Question 1

Consider the probability space $(\Omega, \mathcal{F}, \mathrm{P})$. Recall* the definition of an indicator variable for an event $A \in \mathcal{F}$, denoted $\mathbb{I}_A$ (or $\mathbb{I}(A)$) and defined for $\omega \in \Omega$ by

$$\mathbb{I}_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \notin A. \end{cases}$$

(*See Definition 7.3.2 from Prof. Veraart's notes in Term 1.)

(a) Is $\mathbb{I}_A$ a discrete random variable or a continuous random variable?

(b) If $\mathbb{I}_A$ is discrete, write down its probability mass function, or if it is continuous, write down its probability density function.

(c) Compute $\mathrm{E}(\mathbb{I}_A)$.

Solution to Question 1

Part (a): Since $\mathrm{Im}(\mathbb{I}_A) = \{\mathbb{I}_A(\omega) : \omega \in \Omega\} = \{0, 1\}$, and is therefore countable, the random variable $\mathbb{I}_A$ is a discrete random variable (Definition 7.2.1 in Prof. Veraart's notes from Term 1).

Part (b): Since $\mathbb{I}_A$ is discrete, it has a probability mass function (p.m.f.). One might directly be able to write down the p.m.f. as:

$$p_{\mathbb{I}_A}(x) = \begin{cases} \mathrm{P}(A), & \text{if } x = 1, \\ 1 - \mathrm{P}(A), & \text{if } x = 0. \end{cases}$$

However, **the box on the next page contains a careful derivation**.

Part (c): There are two (similar) approaches to computing the expectation of $\mathbb{I}_A$.

The first approach is to directly use the definition of expectation for a discrete random variable:

$$\begin{aligned} \mathrm{E}(X) &= \sum_{x \in \mathrm{Im}(\mathbb{I}_A)} x p_{\mathbb{I}_A}(x) \\ &= 1 \cdot p_{\mathbb{I}_A}(1) + 0 \cdot p_{\mathbb{I}_A}(0) \\ &= 1 \cdot \mathrm{P}(A) + 0 \cdot (1 - \mathrm{P}(A)) \\ \Rightarrow \mathrm{E}(X) &= \mathrm{P}(A). \end{aligned}$$

The second approach takes a shortcut. One can consider $\mathbb{I}_A$ to be random variable following a Bernoulli distribution with parameter p, i.e. $\mathbb{I}_A \sim \mathrm{Bern}(\lambda)$, where $\lambda = p_{\mathbb{I}_A}(1) = \mathrm{P}(A)$. Therefore, $\mathrm{E}(\mathbb{I}_A) = \lambda = \mathrm{P}(A)$.

(One would usually use "p" as the parameter for a Bernoulli distribution, but "λ" was used here to avoid confusion with the p.m.f. $p_{\mathbb{I}_A}$.)

The first approach is from [3, Chap. 24, p. 203-204] and the second approach is from [1, Theorem 4.4.2, p. 164].

Part (b): (in detail)

Recall the definition of a probability mass function from Prof. Veraart's notes from Term 1:
Definition 7.2.4 (Probability mass function)**.** The **probability mass function** (p.m.f.) of the discrete random variable X is defined as the function $p_X : \mathbb{R} \to [0,1]$ given by

$$p_X(x) = \mathrm{P}\left(\{\omega \in \Omega \,:\, X(\omega) = x\}\right).$$

Now, since $\mathrm{Im}\left(\mathbb{I}_A\right) = \{\mathbb{I}_A(\omega) \,:\, \omega \in \Omega\} = \{0,1\}$, there are only three cases to consider and so one can directly use Definition 7.2.4 and the definition of $\mathbb{I}_A$ to obtain:

$$\begin{aligned}
p_{\mathbb{I}_A}(x) &= \mathrm{P}\left(\{\omega \in \Omega \,:\, \mathbb{I}_A(\omega) = x\}\right) \\
&= \begin{cases} \mathrm{P}\left(\{\omega \in \Omega \,:\, \mathbb{I}_A(\omega) = 1\}\right), & \text{if } x = 1, \\ \mathrm{P}\left(\{\omega \in \Omega \,:\, \mathbb{I}_A(\omega) = 0\}\right), & \text{if } x = 0, \\ \mathrm{P}\left(\{\omega \in \Omega \,:\, \mathbb{I}_A(\omega) = x\}\right), & \text{if } x \notin \{0,1\} \end{cases} \\
&= \begin{cases} \mathrm{P}\left(\{\omega \in \Omega \,:\, \omega \in A\}\right), & \text{if } x = 1, \\ \mathrm{P}\left(\{\omega \in \Omega \,:\, \omega \notin A\}\right), & \text{if } x = 0, \\ \mathrm{P}\left(\varnothing\right), & \text{if } x \notin \{0,1\} \end{cases} \qquad (1) \\
&= \begin{cases} \mathrm{P}\left(A\right), & \text{if } x = 1, \\ \mathrm{P}\left(A^c\right), & \text{if } x = 0, \\ 0, & \text{if } x \notin \{0,1\} \end{cases} \qquad (2) \\
\Rightarrow p_{\mathbb{I}_A}(x) &= \begin{cases} \mathrm{P}\left(A\right), & \text{if } x = 1, \\ 1 - \mathrm{P}\left(A\right), & \text{if } x = 0, \\ 0, & \text{otherwise} \end{cases}
\end{aligned}$$

Equation (1) leads to Equation (2) by recalling that $A \in \mathcal{F} \Rightarrow A \subseteq \Omega$.

Equation (1) uses $\varnothing$ to denote the empty set; the set is empty because there are no elements of Ω with $\mathbb{I}_A(\omega) = x$ and $x \notin \mathrm{Im}\left(\mathbb{I}_A\right) = \{0,1\}$.

Question 2

Prove Theorem 1.1.2 from the notes:

Theorem 1.1.2. Given two arbitrary random variables X and Y with a specified joint distribution, suppose that X and Y both have finite means. Then the function g of X that minimises $\mathrm{E}[(Y - g(X))^2]$ is $g(X) = \mathrm{E}[Y|X]$, i.e.

$$\min_g \mathrm{E}[(Y - g(X))^2] = \mathrm{E}[(Y - \mathrm{E}[Y|X])^2]$$

Solution to Question 2

We need to prove:

$$\min_g \mathrm{E}[(Y - g(X))^2] = \mathrm{E}[(Y - \mathrm{E}[Y|X])^2]$$

where g ranges over all functions of X. Using the alternative notation, this is equivalent to:

$$\min_g \mathrm{E}_{X,Y}[(Y - g(X))^2] = \mathrm{E}_{X,Y}[\left(Y - \mathrm{E}_{Y|X}[Y]\right)^2] \tag{3}$$

Let us follow the approach for Theorem 1.1.1, and try to prove that for any $g(X)$

$$\mathrm{E}_{X,Y}[(Y - g(X))^2] \geq \mathrm{E}_{X,Y}[\left(Y - \mathrm{E}_{Y|X}[Y]\right)^2], \tag{4}$$

which is equivalent to Equation (3) (recall that $\mathrm{E}_{Y|X}[Y]$ is a function of X).

Starting with the expression on the left-hand side of the equation, we add and subtract $\mathrm{E}_{Y|X}[Y]$ inside the parentheses:

$$\begin{aligned}
\mathrm{E}_{X,Y}[(Y - g(X))^2] &= \mathrm{E}_{X,Y}[\left(\left\{Y - \mathrm{E}_{Y|X}[Y]\right\} + \left\{\mathrm{E}_{Y|X}[Y] - g(X)\right\}\right)^2] \\
&= \mathrm{E}_{X,Y}[\left(Y - \mathrm{E}_{Y|X}[Y]\right)^2 \\
&\qquad + 2\left(Y - \mathrm{E}_{Y|X}[Y]\right)\left(\mathrm{E}_{Y|X}[Y] - g(X)\right) \\
&\qquad + \left(\mathrm{E}_{Y|X}[Y] - g(X)\right)^2]
\end{aligned}$$

$$\begin{aligned}
\Rightarrow \mathrm{E}_{X,Y}[(Y - g(X))^2] &= \mathrm{E}_{X,Y}[\left(Y - \mathrm{E}_{Y|X}[Y]\right)^2] \\
&\qquad + 2\mathrm{E}_{X,Y}[\left(Y - \mathrm{E}_{Y|X}[Y]\right)\left(\mathrm{E}_{Y|X}[Y] - g(X)\right)] \\
&\qquad + \mathrm{E}_{X,Y}[\left(\mathrm{E}_{Y|X}[Y] - g(X)\right)^2]
\end{aligned} \tag{5}$$

where the last equality follows from the linearity of the expectation. We now take a closer look at the second term, without the factor of 2:

$$\mathrm{E}_{X,Y}[\left(Y - \mathrm{E}_{Y|X}[Y]\right)\left(\mathrm{E}_{Y|X}[Y] - g(X)\right)].$$

Since $\mathrm{E}_{Y|X}[Y]$ is a function of the random variable X, this implies that $\mathrm{E}_{Y|X}[Y] - g(X)$ is a function only of X, and we can relabel it as $\gamma(X) = \mathrm{E}_{Y|X}[Y] - g(X)$. Then, using Proposition 1.1.8 in the notes, which is a generalisation of the Law of Total Probability, the second term becomes

$$\begin{aligned}
\mathrm{E}_{X,Y}[\left(Y - \mathrm{E}_{Y|X}[Y]\right)\left(\mathrm{E}_{Y|X}[Y] - g(X)\right)] &= \mathrm{E}_{X,Y}[\left(Y - \mathrm{E}_{Y|X}[Y]\right)\gamma(X)] \\
&= \mathrm{E}_X[\mathrm{E}_{Y|X}[\left(Y - \mathrm{E}_{Y|X}[Y]\right)\gamma(X)]],
\end{aligned}$$

(directly using Proposition 1.1.8 with $h(X,Y) = \left(Y - \mathrm{E}_{Y|X}[Y]\right)\gamma(X)$). We now focus our attention on the inner expectation,

$$\mathrm{E}_{Y|X}[\left(Y - \mathrm{E}_{Y|X}[Y]\right)\gamma(X)]. \tag{6}$$

Although Exercise 1.1.6 will prove the relation

$$\mathrm{E}_{Y|X}[g(X)h(Y)] = g(X)\mathrm{E}_{Y|X}[h(Y)],$$

a slight modification of the proof, using $h(X,Y)$ instead of $h(Y)$, will prove

$$\mathrm{E}_{Y|X}[g(X)h(X,Y)] = g(X)\mathrm{E}_{Y|X}[h(X,Y)]. \tag{7}$$

We use Equation (7) on Equation (6), and obtain

$$\mathrm{E}_{Y|X}[\left(Y - \mathrm{E}_{Y|X}[Y]\right)\gamma(X)] = \gamma(X)\mathrm{E}_{Y|X}[\left(Y - \mathrm{E}_{Y|X}[Y]\right)].$$

We next focus on the term $\mathrm{E}_{Y|X}\left(Y - \mathrm{E}_{Y|X}[Y]\right)$. However, we first notice that

$$\mathrm{E}_{Y|X}\left(\mathrm{E}_{Y|X}[Y]\right) = \mathrm{E}_{Y|X}[Y]. \tag{8}$$

This might seem obvious, since when the expectation is not conditional on X, we have a similar result

$$\mathrm{E}_{Y}\left(\mathrm{E}_{Y}[Y]\right) = \mathrm{E}_{Y}[Y].$$

But in this case the inner expectation $\mathrm{E}_Y[Y]$ is simply a number, a constant $c = \mathrm{E}_Y[Y]$, and for any expectation, $\mathrm{E}\,(c) = c$.

On the other hand, $\mathrm{E}_{Y|X}[Y]$ is not a constant, but a function of the random variable X. However, one can see that Equation (8) follows by writing $\alpha(X) = \mathrm{E}_{Y|X}[Y]$ and $\beta(Y) = 1$, and using Exercise 1.1.6:

$$\begin{aligned}
\mathrm{E}_{Y|X}\left(\mathrm{E}_{Y|X}[Y]\right) &= \mathrm{E}_{Y|X}[\alpha(X)] \\
&= \mathrm{E}_{Y|X}[\alpha(X)\cdot 1] \\
&= \mathrm{E}_{Y|X}[\alpha(X)\beta(Y)] \\
&= \alpha(X)\mathrm{E}_{Y|X}[\beta(Y)] \\
&= \alpha(X)\mathrm{E}_{Y|X}[1] \\
&= \alpha(X)\cdot 1 \\
&= \alpha(X) \\
&= \mathrm{E}_{Y|X}[Y]
\end{aligned}$$

(Although a few of the steps are trivial, the above carefully establishes Equation (8).)

We now return to the term $\mathrm{E}_{Y|X}\left(Y - \mathrm{E}_{Y|X}[Y]\right)$. Using the linearity of expectation and Equation (8),

$$\mathrm{E}_{Y|X}\left(Y - \mathrm{E}_{Y|X}[Y]\right) = \mathrm{E}_{Y|X}\left(Y\right) - \mathrm{E}_{Y|X}\left(\mathrm{E}_{Y|X}[Y]\right) = \mathrm{E}_{Y|X}\left(Y\right) - \mathrm{E}_{Y|X}\left(Y\right) = 0. \tag{9}$$

Equation (9) is the key result. We can now return to the expectation of the second term, and including the intermediate steps one obtains

$$\begin{aligned}
\mathrm{E}_{X,Y}[\left(Y - \mathrm{E}_{Y|X}[Y]\right)\left(\mathrm{E}_{Y|X}[Y] - g(X)\right)] &= \mathrm{E}_{X,Y}[\left(Y - \mathrm{E}_{Y|X}[Y]\right)\gamma(X)] \\
&= \mathrm{E}_X[\mathrm{E}_{Y|X}[\left(Y - \mathrm{E}_{Y|X}[Y]\right)\gamma(X)]], \\
&= \mathrm{E}_X[\gamma(X)\mathrm{E}_{Y|X}[\left(Y - \mathrm{E}_{Y|X}[Y]\right)]], \\
&= \mathrm{E}_X[\gamma(X)\cdot(0)] \\
&= \mathrm{E}_X[0] \\
&= 0.
\end{aligned}$$

Now, returning to Equation (5),

$$\begin{aligned}\mathrm{E}_{X,Y}[(Y-g(X))^2] = \mathrm{E}_{X,Y}[\,&\big(Y-\mathrm{E}_{Y|X}[Y]\big)^2] \\ &+ 2\cdot 0 \\ &+ \mathrm{E}_{X,Y}[\big(\mathrm{E}_{Y|X}[Y]-g(X)\big)^2]\end{aligned}$$

$$\Rightarrow \mathrm{E}_{X,Y}[(Y-g(X))^2] = \mathrm{E}_{X,Y}[\,\big(Y-\mathrm{E}_{Y|X}[Y]\big)^2] + \mathrm{E}_{X,Y}[\big(\mathrm{E}_{Y|X}[Y]-g(X)\big)^2]$$

$$\Rightarrow \mathrm{E}_{X,Y}[(Y-g(X))^2] \geq \mathrm{E}_{X,Y}[\,\big(Y-\mathrm{E}_{Y|X}[Y]\big)^2], \tag{10}$$

where the inequality in the last line follows because $\mathrm{E}_{X,Y}[\big(\mathrm{E}_{Y|X}[Y]-g(X)\big)^2] \geq 0$.

To see this, first note that $\mathrm{E}_{Y|X}[Y]-g(X)$ is a function only of the random variable X, and so we can write it as $\delta(X) = \mathrm{E}_{Y|X}[Y]-g(X)$. Even though there are no restrictions on the values X can take, and therefore no restrictions on the values $\delta(X)$ can take, its square $(\delta(X))^2$ will only be able to take nonnegative values, i.e.

$$\begin{aligned}&(\delta(X))^2 \geq 0 \\ \Rightarrow &\mathrm{E}_{X,Y}[(\delta(X))^2] \geq 0 \\ \Rightarrow &\mathrm{E}_{X,Y}[\big(\mathrm{E}_{Y|X}[Y]-g(X)\big)^2] \geq 0.\end{aligned}$$

This proves Equation (10), which is the same as Equation (4), which proves the result.

Question 3

Given a random variable X, the median of the distribution of X is a value m such that $\mathrm{P}(X \leq m) \geq \frac{1}{2}$ and $\mathrm{P}(X \geq m) \geq \frac{1}{2}$. Show that if X is a **continuous** random variable with probability density function $f_X(x)$ then

$$\min_c \mathrm{E}(\,|X - c|\,) = \mathrm{E}(\,|X - m|\,)\,.$$

Solution to Question 3

In this case, we assume that X has a continuous pdf f_X (which we shall just write as f below). Then one can define the median of X as the value m such that

$$\int_{-\infty}^{m} f(x)\,\mathrm{d}x = \int_{m}^{\infty} f(x)\,\mathrm{d}x = \frac{1}{2}.$$

For the random variable X, let us define the function H such that for a given value $c \in \mathbb{R}$, $H(X) = \mathrm{E}(|X - c|)$. Then,

$$\begin{aligned} H(c) &= \int_{-\infty}^{\infty} |x - c|\, f(x)\,\mathrm{d}x \\ &= \int_{-\infty}^{c} |x - c|\, f(x)\,\mathrm{d}x + \int_{c}^{\infty} |x - c|\, f(x)\,\mathrm{d}x \\ \Rightarrow H(c) &= -\int_{-\infty}^{c} (x - c)\, f(x)\,\mathrm{d}x + \int_{c}^{\infty} (x - c)\, f(x)\,\mathrm{d}x \end{aligned}$$

At this point, in order to find the value of c which minimises H (and hence minimises $\mathrm{E}(|X - c|)$), we decide to compute the derivative $\frac{\mathrm{d}H}{\mathrm{d}c}$. However, since c occurs in the integral limits, the Leibniz integral rule [2] will be helpful:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{a(t)}^{b(t)} g(x,t)\mathrm{d}x\right) = g(b(t),t)\cdot\left(\frac{\mathrm{d}}{\mathrm{d}t}b(t)\right) - g(a(t),t)\cdot\left(\frac{\mathrm{d}}{\mathrm{d}t}a(t)\right) + \int_{a(t)}^{b(t)}\left(\frac{\partial}{\partial t}g(x,t)\right)\mathrm{d}x.$$

Looking at the first term in $H(c)$, using the variable c rather than t, and defining the functions $g(x,c) = (x - c)\, f(x)$, $b(c) = c$, $a(c) = -\infty$ (so a is a constant function):

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}c}\left[-\int_{-\infty}^{c} (x - c)\, f(x)\,\mathrm{d}x\right] \\ &= -\frac{\mathrm{d}}{\mathrm{d}c}\left[\int_{-\infty}^{c} (x - c)\, f(x)\,\mathrm{d}x\right] \\ &= -\left[(c - c)f(c)\cdot 1 - 0 + \int_{-\infty}^{c} \frac{\partial}{\partial c}\left((x - c)\, f(x)\right)\mathrm{d}x\right] \\ &= -\left[0 + \int_{-\infty}^{c} (-1\cdot f(x))\,\mathrm{d}x\right] \\ &= \int_{-\infty}^{c} f(x)\,\mathrm{d}x \end{aligned}$$

Again, note that here only the upper limit is a function of c, while the lower limit is a constant (hence the zero in the third line). Similarly,

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}c}\left[\int_c^\infty (x-c)\,f(x)\,\mathrm{d}x\right] \\
&= 0 - (c-c)f(c)\cdot 1 + \int_c^\infty \frac{\partial}{\partial c}\left((x-c)\,f(x)\right)\mathrm{d}x \\
&= 0 + \int_c^\infty \left(-1\cdot f(x)\right)\mathrm{d}x \\
&= -\int_c^\infty f(x)\,\mathrm{d}x
\end{aligned}$$

Therefore,

$$\frac{\mathrm{d}H}{\mathrm{d}c} = \int_{-\infty}^c f(x)\,\mathrm{d}x - \int_c^\infty f(x)\,\mathrm{d}x = \mathrm{P}(X \le c) - \mathrm{P}(X > c)\,.$$

Suppose $c = c^*$ is a value that satisfies $\frac{\mathrm{d}H}{\mathrm{d}c}(c^*) = 0$. Then

$$\mathrm{P}(X \le c^*) - \mathrm{P}(X > c^*) = 0.$$

However, for every random variable X:

$$\mathrm{P}(X \le c^*) + \mathrm{P}(X > c^*) = 1,$$

so adding the two equations together yields

$$\begin{aligned}
2\mathrm{P}(X \le c^*) &= 1 \\
\Rightarrow \mathrm{P}(X \le c^*) &= \frac{1}{2}
\end{aligned}$$

which shows that c^* is a median , i.e. $c^* = m$. Note that we also have $\mathrm{P}(X > c^*) = \frac{1}{2} = \mathrm{P}(X > m)$.

Although we have shown that $\frac{\mathrm{d}H}{\mathrm{d}c}(m) = 0$, we have not shown whether or not it is a maximum or a minimum. One could try computing the second derivative, but $\frac{\mathrm{d}^2 H}{\mathrm{d}c^2} = 0$. Rather, define the function

$$L(c) = \frac{\mathrm{d}H}{\mathrm{d}c} = \mathrm{P}(X \le c) - \mathrm{P}(X > c)\,.$$

Then, $L(m) = 0$. Consider a value $\epsilon > 0$. Then $L(m-\epsilon) < 0$, because $\mathrm{P}(X \le m-\epsilon) < \frac{1}{2}$ and $\mathrm{P}(X > m-\epsilon) > \frac{1}{2}$. Similarly, one can show $L(m+\epsilon) > 0$. Therefore, the gradient of H is decreasing before m, zero at m and increasing after m, and so m is a minimum for H. This completes the proof of the result.

Question 4 (Challenge)

Question 1 proves the result in a special case, since it is assumed that X is continuous and it is assumed that a p.d.f. exists. Prove the result in general:

$$\min_c \mathrm{E}\left(|X - c| \right) = \mathrm{E}\left(|X - m| \right),$$

when X is either discrete or continuous, only using the properties of the expectation. As before, m is the median of the distribution of X.

Solution to Question 4

The goal is to show that, for a random variable X, with a median m defined by $\mathrm{P}\left(X \leq m\right) \geq \frac{1}{2}$ and $\mathrm{P}\left(X \geq m\right) \geq \frac{1}{2}$, then for any value a, $\mathrm{E}\left(|X-a|\right) \geq \mathrm{E}\left(|X-m|\right)$. This is equivalent to showing that $\mathrm{E}\left(|X-a| - |X-m|\right) \geq 0$, which leads us to consider the function

$$G\left(X\right) = |X-a| - |X-m|.$$

There are two cases: (i) $a \leq m$ and (ii) $a > m$. For the moment, let us consider case (i), and assume $a \leq m$.

In order to investigate the value of $G\left(X\right)$, we need to look at the three cases:

1. $X < a$,
2. $a \leq X \leq m$,
3. $X > m$.

Case 1: For $X < a (\leq m)$:

$$\begin{aligned} |X-a| - |X-m| &= -\left(X-a\right) - \left[-\left(X-m\right)\right] \\ &= -X + a + X - m \\ &= a - m. \end{aligned}$$

Case 3: For $X > m (\geq a)$:

$$\begin{aligned} |X-a| - |X-m| &= \left(X-a\right) - \left(X-m\right) \\ &= X - a - X + m \\ &= -\left(a-m\right). \end{aligned}$$

Case 2: For $a \leq X \leq m$, the situation is slightly different:

$$\begin{aligned} |X-a| - |X-m| &= \left(X-a\right) - \left[-\left(X-m\right)\right] \\ &= X - a + \left(X-m\right) \\ &= 2X - a - m, \end{aligned}$$

and then since $X \geq a \Rightarrow 2X \geq 2a$,

$$|X-a| - |X-m| \geq 2a - a - m = a - m.$$

We can summarise these three cases in two cases:

$$\begin{aligned} X \leq m: \qquad & G\left(X\right) = |X-a| - |X-m| \geq a - m \\ X > m: \qquad & G\left(X\right) = |X-a| - |X-m| = -(a-m) \end{aligned}$$

Now we use the indicator variable $\mathbb{I}_{X\leq m} = \mathbb{I}(X \leq m)$ (see Question 1):

$$\begin{aligned}
\text{if } X \leq m,\ G(X) \geq (a-m)& \\
&\Rightarrow G(X)\,\mathbb{I}(X \leq m) \geq (a-m)\mathbb{I}(X \leq m) \\
&\Rightarrow \mathrm{E}(G(X)\,\mathbb{I}(X \leq m)) \geq \mathrm{E}((a-m)\mathbb{I}(X \leq m)) \\
&\Rightarrow \mathrm{E}(G(X)\,\mathbb{I}(X \leq m)) \geq (a-m)\mathrm{E}(\mathbb{I}(X \leq m)) \\
&\Rightarrow \mathrm{E}(G(X)\,\mathbb{I}(X \leq m)) \geq (a-m)\mathrm{P}(X \leq m) \qquad (11)
\end{aligned}$$

Similarly, using the indicator variable $\mathbb{I}(X > m)$, we can show that

$$\text{if } X > m \Rightarrow \mathrm{E}(G(X)\,\mathbb{I}(X > m)) = -(a-m)\mathrm{P}(X > m) \qquad (12)$$

Noticing that we use the indicator variables to partition 1:

$$\begin{aligned}
1 &= \mathbb{I}(X \leq m) + \mathbb{I}(X > m) \\
\Rightarrow G(X) &= G(X)\mathbb{I}(X \leq m) + G(X)\mathbb{I}(X > m) \\
\Rightarrow \mathrm{E}[G(X)] &= \mathrm{E}[G(X)\mathbb{I}(X \leq m)] + \mathrm{E}[G(X)\mathbb{I}(X > m)] \qquad (13)
\end{aligned}$$

where the last line follows from the linearity of expectation.

Now, recalling Inequality (11) and Equation (12) above:

$$\begin{aligned}
\mathrm{E}(G(X)\,I(X \leq m)) &\geq \quad (a-m)\,\mathrm{P}(X \leq m) \\
\mathrm{E}(G(X)\,I(X > m)) &= -(a-m)\,\mathrm{P}(X > m),
\end{aligned}$$

adding these two equations and using Equation (13) gives

$$\mathrm{E}(G(X)) \geq (a-m)\left[\mathrm{P}(X \leq m) - \mathrm{P}(X > m)\right]. \qquad (14)$$

We are almost there. Recall that for any random variable X,

$$\begin{aligned}
&\mathrm{P}(X \leq m) + \mathrm{P}(X > m) = 1 \\
\Rightarrow &\mathrm{P}(X \leq m) + \mathrm{P}(X > m) - 1 = 0.
\end{aligned}$$

Then, looking at the expression in brackets in Inequality (14),

$$\begin{aligned}
\mathrm{P}(X \leq m) - \mathrm{P}(X > m) &= \mathrm{P}(X \leq m) - \mathrm{P}(X > m) + 0 \\
&= \mathrm{P}(X \leq m) - \mathrm{P}(X > m) + (\mathrm{P}(X \leq m) + \mathrm{P}(X > m) - 1) \\
&= 2\mathrm{P}(X \leq m) - 1.
\end{aligned}$$

Now, recall from the definition of the median m that $\mathrm{P}(X \leq m) \geq \frac{1}{2}$. Then,

$$\begin{aligned}
&\mathrm{E}(G(X)) \geq (a-m)\left[\mathrm{P}(X \leq m) - \mathrm{P}(X > m)\right] = (a-m)\left[2\mathrm{P}(X \leq m) - 1\right] \\
\Rightarrow &\mathrm{E}(G(X)) \geq (a-m)\left[2 \cdot \left(\tfrac{1}{2}\right) - 1\right] \\
\Rightarrow &\mathrm{E}(G(X)) \geq 0.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathrm{E}(|X-a| - |X-m|) \geq 0 \\
\Rightarrow &\mathrm{E}(|X-a|) - \mathrm{E}(|X-m|) \geq 0 \\
\Rightarrow &\mathrm{E}(|X-a|) \geq \mathrm{E}(|X-m|),
\end{aligned}$$

using the linearity of the expectation. This shows that m minimises the function $\mathrm{E}(|X-a|)$ (as a function of a), for the case $a \leq m$. However, the same argument can be used similarly for the case $a > m$, which proves the result.

(Actually, the above only proves the result for case (i), when $a \leq m$. Fortunately, case (ii), when $a > m$ is virtually identical. Defining $G(X) = |X - a| - |X - m|$, we consider the cases:

$$\begin{aligned}
(1)\ X \leq m < a \Rightarrow |X - a| - |X - m| &= -(X - a) - [-(X - m)] \\
&= a - m \\
(2)\ m < X \leq a \Rightarrow |X - a| - |X - m| &= -(X - a) - (X - m) \\
&= -2X + a + m \\
&\geq -2a + a + m = -(a - m) \\
(3)\ X > a > m \Rightarrow |X - a| - |X - m| &= (X - a) - (X - m) \\
&= -(a - m)
\end{aligned}$$

which can be summarised by

$$\begin{aligned}
X \leq m: &\qquad G(X) = |X - a| - |X - m| = a - m \\
X > m: &\qquad G(X) = |X - a| - |X - m| \geq -(a - m)
\end{aligned}$$

and from here the result follows as in the first case.)

Some comments

The above proof seems very long, but only because all the steps have been carefully written out, since this is perhaps the first time we have used indicator variables in this way. In hindsight, the proof itself is actually quite easy; (1) we want to prove something is a minimum, so (2) we rewrite it as a function $G(X)$ that we want to show has some simpler property (nonnegative expectation), (3) we check a few cases to see the value of this function over the whole real line, (4) it turns out there are only really two special regions worth considering ($X \leq m$ and $X > m$), (5) use the indicator function to split the definition of $G(X)$ over the two regions, (6) some straightforward algebra, and we have the result.

That being said, this proof is only easy if you know the technique—if one has never seen this sort of technique before, it requires a lot of thought!

Question 5 (using R)

Suppose there is a file named `file1.txt` which contains the following data:

```
x,y
2,3
4,6
6,9
8,12
```

(Either download the file from Blackboard, or copy-paste the data into a file and name it `file1.txt`.)

(a) Use the function `read.table` to read the data from `file1.txt` into a data frame object named `df`.

(b) Extract a vector named `x`, containing values $(2, 4, 6, 8)$ from the data frame `df`. Similarly, extract a vector named `y`, containing values $(3, 6, 9, 12)$ from the data frame `df`.

(c) Create a vector named `z` which is the mean of the two vectors `x` and `y`, i.e. `z` contains four values, the first of which is $(2 + 3)/2 = 2.5$.

(d) Add the vector `z` to the data frame `df` so that `df` contains three columns, `x`, `y` and `z`.

(e) Write the data frame `df` to a file named `file2.txt`, so that this file contains:

```
x,y,z
2,3,2.5
4,6,5
6,9,7.5
8,12,10
```

Solution to Question 5

The following script contains all the lines needed to complete the exercise.

```r
q5 <- function(){
    # 5(a)
    df <- read.table("file1.txt", sep=",", header=T)

    # 5(b)
    x <- df$x
    y <- df$y

    # 5(c)
    z <- (x + y)/2

    # 5(d)
    df["z"] <- z

    # 5(e)
    write.csv(df, "file2.txt", quote=F, row.names=F)
}
```

These lines can be run individually in the terminal, or the script (if it is saved as a file called `q5script.R`) can be called from the terminal using:

```r
source("q5script.R")
q5()
```

Note there are alternatives for Questions 5(a) and 5(e):

```
# 5(a) alternative
df <- read.csv("file1.txt")

# 5(e) alternative
write.table(df, file="file2.txt", col.names=T, row.names=F, quote=F, sep=",")
```

However, the reason that I suggest using the `read.table` command, rather than the `read.csv` command, is that learning to use `read.table` and how to set its parameters is more useful. For example, if instead of giving the data in `file1.txt` as above, one had to read the following data contained in a file named `file3.txt`:

```
2 3
4 6
6 9
8 12
```

Then, in this case, the file is not in csv format (the data are not comma-separated) and the `read.csv` command is not useful. Rather, one should use the commands:

```
# reads in the data
df3 <- read.table("file3.txt", sep=" ")

# sets the column names to be "x" and "y"
colnames(df3) <- c("x", "y")
```

R scripts and using the source function

The solution presented above creates an R script named `q5script.R`, which contains the function `q5()`. Then, after using `source("q5script.R")`, one still needs to call the function `q5()`. The `source` function essentially makes R read all the lines in the script. But, why bother to create a function? Why not just have something such as the following in `q5script2.R`:

```
## q5script2.R
df <- read.table("file1.txt", sep=",", header=T)
df["z"] <- (df$x + df$y)/2
write.csv(df, file="file2.txt", quote=F, row.names=F)
```

and then once one calls `source("q5script2.R")`, it will run all the lines—surely this is just as good, and saves us a line calling the function `q5()`? In fact, this second option would work, but it is **bad practice**. The `source` function indeed reads in all the lines, and (a) if the line is inside a function, it checks if any of the lines doesn't make sense (e.g. there is a syntax error), or (b) any line not inside a function is executed by R.

If the last lines of `q5script.R` and `q5script2.R` both had the same syntax error

```
wrrrrritte.csv(df, file="file2.txt", quote=F, row.names=F)
```

then `source("q5script.R")` would identify the error without running the lines inside the function `q5()`, but `source("q5script2.R")` would execute the previous lines before identifying the error—if your script takes an hour to run, it isn't ideal to find out at the end of the hour that the last line contained an error!

References

[1] J. K. Blitzstein and J. Hwang. *Introduction to probability.* Chapman and Hall/CRC, 2nd edition, 2014.

[2] H. Flanders. Differentiation under the integral sign. *The American Mathematical Monthly*, 80(6):615–627, 1973.

[3] M. Taboga. *Lectures on probability theory and mathematical statistics*. CreateSpace Independent Publishing Platform, 3rd edition, 2017.