\tau_j^2}} \exp\left(-\frac{\Omega_j^2}{\tau_j^2\sigma^2}\left(u - \frac{\tau_j^2}{\Omega_j^2}w\right)^2\right) du + \left(1-\pi_j\right) I(w \geq 0)\right) \\
= & C^{\dagger}\left(\pi_j e^{\frac{\tau_j^2}{\Omega_j^2\sigma^2}w^2} \frac{\sigma}{\Omega_j} \int_{-\infty}^{\frac{\Omega_j}{\tau_j\sigma}\left(u - \frac{\tau_j^2 w}{\Omega_j^2}\right)} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\xi^2\right) d\xi + \left(1-\pi_j\right) I(w \geq 0)\right) \\
= & C^{\dagger}\left(\pi_j e^{\frac{\tau_j^2}{\Omega_j^2\sigma^2}w^2} \frac{\sigma}{\Omega_j} \Phi\left(\frac{\Omega_j}{\tau_j\sigma}\left(u - \frac{\tau_j^2 w}{\Omega_j^2}\right)\right) + \left(1-\pi_j\right) I(w \geq 0)\right)
\end{aligned}
$$

 where we define

$$
I(w \geq 0) = \begin{cases} 1 & \text{if} \quad w \geq 0 \\ 0 & \text{if} \quad w < 0 \end{cases}
$$

We now have to determine the constant of proportionality. This is done by

$$\lim_{u \to \infty} P\left(U_j \leq u \mid W_j = w\right) = 1$$

Or

$$C^{\dagger}\left(\pi_j e^{\frac{\tau_j^2}{\Omega_j^2 \sigma^2} w^2} \frac{\sigma}{\Omega_j} + (1-\pi_j)\right) = 1$$

This of course means

$$\begin{aligned} C^{\dagger} &= \left(\pi_j e^{\frac{\tau^2}{\Omega_j^2 \sigma^2} w^2} \frac{\sigma}{\Omega_j} + (1-\pi_j)\right)^{-1} \\ &= \left(\xi_j^{-1}(1-\pi_j) + (1-\pi_j)\right)^{-1} \\ &= \frac{\xi_j}{1+\xi_j} \frac{1}{1-\pi_j} \end{aligned}$$

Hence we have

$$\begin{aligned} F_{U_j|W_j}(u \mid w) &= \frac{\xi_j}{1+\xi_j} \frac{1}{1-\pi_j}\left(\frac{1-\pi_j}{\xi_j} \Phi\left(\frac{\Omega_j}{\tau_j \sigma}\left(u - \frac{\tau_j^2 w}{\Omega_j^2}\right)\right) + (1-\pi_j)\right) \\ &= \frac{1}{1+\xi_j} \Phi\left(\frac{\Omega_j}{\tau_j \sigma}\left(u - \frac{\tau_j^2 w}{\Omega_j^2}\right)\right) + \frac{\xi_j}{1+\xi_j} I(w \geq 0) \end{aligned}$$

as we were supposed to.

(c) From the previous question we clearly have

$$\begin{aligned} E_{U_j|W_j}\left(U_j \mid W_j\right) &= \frac{1}{1+\xi_j} E_{N\left(\frac{\tau_j^2 w}{\Omega_j^2}, \frac{\tau_j^2 \sigma^2}{\Omega_j^2}\right)}\{U\} + 0 \times \frac{\xi_j}{1+\xi_j} \\ &= \frac{1}{1+\xi_j} \frac{\tau_j^2 w}{\Omega_j^2} \end{aligned}$$

So we are effectively shrinking the estimates $w$ towards zero as $\xi_j > 0$, and $\tau_j^2 < \Omega_j^2$, unlike the thresholding rule given in class, where we either kept or killed (setting $\hat{u}_j = 0$ ) our parameters.

(d) Well from above it is clear that (sorry that should have been $U_j$ )

$$P\left(U_j = 0 \mid W_j = w_j\right) = \frac{\xi_j}{1+\xi_j}$$

and

$$P\left(U_j \neq 0 \mid W_j = w_j\right) = 1 - P\left(U_j = 0 \mid W_j = w_j\right) = \frac{1}{1+\xi_j}$$

and so

$$\chi_j = \xi = \frac{1-\pi_j}{\pi_j} \frac{\Omega_j}{\sigma} \exp\left(-\frac{\tau_j^2}{2\sigma^2 \Omega_j^2} w_j^2\right)$$

We see that this expression depends on the hyperparameters of the problem (our initial belief that the wavelet coefficient is zero $(1-\pi_j)$, and our prior belief of the spread of the non-zero coefficients $\tau$ ). Strictly speaking we should also put a prior on $\sigma$, the variance of $\epsilon_j$. Decreasing $\pi_j$ and $\tau_j$ will mean setting more coefficients to zero. Compare this very brief discussion with the final chapter of your notes.

15. This question is extremely straightforward, as in the previous question

$$\begin{gathered} \mathbf{W} = \mathbf{U} + \epsilon \\ \epsilon_j \sim N\left(0, \sigma^2\right) \Rightarrow W_j \mid U_j = u_j \sim N\left(u_j, \sigma^2\right) \end{gathered}$$

Hence

$$f_{W_j|U_j}(w \mid u) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(w-u)^2\right) w \in \mathbb{R}$$

As noted in class (on multiple occasions)

$$f_{U_j|W_j}(u \mid w) = C f_{U_j,W_j}(u,w)$$

for some constant (wrt $u$ ) $C$. We then have

$$\begin{aligned} f_{U_j,W_j}(u,w) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(w-u)^2\right) \frac{\lambda}{2\sigma^2} \exp\left(-\frac{\lambda}{\sigma^2}|u|\right) \\ &= \frac{\lambda}{2\sigma^3\sqrt{2\pi}} \exp\left(-\frac{1}{\sigma^2}\left((w-u)^2/2 + \lambda|u|\right)\right) \end{aligned}$$

Then

$$f_{U_j|W_j}(u \mid w) = C \frac{\lambda}{2\sigma^3\sqrt{2\pi}} \exp\left(-\frac{1}{\sigma^2}\left((w-u)^2/2 + \lambda|u|\right)\right)$$

Clearly this function is at a maximum, when the argument of the exponential is at a maximum as exp is an increasing function, which is when the negative of the argument of the exponential is at a minimum or

$$(w-u)^2/2 + \lambda|u|$$

is minimised as we may ignore positive multiplicative constants. Let $s = |u|$ and use the hint given on the sheet to mimise (over $s$ )

$$k(s) = (|w| - s)^2/2 + \lambda s$$

we do this using calculus:

$$\begin{aligned} \tfrac{dk}{ds} &= s - |w| + \lambda \\ \tfrac{d^2k}{ds^2} &= 1 + \lambda > 0 \Rightarrow \min \end{aligned}$$

This is zero when

$$s = |w| - \lambda$$

if the latter quantity is positive, (as $s$ is constrained to be positive) so for $|w| > \lambda$, the function is minimised by

$$|u| = |w| - \lambda$$

If $|w| < \lambda$ then

$$\frac{dk}{ds} = s - |w| + \lambda$$

is positive, and so we should choose $s$ as small as it can be, namely $s = 0$. This then as $u$ has the same sign as $w$ means we maximise the posterior via

$$u^{(st)} = \begin{cases} w - \lambda & \text{for} \quad w > \lambda \\ 0 & \text{for} \quad |w| \leq \lambda \\ -(|w| - \lambda) & \text{for} \quad w < -\lambda \end{cases}$$

as we were asked to demonstrate.