Question 1

Assume that continuous random variable X has PDF that is non-zero only on segment $[a, b]$ and strictly positive on open interval (a, b). Median of random variable X is value $m \in (a, b)$ such that $P(X < m) = P(X > m) = 1/2$.

Prove that for symmetrical PDFs median is equal to expected value.

Provide an example of PDF such that median is larger than the expected value.

Let X be random variable, f be strictly increasing function and $Y = f(X)$. What can you say about medians of X and Y?

1) First let us prove that for symmetrical PDF expected value of X is equal to the point of symmetry $X = x_0$.

Let us consider new variable $Y = X - x_0$.
PDF of Y is related to PDF of X, we are shifting the graph of $pdf_X(x)$ to the left so that the line of symmetry for $pdf_Y(y)$ matches the ordinate axis.
As the function of $pdf_Y(y)$ is symmetric with respect to the ordinate axis we can state that this function is even, which means that the distributin of Y is the same as the distribution of $-Y$, which, in turn, means that $E(Y) = E(-Y)$.
Considering further and using properties of expected value we get:

$$\begin{gathered} E(Y) = E(-Y) \\ E(Y) = E((-1) \cdot Y) \\ E(Y) = (-1) \cdot E(Y) \\ \Rightarrow E(Y) = 0 \end{gathered} \tag{1}$$

Thus, the only possible value of Y is 0.
Returning to variable X and using properties of expected value we get:

$$E(Y) = E(X - x_0) = E(X) - x_0$$

Using (1) above we arrive at:

$$\begin{gathered} E(X) - x_0 = 0 \\ E(X) = x_0 \end{gathered} \tag{2}$$

Now we can proceed to proving that $m = x_0$.
Since the PDF of X is symmetric around x_0, the probability of X being less than some m is equal to the probability of X being greater than $x_0 + \Delta$, where $\Delta = x_0 - m$. This is because the area under the PDF curve to the left of m is the same as the area under the PDF curve to the right of $x_0 + \Delta$, due to the symmetry around x_0. Mathematically, we can express this as:

$$\begin{gathered} P(X < m) = P(X > x_0 + (x_0 - m)), \text{ or} \\ P(X < m) = P(X > 2x_0 - m) \end{gathered} \tag{3}$$

Similarly, for any value m greater than x_0, the probability of X being greater than m is equal to the probability of X being less than $x_0 - \Delta$, where $\Delta = m - x_0$. That is:

$$\begin{gathered} P(X > m) = P(X < x_0 - (m - x_0)), \text{ or} \\ P(X > m) = P(X < 2x_0 - m) \end{gathered} \tag{4}$$

Now, given m be the median of X, by definition, we have:

$$P(X < m) = 1/2 \text{ and } P(X > m) = 1/2$$

Therefore, using the equations (3) and (4) above, we get:

$$P(X < m) = P(X > 2x_0 - m) = 1/2$$

and

$$P(X < 2x_0 - m) = P(X > m) = 1/2$$

$2x_0 - m$ is the value that is symmetric to m around x_0. So, from the two equations above by the symmetry of the PDF and taking into account equal probabilities and the conclusion (2) above, we have:

$$2x_0 - m = m$$
$$m = x_0 = E(X)$$

2) We can consider the following PDF of continuous random variable X:

$$pdf_X(x) = \begin{cases} x & : \quad 0 \le x \le \sqrt{2} \\ 0 & : \quad otherwise \end{cases}$$

Checking that the area under the PDF line is equal to 1:

$$\int_0^{\sqrt{2}} pdf_X(x)\, dx = \int_0^{\sqrt{2}} x\, dx = \frac{x^2}{2}\bigg|_0^{\sqrt{2}} = \frac{\left(\sqrt{2}\right)^2}{2} = 1$$

It is the same as finding the area of the rectangular isosceles triangle we have under the graph line: $\sqrt{2} \cdot \sqrt{2} \cdot \frac{1}{2} = 1$.

Then the CDF of X is:

$$F_X(x) = \begin{cases} 0 & : \quad x < 0 \\ \frac{x^2}{2} & : \quad 0 \le x \le \sqrt{2} \\ 1 & : \quad x > \sqrt{2} \end{cases}$$

To find the median we can solve the following for x :

$$\frac{x^2}{2} = \frac{1}{2},$$

which gives median m at $x = 1$.
Again, looking for a half of the triangle area equaling to 1, we just need to find a smaller similar triange, which area is $1/2$. In our case it is $1 \cdot 1 \cdot \frac{1}{2}$ at $x = 1$.

Now, let us find expected value:

$$E(X) = \int_0^{\sqrt{2}} x \cdot pdf_X(x)\, dx = \int_0^{\sqrt{2}} x^2\, dx = \frac{x^3}{3}\bigg|_0^{\sqrt{2}} = \frac{\left(\sqrt{2}\right)^3}{3} \approx 0.94$$

Thus for the considered PDF median $m = 1$ is larger than the expected value $E(X) \approx 0.94$.

3) If we take a random variable X, and apply a strictly increasing function f to it to get a new random variable Y, meaning Y is the result of applying f to the possible values of X, then the median of Y will also be the result of applying f to the median of X.

This means that if the median of X is m, then the median of Y will be $f(m)$.