ISLA : TD1 (2024/2025)

-- Exercise 1

U and V are independent RVs and uniform in $[0,1]$

$X = U + V$ and $Y = U - V$

(a) $Z = \begin{bmatrix} X \\ Y \end{bmatrix} \Rightarrow \mathbb{E}[Z] = \begin{bmatrix} \mathbb{E}[X] \\ \mathbb{E}[Y] \end{bmatrix} = \begin{bmatrix} \mathbb{E}[U+V] \\ \mathbb{E}[U-V] \end{bmatrix} = \begin{bmatrix} \mathbb{E}[U] + \mathbb{E}[V] \\ \mathbb{E}[U] - \mathbb{E}[V] \end{bmatrix}$

Remember that

$$\mathbb{E}[U] = \int_{\mathbb{R}} u \, p_U(u) \mathrm{d}u = \int_0^1 u \, \mathrm{d}u = \frac{1}{2} = \mathbb{E}[V] \quad \text{therefore} \quad \mathbb{E}[Z] = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

We also have that

$$\mathrm{cov}(Z) = \mathbf{\Sigma}_Z = \mathbb{E}\Big[\big(Z - \mathbb{E}[Z]\big)\big(Z - \mathbb{E}[Z]\big)^\top\Big] = \begin{bmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X,Y) & \mathrm{Var}(Y) \end{bmatrix}$$

Since U and V are independent, we can write

$$\begin{aligned} \mathrm{Var}(X) &= \mathrm{Var}(U+V) = \mathrm{Var}(U) + \mathrm{Var}(V) \\ \mathrm{Var}(Y) &= \mathrm{Var}(U-V) = \mathrm{Var}(U) + \mathrm{Var}(V) \end{aligned}$$

and remember the definition of variance

$$\mathrm{Var}(U) = \int_{\mathbb{R}} (u - \mathbb{E}[u])^2 \, p_U(u) \mathrm{d}u = \frac{1}{12} = \mathrm{Var}(V)$$

Now we have to calculate the cross-covariance,

$$\mathrm{Cov}(X,Y) = \mathbb{E}_{XY}\Big[\big(X - \mathbb{E}[X]\big)\big(Y - \mathbb{E}[Y]\big)\Big] = \mathbb{E}\Big[\big(X-1\big)Y\Big] = \mathbb{E}\Big[XY\Big] - \mathbb{E}[Y] = \mathbb{E}[U^2 - V^2]$$

and finally

$$\mathrm{Cov}(X,Y) = \mathbb{E}[U^2] - \mathbb{E}[V^2] = 0 \quad \text{therefore} \quad \mathbf{\Sigma}_Z = \begin{bmatrix} \frac{1}{6} & 0 \\ 0 & \frac{1}{6} \end{bmatrix}$$

Note: It would have been much easier to do everything using matrices. For this, one would have to start by noticing that we can write

$$Z = \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}}_{\mathbf{A}} \begin{bmatrix} U \\ V \end{bmatrix} \quad \text{thus} \quad \mathbb{E}[Z] = \mathbf{A} \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} \quad \text{and} \quad \mathbf{\Sigma}_Z = \mathbf{A} \begin{bmatrix} \frac{1}{12} & 0 \\ 0 & \frac{1}{12} \end{bmatrix} \mathbf{A}^\top$$

(b) From the result in (a) we see that $\mathrm{Cov}(X,Y) = 0$ so they are indeed uncorrelated.

To check whether two random variables are independent we have to first calculate their joint pdf $p_{XY}(x,y)$ and compare it to the marginal pdfs $p_X(x)$ and $p_Y(y)$. The random variables will be independent **if, and only if,** we can write $p_{XY}(x,y) = p_X(x)p_Y(y)$

First, recall that

$$\begin{array}{rcl} X & = & U+V \\ Y & = & U-V \end{array} \iff \begin{array}{rcl} U & = & \frac{1}{2}(X+Y) \\ Y & = & \frac{1}{2}(X-Y) \end{array}$$

So let's first check what is the joint pdf. We can simply use the transformation method, which writes

$$p_{XY}(x,y) = p_{UV}(u,v) \times |\det(J)|^{-1}$$

where J is the jacobian of the transformation from (U,V) to (X,Y):

$$J = \begin{bmatrix} \frac{\partial X}{\partial U} & \frac{\partial X}{\partial V} \\ \frac{\partial Y}{\partial U} & \frac{\partial Y}{\partial V} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \Rightarrow |\det(J)| = 2$$

Then

$$p_{XY}(x,y) = p_U\left(\tfrac{1}{2}(x+y)\right) p_V\left(\tfrac{1}{2}(x-y)\right) \times \frac{1}{2}$$

We thus identify a pdf which is constant on a region $\mathcal{S}$ defined by the two uniform marginals. Let's see what $\mathcal{S}$ looks like:

Since $0 \le U \le 1$ and $0 \le V \le 1$, then we have $0 \le X+Y \le 2$ and $0 \le X-Y \le 2$, therefore $0 \le X \le 2$ and $-X \le Y \le X$

So we can write the joint pdf as per

$$p_{XY}(x,y) = \begin{cases} \frac{1}{2}, & 0 \le x \le 2 \quad -x \le y \le x \\ 0, & \text{otherwise} \end{cases}$$

OK, we have the joint pdf. Let's see now what the marginal for X looks like.

Remember that $X = U+V$ with U and V independent. One way of calculating the pdf for the sum of two general RVs is to begin with the CDF and then taking its derivative. Let's see where this gets us:

$$F_X(x) = \mathbb{P}(X \le x) = \mathbb{P}(U+V \le x) = \iint\limits_{u+v \le x} p_{UV}(u,v)\mathrm{d}u\mathrm{d}v = \int_{-\infty}^{+\infty}\int_{-\infty}^{x-u} p_U(u)p_V(v)\mathrm{d}u\mathrm{d}v$$

Rearranging things, we get

$$F_X(x) = \int_{-\infty}^{+\infty} p_U(u)\left(\int_{-\infty}^{x-u} p_V(v)\mathrm{d}v\right)\mathrm{d}u = \int_{-\infty}^{+\infty} p_U(u)F_V(x-u)\mathrm{d}u$$

So the pdf of X can be calculated as

$$p_X(x) = \frac{d}{dx}F_X(x) = \int_{-\infty}^{+\infty} p_U(x)\left(\frac{d}{dx}F_V(x-u)\right)\mathrm{d}u = \int_{-\infty}^{+\infty} p_U(x)p_V(x-u)\mathrm{d}u$$

For our specific case with two uniform distributions, this gives

$$p_X(x) = \int_0^1 p_V(x-u)\mathrm{d}u = \int_0^1 \mathbf{1}_{[0,1]}(x-u)\mathrm{d}u = \begin{cases} 0 & x \le 0 \quad \text{or} \quad x \ge 2 \\ x & 0 \le x \le 1 \\ 2-x & 1 \le x \le 2 \end{cases}$$

We see right away that $p_X(x)$ is not a constant, so we don't even need to calculate the marginal for Y to conclude that the we can not factorize the joint pdf into the marginals. Conclusion, the two RVs are not independent.

-- Exercise 2

$Z = \begin{bmatrix} X \\ Y \end{bmatrix}$ with $Z \sim \mathcal{N}(\mathbf{z} \mid \mu, \Sigma)$ and $\mu = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$

(a) The pdf of a bivariate normal distribution is

$$p_Z(\mathbf{z}) = \frac{1}{2\pi \det(\Sigma)} \exp\left(-\frac{1}{2}(\mathbf{z}-\mu)^\top \Sigma^{-1}(\mathbf{z}-\mu)\right)$$

With $\det(\Sigma) = 1$ and $\Sigma^{-1} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$

We end up with

$$p_Z(x,y) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(2x^2 + y^2 + 2xy - 8x - 6y + 10)\right)$$

(b) The conditional expectation can be written as

$$p_{Y|X=x}(y) = \frac{p_{XY}(x,y)}{p_X(x)} \quad \text{with} \quad p_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-1)^2\right)$$

Which after lots of simplifications gives us:

$$p_{Y|X=x}(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\Big(y - (3-x)\Big)^2\right) = \mathcal{N}(y \mid 3-x, 1)$$

(c) Remember from the CM1 that the best prediction of Y given $X = x$ is the conditional expectation as per

$$\hat{Y} = \mathbb{E}\Big[Y \mid X = x\Big] = 3 - x$$

-- Exercise 3

We can rewrite the loss function to minimize as per

$$\begin{aligned}
\mathbb{E}_{(X,Y)}\Big[(Y - f(X))^2\Big] &= \mathbb{E}_X\Big[\mathbb{E}_{Y|X}\big[(Y-f(X))^2 \mid X\big]\Big] \\
&= \mathbb{E}_X\Big[\mathrm{Var}(Y - f(X) \mid X) + \Big(\mathbb{E}_{Y|X}\big[(Y - f(X) \mid X\big]\Big)^2\Big] \\
&= \mathbb{E}_X\Big[\mathrm{Var}(Y \mid X) + \Big(\mathbb{E}_{Y|X}\big[(Y - f(X) \mid X\big]\Big)^2\Big] \\
&\geq \mathbb{E}_X\Big[\mathrm{Var}(Y \mid X)\Big]
\end{aligned}$$

so we see that to minimize the loss we should choose $f(x) = \mathbb{E}_{Y|X}[Y \mid X = x]$

-- Exercise 4

The model is $Y = \beta_0 + \beta_1 X_1 + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, \sigma^2)$

(a) From the CM we have the expressions

$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ and $\hat{\beta}_1 = \dfrac{c_{XY}}{s_X^2}$ with $c_{XY} = \dfrac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$ and $s_X^2 = \dfrac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$

We first check the unbiasedness of $\hat{\beta}_1$

$\mathbb{E}\big[\hat{\beta}_1\big] = \mathbb{E}_{X_1,\ldots,X_N}\Big[\mathbb{E}\big[\hat{\beta}_1 \mid X_1 = x_1, \ldots, X_N = x_N\big]\Big]$ (explain why we first consider the conditional expectation)

$\mathbb{E}\big[\hat{\beta}_1 \mid X_1 = x_1, \ldots, X_N = x_N\big] = \dfrac{1}{s_X^2} \dfrac{1}{N} \sum_{i=1}^{N} \Big(x_i - \bar{x}\Big)\Big(\mathbb{E}[Y_i \mid X_i = x_i] - \mathbb{E}[\bar{y} \mid X_1 = x_1, \ldots, X_N = x_N]\Big)$
(the X are fixed but the Y are random variables)

Note that,

$$\mathbb{E}\Big[Y_i \mid X_i = x_i\Big] = \beta_0 + \beta_1 x_i + \mathbb{E}[\varepsilon_i] = \beta_0 + \beta_1 x_i$$

and that

$$\mathbb{E}\Big[\bar{y} \mid X_1 = x_1, \ldots, X_N = x_N\Big] = \mathbb{E}\left[\frac{1}{N} \sum_{i=1}^{N} (\beta_0 + \beta_1 x_i + \varepsilon_i)\right] = \beta_0 + \beta_1 \bar{x}$$

so we get

$$\mathbb{E}\big[\hat{\beta}_1 \mid X_1 = x_1, \ldots, X_N = x_N\big] = \frac{1}{s_X^2} \frac{1}{N} \sum_{i=1}^{N} \Big(x_i - \bar{x}\Big)\Big(\beta_0 + \beta_1 x_i - \beta_0 - \beta_1 \bar{x}\Big) = \beta_1$$

Then taking the expectacion along all possible datasets, we get

$$\mathbb{E}[\hat{\beta}_1] = \mathbb{E}_{X_1,\ldots,X_N}\Big[\beta_1\Big] = \beta_1$$

The bias for $\hat{\beta}_0$ is checked similarly.

$$\mathbb{E}[\hat{\beta}_0 \mid X_1 = x_1, \ldots, X_N = x_N] = \mathbb{E}[\bar{Y} - \hat{\beta}_1 \bar{X} \mid X_1 = x_1, \ldots, X_N = x_N]$$

Which then gives us

$$\mathbb{E}[\bar{Y} \mid X_1 = x_1, \ldots, X_N = x_N] = \beta_0 + \beta_1 \bar{x}$$

and

$$\mathbb{E}[\hat{\beta}_1 \bar{X} \mid X_1 = x_1, \ldots, X_N = x_N] = \bar{x}\, \mathbb{E}[\hat{\beta}_1 \mid X_1 = x_1, \ldots, X_N = x_N] = \beta_1 \bar{x}$$

so in the end we get

$$\mathbb{E}[\hat{\beta}_0 \mid X_1 = x_1, \ldots, X_N = x_N] = \beta_0 + \beta_1 \bar{x} - \beta_1 \bar{x} = \beta_0$$

(b) OK, now let's get the variances.

Remember that $\hat{\beta}_1 = \dfrac{c_{XY}}{s_X^2} = \left(\dfrac{1}{N} \sum_i x_i y_i - \bar{x}\bar{y}\right) \dfrac{1}{s_X^2}$ so we will first rewrite it so that things get easier later

Note that

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \text{and} \quad \bar{y} = \beta_0 + \beta_1 \bar{x} + \bar{\varepsilon} \quad \text{where} \quad \bar{\varepsilon} = \frac{1}{N} \sum_i \varepsilon_i$$

and that

$$\frac{1}{N}\sum_i x_i y_i = \frac{1}{N}\sum_i x_i(\beta_0 + \beta_1 x_i + \varepsilon_i) = \beta_0 \bar{x} + \beta_1 \frac{1}{N}\sum_i x_i^2 + \frac{1}{N}\sum_i x_i \varepsilon_i$$

and

$$\bar{x}\bar{y} = \beta_0 \bar{x} + \beta_1 (\bar{x})^2 + \bar{x}\bar{\varepsilon}$$

Then

$$\hat{\beta}_1 = \left(\beta_1 s_X^2 + \frac{1}{N}\sum_i x_i \varepsilon_i - \bar{x}\bar{\varepsilon}\right)\frac{1}{s_X^2} = \beta_1 + \frac{1}{s_X^2}\frac{1}{N}\sum_i (x_i - \bar{x})\varepsilon_i$$

This way of rewriting the estimator $\hat{\beta}_1$ is very insightful, since now we can easily write that

$$\mathrm{Var}(\hat{\beta}_1) = \mathbb{E}_{X_1,\ldots,X_N}\left[\mathrm{Var}_X\left(\hat{\beta}_1 \mid X_1 = x_1, \ldots, X_N = x_N\right)\right] = \mathbb{E}_{X_1,\ldots,X_N}\left[\frac{1}{s_X^4}\frac{1}{N^2}\sum_i (x_i - \bar{x})^2 \sigma^2\right]$$

so we get

$$\mathrm{Var}(\hat{\beta}_1) = \frac{1}{N^2}\frac{1}{s_X^4} s_X^2 N s_X^2 \sigma^2 = \frac{1}{N}\frac{\sigma^2}{s_X^2}$$

What about $\hat{\beta}_0$?

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \left(\beta_0 + \beta_1 \bar{x} + \bar{\varepsilon}\right) - \left(\beta_1 + \frac{1}{s_X^2}\frac{1}{N}\sum_i (x_i - \bar{x})\varepsilon_i\right)\bar{x}$$

which gives us

$$\hat{\beta}_0 = \beta_0 + \bar{\varepsilon} - \frac{1}{s_X^2}\frac{1}{N}\sum_i (x_i - \bar{x})\bar{x}\varepsilon_i$$

Theferore,

$$\mathrm{Var}\left(\hat{\beta}_0\right) = \frac{\sigma^2}{N} + \frac{1}{N^2 s_X^4}\sum_i (x_i - \bar{x})^2 \bar{x}^2 \sigma^2 = \frac{1}{N}\sigma^2\left(1 + \frac{\bar{x}^2}{s_X^2}\right)$$

(c)

$$\hat{m}(x) = \hat{\beta}_0 + \hat{\beta}_1 x = \left(\bar{y} - \hat{\beta}_1 \bar{x}\right) + \hat{\beta}_1 x = \bar{y} + \hat{\beta}_1 (x - \bar{x})$$

Taking back the expression from item (b)

$$\begin{aligned}
\hat{m}(x) &= \bar{y} + \left(\beta_1 + \frac{1}{s_X^2}\frac{1}{N}\sum_i (x_i - \bar{x})\varepsilon_i\right)(x - \bar{x}) \\
&= \frac{1}{N}\sum_i \Big(\beta_0 + \beta_1 x_i + \varepsilon_i\Big) + \Big(\beta_1 + \frac{1}{s_X^2}\frac{1}{N}\sum_i (x_i - \bar{x})\varepsilon_i\Big)(x - \bar{x}) \\
&= \beta_0 + \beta_1\bar{x} + \bar{\varepsilon} + \beta_1 x - \beta_1\bar{x} + \frac{(x-\bar{x})}{N s_X^2}\sum_i (x_i - \bar{x})\varepsilon_i \\
&= \beta_0 + \beta_1 x + \bar{\varepsilon} + \frac{(x-\bar{x})}{N s_X^2}\sum_i (x_i - \bar{x})\varepsilon_i \\
\mathbb{E}[\hat{m}(x)] &= \beta_0 + \beta_1 x
\end{aligned}$$

(d) Finally, the conditional variance on a given choice of dataset $x_1, \ldots, x_N$ is

$$\mathrm{Var}_X\Big(\hat{m}(x)\Big) = \mathrm{Var}_X\left(\beta_0 + \beta_1 x + \bar{\varepsilon} + \frac{(x-\bar{x})}{N s_X^2}\sum_i (x_i - \bar{x})\varepsilon_i\right)$$

which can be simplified as per

$$\begin{aligned}
\mathrm{Var}_X\Big(\hat{m}(x)\Big) &= \mathrm{Var}\left(\frac{1}{N}\sum_i \varepsilon_i + \tfrac{(x-\bar{x})}{N s_X^2}\sum_i (x_i - \bar{x})\varepsilon_i\right) \\
&= \mathrm{Var}\left(\frac{1}{N}\sum_i \Big[1 + \tfrac{(x-\bar{x})}{s_X^2}(x_i - \bar{x})\Big]\varepsilon_i\right) \\
&= \mathrm{Var}\left(\frac{1}{N}\sum_i \Big[1 + \tfrac{(x-\bar{x})}{s_X^2}(x_i - \bar{x})\Big]\varepsilon_i\right) \\
&= \frac{1}{N^2}\sum_i \Big[1 + \tfrac{(x-\bar{x})}{s_X^2}(x_i - \bar{x})\Big]^2 \mathrm{Var}(\varepsilon_i) \\
&= \frac{\sigma^2}{N^2}\sum_i \left(1 + \frac{2(x-\bar{x})(x_i-\bar{x})}{s_X^2} + \frac{(x-\bar{x})^2(x_i-\bar{x})^2}{s_X^4}\right) \\
&= \frac{\sigma^2}{N}\sum_i \left(1 + \frac{(x-\bar{x})^2}{s_X^2}\right)
\end{aligned}$$

Notice that to get the unconditional variance we would have to take the expectation with respect to the x_i, which can be quite cumbersome and not that much insightful. We won't do that.