1 Question 2

Show that $o_P(1) + \mathcal{O}_P(1) = \mathcal{O}(1)$.

Note that $X_n = o_P(1)$ if $X_n \xrightarrow{p} 0$ and $X_n = \mathcal{O}_P(1)$ if X_n is tight. Note that $X_n \xrightarrow{p} 0$ implies that $X_n \xrightarrow{d} 0$ and therefore X_n is tight.

Let $X_n = o_P(1)$ and $Y_N = \mathcal{O}(1)$. By the above logic, X_n is tight. So $\forall \epsilon > 0, \exists B_x, B_y$ such that:

$$\inf_n \Pr(|X_n| \leq B_x) \geq 1 - \frac{\epsilon}{2}$$
$$\inf_n \Pr(|Y_n| \leq B_y) \geq 1 - \frac{\epsilon}{2}$$

For any such $\epsilon > 0$, choose M such that

$$\frac{M}{2} > B_x \qquad \frac{M}{2} > B_y$$

Define A and B such that:

$$A := \{|X_n| + |Y_n| > M\} \Rightarrow \left\{|X_n| > \frac{M}{2}\right\} \cup \left\{|Y_n| > \frac{M}{2}\right\} =: B$$

Note that:

$$\Pr(A) \leq \Pr(B) \leq \Pr\left(|X_n| > \frac{M}{2}\right) + \Pr\left(|Y_n| > \frac{M}{2}\right) < \epsilon$$

From the definition of tightness, and our choice of M.

Define $C := \{|X_n + Y_n| > M\}$. From the triangle inequality we know that $|X_n + Y_n| \leq |X_n| + |Y_n|$

Thus:

$$|X_n + Y_n| > M \Rightarrow |X_n| + |Y_n| > M$$
$$\Pr(|X_n + Y_n| > M) \leq \Pr(|X_n| + |Y_n| > M) < \epsilon$$
$$\Pr(|X_n + Y_n| \leq M) \geq 1 - \epsilon$$

This tells us that $X_n + Y_n$ is tight, and therefore $\mathcal{O}_P(1) + \mathcal{O}_P(1) = \mathcal{O}_P(1)$. So:

$$o_P(1) + \mathcal{O}_P(1) = \mathcal{O}_P(1)$$

2 Question 6

$$\mathbb{V}(Y|X) = \mathbb{E}\left[(Y - \mathbb{E}[Y|X])^2 |X\right]$$

Let $Z = Y^2$. Then

$$\mathbb{E}\left[\mathbb{E}\left[Y^2|X\right]\right] = \mathbb{E}[\mathbb{E}[Z|X]] = \mathbb{E}[Z] = \mathbb{E}\left[Y^2\right]$$

2.1 A

$$\begin{aligned}\mathbb{V}(Y|X) &= \mathbb{E}\left[Y^2|X\right] - 2\mathbb{E}\left[Y\mathbb{E}\left[Y|X\right]|X\right] + \mathbb{E}\left[\mathbb{E}\left[Y|X\right]^2|X\right] \\ &= \mathbb{E}\left[Y^2|X\right] - 2\mathbb{E}\left[Y|X\right]\mathbb{E}\left[Y|X\right] + \mathbb{E}\left[Y|X\right]^2 \\ &= \mathbb{E}\left[Y^2|X\right] - \mathbb{E}\left[Y|X\right]^2\end{aligned}$$

2.2 B

$$\begin{aligned}\mathbb{E}\left[\mathbb{V}(Y|X)\right] &= \mathbb{E}\left[\mathbb{E}\left[Y^2|X\right]\right] - \mathbb{E}\left[\mathbb{E}\left[Y|X\right]^2\right] \\ \mathbb{V}\left(\mathbb{E}\left[Y|X\right]\right) &= \mathbb{E}\left[\mathbb{E}\left[Y|X\right]^2\right] - \mathbb{E}\left[\mathbb{E}\left[Y|X\right]\right]^2\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[\mathbb{V}(Y|X)\right] + \mathbb{V}\left(\mathbb{E}\left[Y|X\right]\right) &= \mathbb{E}\left[\mathbb{E}\left[Y^2|X\right]\right] - \mathbb{E}\left[\mathbb{E}\left[Y|X\right]\right]^2 \\ &= \mathbb{E}\left[\mathbb{E}\left[Y^2|X\right]\right] - \mathbb{E}\left[Y\right]^2 \\ &= \mathbb{E}\left[Y^2\right] - \mathbb{E}\left[Y\right]^2 \\ &= \mathbb{V}(Y)\end{aligned}$$

3 Question 10

Let (Y, X) be a bivariate normal random variable. Find $\mathbb{E}\left[Y|X\right]$.

$$(Y, X) \sim \mathcal{N}\left(\begin{pmatrix}\mu_Y \\ \mu_X\end{pmatrix}, \begin{pmatrix}\sigma_Y^2 & \rho\sigma_Y\sigma_X \\ \rho\sigma_Y\sigma_X & \sigma_X^2\end{pmatrix}\right)$$

Any bivariate normal random variable can be rewritten as:

$$\begin{aligned}X &= \sigma_X Z_1 + \mu_X \\ Y &= \sigma_Y \rho Z_1 + Z_2\sqrt{1-\rho^2} + \mu_Y\end{aligned}$$

This allows us to rewrite Z_1 and then Y.

$$\begin{aligned}Z_1 &= \frac{X - \mu_X}{\sigma_X} \\ Y &= \sigma_Y \rho\left(\frac{X-\mu_X}{\sigma_X}\right) + Z_2\sqrt{1-\rho^2} + \mu_Y\end{aligned}$$

Taking the expectation conditioned on X.

$$
\begin{aligned}
\mathbb{E}\left[Y|X\right] &= \mathbb{E}\left[\sigma_Y \rho \frac{X-\mu_X}{\sigma_X}|X\right] + \mathbb{E}\left[\sqrt{1-\rho^2} Z_2|X\right] + \mathbb{E}\left[\mu_Y|X\right] \\
&= \frac{\sigma_Y \rho}{\sigma_X}\mathbb{E}\left[X|X\right] - \frac{\sigma_Y \rho \mu_X}{\sigma_X} + \sqrt{1-\rho^2}\mathbb{E}\left[Z_2|X\right] + \mu_Y \\
&= \frac{\sigma_Y \rho}{\sigma_X} X - \frac{\sigma_Y \rho \mu_X}{\sigma_X} + \mu_Y
\end{aligned}
$$

where $\mathbb{E}\left[Z_2|X\right] = \mathbb{E}\left[Z_2\right] = 0$ by the fact that Z_1, Z_2 are independent, and X is a function of Z_1 only.

4 Question 14

The best linear predictor of Y conditioned on $\boldsymbol{X}$ is given by:

$$\min_{\boldsymbol{b} \in \mathbb{R}^3} \mathbb{E}\left[\left[Y - X'\boldsymbol{b}\right]^2\right]$$

Note that $X_1 b_1 + X_2 b_2 + X_3 b_3 = X_1\left(b_1 + \alpha_1 b_3\right) + X_2\left(b_2 + \alpha_2 b_3\right) := (\gamma_1, \gamma_2)$.

The best linear predictor of Y given (X_1, X_2) is given by:

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^2} \mathbb{E}\left[\left[Y - X_1\beta_1 - X_2\beta_2\right]^2\right]$$

It would not be possible to minimize over two dimensions and do better than minimizing over three. One could fix $b_3 = 0$ and then reach the same problem as minimizing over two dimensions.

This tells us that

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^2} \mathbb{E}\left[\left[Y - X_1\beta_1 - X_2\beta_2\right]^2\right] \geq \min_{\boldsymbol{b} \in \mathbb{R}^3} \mathbb{E}\left[\left[Y - X'\boldsymbol{b}\right]^2\right]$$

One cannot do any worse minimizing over the two dimensions either. For any value of $\boldsymbol{b}$, choose $\boldsymbol{\gamma}$ as above, and $\mathbb{E}\left[\left[Y - X'\boldsymbol{b}\right]^2\right] = \mathbb{E}\left[\left[Y - (X_1, X_2)'\boldsymbol{\gamma}\right]^2\right]$. Thus the two dimensional case can always do as well as the three dimensional case and:

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^2} \mathbb{E}\left[\left[Y - X_1\beta_1 - X_2\beta_2\right]^2\right] \leq \min_{\boldsymbol{b} \in \mathbb{R}^3} \mathbb{E}\left[\left[Y - X'\boldsymbol{b}\right]^2\right]$$

This leads us to conclude that:

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^2} \mathbb{E}\left[\left[Y - X_1\beta_1 - X_2\beta_2\right]^2\right] = \min_{\boldsymbol{b} \in \mathbb{R}^3} \mathbb{E}\left[\left[Y - X'\boldsymbol{b}\right]^2\right]$$

This means the best linear predictor of Y given $\boldsymbol{X}$ is equivalent to the best linear predictor of Y given (X_1, X_2). Since we know that there is no perfect colinearity between (X_1, X_2) we may apply the standard Linear Regression approach.

$$\boldsymbol{\beta} = \mathbb{E}\left[(X_1, X_2)(X_1, X_2)'\right]\mathbb{E}\left[(X_1, X_2)Y\right]$$

The solution to the minimization problem over all X is any combination of b_1, b_2, b_3 such that $\boldsymbol{\beta} = (b_1 + \alpha_1 b_3, b_2 + \alpha_2 b_3)'$.