Problem 1 [20 pts.]
First consider $X_1$ and $X_2$ with joint probability function $p(x_1, x_2)$.

\[
\mathbb{E}[a_1 X_1 + a_2 X_2] = \sum_{x_1} \sum_{x_2} (a_1 x_1 + a_2 x_2) \cdot p(x_1, x_2) \tag{1}
\]

\[
= a_1 \sum_{x_1} \sum_{x_2} x_1 \cdot p(x_1, x_2) + a_2 \sum_{x_1} \sum_{x_2} x_2 \cdot p(x_1, x_2) \tag{2}
\]

\[
= a_1 \sum_{x_1} x_1 \sum_{x_2} p(x_1, x_2) + a_2 \sum_{x_2} x_2 \sum_{x_1} p(x_1, x_2) \tag{3}
\]

\[
= a_1 \sum_{x_1} x_1 \cdot p(x_1) + a_2 \sum_{x_2} x_2 \cdot p(x_2) \tag{4}
\]

\[
= a_1 \mathbb{E}[X_1] + a_2 \mathbb{E}[X_2] \tag{5}
\]

Now assume

\[
\mathbb{E}\left[ \sum_{j=1}^{k} a_j X_j \right] = \sum_{j=1}^{k} a_j \mathbb{E}[X_j]
\]

holds for some $k \in \mathbb{Z}^+$, and define

\[
Y := \sum_{j=1}^{k} a_j X_j.
\]

Then

\[
\mathbb{E}\left[ \sum_{j=1}^{k+1} a_j X_j \right] = \mathbb{E}\left[ Y + a_{k+1} X_{k+1} \right]
\]

\[
= \mathbb{E}[Y] + a_{k+1} \mathbb{E}[X_{k+1}] \tag{6}
\]

\[
= \sum_{j=1}^{k+1} a_j \mathbb{E}[X_j],
\]

where (6) follows from (1)-(5). Therefore,

\[
\mathbb{E}\left[ \sum_{j=1}^{m} a_j X_j \right] = \sum_{j=1}^{m} a_j \mathbb{E}[X_j]
\]

for any $m \in \mathbb{Z}^+$, by induction.

Notice (5) also implies

\[
\mathbb{E}[a_1 X + a_2] = a_1 \cdot \mathbb{E}[X] + a_2
\]

by letting $X_2$ be a degenerate random variable with $P(X_2 = 1) = 1$. 

1
Problem 2 [20 pts.]

\[
\sum_{j=1}^{\infty} P(X \geq j) = \sum_{j=1}^{\infty} \sum_{k=j}^{\infty} P(X = k) \quad (7)
\]

\[
= \sum_{k=1}^{\infty} \sum_{j=1}^{k} P(X = k) \quad (8)
\]

\[
= \sum_{k=1}^{\infty} k \cdot P(X = k)
\]

\[
= \mathbb{E}[X]
\]

To understand what is happening with the interchange of summations in (8) it may help to write (7) as

\[
\sum_{j=1}^{\infty} \sum_{k=j}^{\infty} P(X = k) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) + \ldots
\]

\[
+ P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) + \ldots
\]

\[
+ P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) + \ldots
\]

\[
+ P(X = 4) + P(X = 5) + P(X = 6) + \ldots
\]

\[
+ P(X = 5) + P(X = 6) + \ldots
\]

\[
+ P(X = 6) + \ldots
\]

(7) sums over this “matrix” by row and (8) sums over it by column.

Alternate approach

One could also sum over all entries of the above matrix with zeros plugged into the lower triangle, i.e.

\[
\sum_{j=1}^{\infty} P(X \geq j) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} P(X = k) \cdot 1_{\{k \geq j\}}
\]

\[
= \sum_{k=1}^{\infty} \sum_{j=1}^{k} P(X = k) \cdot 1_{\{k \geq j\}}
\]

\[
= \sum_{k=1}^{\infty} k \cdot P(X = k)
\]

\[
= \mathbb{E}[X].
\]
Problem 3 [20 pts.]

(a)  
\[ \text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] \]
\[ = \int_{\mathbb{R}} (x - \mathbb{E}[X])^2 \cdot p(x) \, dx \]
\[ = (a - \mathbb{E}[X])^2 \cdot p(a) + \int_{\mathbb{R}\{a\}} (x - \mathbb{E}[X])^2 \cdot p(x) \, dx \]
\[ = (a - \mathbb{E}[X])^2 \cdot 1 \]
\[ = (a - \int_{\mathbb{R}} x \cdot p(x) \, dx)^2 \]
\[ = (a - a \cdot p(a))^2 \]
\[ = (a - a)^2 \]
\[ = 0. \]

(b) Here we assume $X$ is discrete and $\text{Var}(X) = 0$, i.e.

\[ \text{Var}(X) = \sum_{x} (x - \mathbb{E}[X])^2 \cdot p(x) \]
\[ = 0. \tag{9} \]

Since every term in (9) is nonnegative, the above implies

\[ (x - \mathbb{E}[X])^2 \cdot p(x) = 0 \]
\[ \text{for all } x. \tag{10} \]

For (10) to hold, then any time $p(x) > 0$, we must have $(x - \mathbb{E}[X])^2 = 0$.

Now assume there are two distinct values $x_1$ and $x_2$ for which $p(x_1) > 0$ and $p(x_2) > 0$.
But (10) implies

\[ x_1 = x_2 = \mathbb{E}[X], \]

a contradiction. Therefore,

\[ P(X = a) = 1, \]

where $a = \mathbb{E}[X]$. 

---

3
Problem 4 [20 pts.]

\[
\text{Cov}(a + bX, c + dY) = \mathbb{E}\left[ (a + bX - \mathbb{E}[a + bX])(c + dY - \mathbb{E}[c + dY]) \right] \\
= \mathbb{E}\left[ (a + bX - a - b \cdot \mathbb{E}[X])(c + dY - c - d \cdot \mathbb{E}[Y]) \right] \\
= \mathbb{E}\left[ b \cdot (X - \mathbb{E}[X]) \cdot d \cdot (Y - \mathbb{E}[Y]) \right] \\
= bd \cdot \mathbb{E}\left[ (X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) \right] \\
= bd \cdot \text{Cov}(X, Y),
\]

where we have used the linearity of expectation, established in Problem 1.
Problem 5 [20 pts.]

(a) 
\[ E[Y] = 5E[X] + E[\epsilon] \] (by linearity of expectation)
\[ = 5 \cdot 0 + 0 \]
\[ = 0 \]

\[ \text{Var}(Y) = E[(Y - E[Y])^2] \]
\[ = E[(5X + \epsilon)^2] \]
\[ = E[25X^2 + 10X \cdot \epsilon + \epsilon^2] \]
\[ = 25 \cdot E[X^2] + 10 \cdot E[X] \cdot E[\epsilon] + E[\epsilon^2] \]
\[ = 25 \cdot \text{Var}(X) + E[X] \cdot E[\epsilon] + 10 \cdot E[X] \cdot E[\epsilon] + E[\epsilon^2] \]
\[ = 25 \cdot (\frac{1}{3} + 0^2) + 10 \cdot 0 \cdot 0 + 1 + 0^2 \]
\[ = \frac{28}{3} \]

(b) 
\[ E[Y^2] = \text{Var}(Y) + E[Y]^2 \]
\[ = \frac{28}{3} + 0^2 \]
\[ = \frac{28}{3} \]

(c) 
\[ E[Y \mid X = x] = E[5X + \epsilon \mid X = x] \]
\[ = E[5X \mid X = x] + E[\epsilon \mid X = x] \] (by linearity of conditional expectation)
\[ = 5x + E[\epsilon \mid X = x] \]
\[ = 5x \]
\[ \text{by independence of } X \text{ and } \epsilon \]

(d) 
\[ E[Y^3] = E[(5X + \epsilon)^3] \]
\[ = E[125X^3] + E[75X^2 \cdot \epsilon] + E[15X \cdot \epsilon^2] + E[\epsilon^3] \]
\[ = 125E[X^3] + 75E[X^2 \cdot \epsilon] + 15E[X \cdot \epsilon^2] + E[\epsilon^3] \]
\[ = 125 \int_{-1}^{1} x^3 \cdot \frac{1}{2} dx + 75E[X^2 \cdot \epsilon] + 15E[X \cdot \epsilon^2] + \text{Skew}[Z] \]
\[ = 75E[X^2 \cdot \epsilon] + 15E[X \cdot \epsilon^2] \]
\[ = 75E[X^2] \cdot E[\epsilon] + 15E[X] \cdot E[\epsilon^2] \]
\[ \text{by independence; see A1} \]
\[ = 0 \]
(e)

\[ \text{Cov}(\epsilon, \epsilon^2) = \mathbb{E}[(\epsilon - \mathbb{E}[\epsilon])(\epsilon^2 - \mathbb{E}[\epsilon^2])] \]
\[ = \mathbb{E}[\epsilon \cdot (\epsilon^2 - 1)] \]
\[ = \mathbb{E}[\epsilon^3] - \mathbb{E}[\epsilon] \]
\[ = \text{Skew}[Z] \]
\[ = 0. \]

Although \( \text{Cov}(\epsilon, \epsilon^2) = 0 \), this does not imply that they are independent! For example, consider that

\[ P(\epsilon \leq 1, \epsilon^2 \leq 1) = P(\epsilon \leq 1) \]
\[ \neq P(\epsilon \leq 1) \cdot P(\epsilon^2 \leq 1). \]

Hence, \( \epsilon \) and \( \epsilon^2 \) are not independent.
Appendix

(A1) Let $X$ and $Y$ be independent random variables. Then, by definition, for any $(x, y)$

$$P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y).$$

Now consider $X^2$ and $Y$. We have

$$P(X^2 \leq x, Y \leq y) = P(-\sqrt{x} \leq X \leq \sqrt{x}, Y \leq y)$$
$$= P(X \leq \sqrt{x}, Y \leq y) - P(X \leq -\sqrt{x}, Y \leq y)$$
$$= P(X \leq \sqrt{x}) \cdot P(Y \leq y) - P(X \leq -\sqrt{x}) \cdot P(Y \leq y)$$
$$= (P(X \leq \sqrt{x}) - P(X \leq -\sqrt{x})) \cdot P(Y \leq y)$$
$$= P(X^2 \leq x) \cdot P(Y \leq y).$$