Mathematical Expectation: Examples

- Consider the following game of chance. You pay 2 dollars and roll a fair die. Then you receive a payment according to the following schedule. If the event $A = \{1, 2, 3\}$ occurs, then you will receive 1 dollar. If the event $B = \{4, 5\}$ occurs, you receive 2 dollars. If the event $C = \{6\}$ occurs, then you will receive 6 dollars. What is the average profit you can make if you participate this game?

If $A$ occurs, then a profit will be $1 - 2 = -1$ dollar, i.e., you will lose 1 dollar. If $B$ occurs, a profit will be $2 - 2 = 0$. If $C$ occurs, a profit will be $6 - 2 = 4$ dollars. Therefore, we may compute the average profit as follows:

$$\text{average profit} = (1/6+1/6+1/6)\times(-1)+(1/6+1/6)\times0+(1/6)\times4 = (1/6)\times(-3+0+4) = 1/6.$$  

That is, you can expect to make $1/6$ dollar on the average every time you play this game. This is the mathematical expectation of the payment.

We can define a random variable $X$ which represents a profit, where $X$ takes a value of $-1$, 0, and 4 with probabilities $1/2$, $1/3$, and $1/6$, respectively. Namely, $P(X = -1) = 1/2$, $P(X = 0) = 1/3$, and $P(X = 4) = 1/6$. Then this mathematical expectation is written as

$$E(X) = \sum_{x\in\{-1,0,4\}} xP(X = x) = (-1)\times(1/2) + 0 \times (1/3) + 4 \times (1/6) = 1/6.$$  

- Roll a die twice. Let $X$ be the number of times 4 comes up. $X$ takes three possible values 0, 1, or 2. $X = 0$ when the event $\{1, 2, 3, 5, 6\}$ occurs for both cases so that $P(X = 0) = (5/6)\times(5/6) = 25/36$. $X = 1$ either when the event $\{1, 2, 3, 5, 6\}$ occurs for the first die and the event $\{4\}$ occurs for the second die or when the event $\{4\}$ occurs for the first die and the event $\{1, 2, 3, 5, 6\}$ occurs for the second die so that $P(X = 1) = (5/6)\times(1/6) + (1/6)\times(5/6) = 10/36$. Finally, $X = 2$ when the event $\{4\}$ for both dies so that $P(X = 2) = (1/6)\times(1/6) = 1/36$. Note that $P(X = 0) + P(X = 1) + P(X = 2) = 1$. Therefore, the mathematical expectation of $X$ is

$$E(X) = \sum_{x=0,1,2} xP(X = x) = 0 \times (25/36) + 1 \times (10/36) + 2 \times (1/36) = 1/3.$$  

- Toss a coin 3 times. Let $X$ be the number of heads. There are 8 possible outcomes: $\{TTT, TTH, THT, THH, HTT, HTH, HHT, HHH\}$, where $H$ indicates “Head” and $T$ indicates “Tail”. $X$ takes four possible values 0, 1, 2, and 3 with probabilities $P(X = 0) = 1/8$, $P(X = 1) = 3/8$, $P(X = 2) = 3/8$, and $P(X = 3) = 1/8$. Therefore, the mathematical expectation of $X$ is

$$E(X) = \sum_{x=0,1,2,3} xP(X = x) = 0 \times (1/8) + 1 \times (3/8) + 2 \times (3/8) + 3 \times (1/8) = (0+3+6+3)/8 = 12/8 = 3/2.$$
Properties of Mathematical Expectation

Let $X$ be a random variable and suppose that the mathematical expectation of $X$, $E(X)$, exists.

1. If $a$ is a constant, then 
   \[ E(a) = a. \]

2. If $b$ is a constant, then 
   \[ E(bX) = bE(X). \]

3. If $a$ and $b$ are constants, then 
   \[ E(a + bX) = a + bE(X). \] (1)

**Proof:** Let $X$ be a discrete random variable, where possible values for $X$ is \{ $x_1, \ldots, x_n$ \} with probability mass function of $X$ given by 
\[ p_i^X = P(X = x_i), \quad i = 1, \ldots n. \]

For the proof of 1, we have 
\[ E(a) = \sum_{i=1}^{n} ap_i^X \]
\[ = (ap_1^X + ap_2^X + \ldots + ap_n^X) \]
\[ = a \times (p_1^X + p_2^X + \ldots + p_n^X) \]
\[ = a \sum_{i=1}^{n} p_i^X \]
\[ = a \]
where the last equality holds because \( \sum_{i=1}^{n} p_i^X = 1 \).

For the proof of 2, we have 
\[ E(bX) = \sum_{i=1}^{n} bx_i p_i^X \]
\[ = (bx_1 p_1^X + bx_2 p_2^X + \ldots + bx_n p_n^X) \]
\[ = b \times (x_1 p_1^X + x_2 p_2^X + \ldots + x_n p_n^X) \]
\[ = b \sum_{i=1}^{n} x_i p_i^X \]
\[ = bE(X). \]

For the proof of 3, we have 
\[ E(a + bX) = \sum_{i=1}^{n} (a + bx_i) p_i^X \]
\[ = (a + bx_1) p_1^X + (a + bx_2) p_2^X + \ldots + (a + bx_n) p_n^X \]
\[ = (ap_1^X + ap_2^X + \ldots + ap_n^X) + (bx_1 p_1^X + bx_2 p_2^X + \ldots + bx_n p_n^X) \]
\[ = a \times (p_1^X + p_2^X + \ldots + p_n^X) + b \times (x_1 p_1^X + x_2 p_2^X + \ldots + x_n p_n^X) \]
\[ = a \sum_{i=1}^{n} p_i^X + b \sum_{i=1}^{n} x_i p_i^X \]
\[ = a + bE(X). \]
Variance and Covariance

Let $X$ and $Y$ be two discrete random variables. The set of possible values for $X$ is $\{x_1, \ldots, x_n\}$; and the set of possible values for $Y$ is $\{y_1, \ldots, y_m\}$. The joint probability function is given by

$$p_{ij}^{X,Y} = P(X = x_i, Y = y_j), \quad i = 1, \ldots, n; j = 1, \ldots, m.$$ 

The marginal probability function of $X$ is

$$p_i^X = P(X = x_i) = \sum_{j=1}^{m} p_{ij}^{X,Y}, \quad i = 1, \ldots, n,$$

and the marginal probability function of $Y$ is

$$p_j^Y = P(Y = y_j) = \sum_{i=1}^{n} p_{ij}^{X,Y}, \quad j = 1, \ldots, m.$$

1. 

$$E[X + Y] = E[X] + E[Y]. \tag{2}$$

Proof:

$$E(X + Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} (x_i + y_j)p_{ij}^{X,Y}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} (x_i p_{ij}^{X,Y} + y_j p_{ij}^{X,Y})$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} x_i p_{ij}^{X,Y} + \sum_{i=1}^{n} \sum_{j=1}^{m} y_j p_{ij}^{X,Y} \tag{3}$$

$$= \sum_{i=1}^{n} \left( x_i \cdot \left( \sum_{j=1}^{m} p_{ij}^{X,Y} \right) \right) + \sum_{j=1}^{m} \left( y_j \cdot \left( \sum_{i=1}^{n} p_{ij}^{X,Y} \right) \right) \tag{4}$$

because we can take $x_i$ out of $\sum_{j=1}^{m}$ because $x_i$ does not depend on $j$’s

$$= \sum_{i=1}^{n} x_i \cdot p_i^X + \sum_{j=1}^{m} y_j \cdot p_j^Y$$

because $p_i^X = \sum_{j=1}^{m} p_{ij}^{X,Y}$ and $p_j^Y = \sum_{i=1}^{n} p_{ij}^{X,Y}$

$$= E(X) + E(Y)$$

Equation (3): To understand $\sum_{i=1}^{n} \sum_{j=1}^{m} (x_i p_{ij}^{X,Y} + y_j p_{ij}^{X,Y}) = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i p_{ij}^{X,Y} + \sum_{i=1}^{n} \sum_{j=1}^{m} y_j p_{ij}^{X,Y}$, consider the case of $n = m = 2$. Then,

$$\sum_{i=1}^{2} \sum_{j=1}^{2} (x_i p_{ij}^{X,Y} + y_j p_{ij}^{X,Y})$$

$$= (x_1 p_{11}^{X,Y} + y_1 p_{11}^{X,Y}) + (x_1 p_{12}^{X,Y} + y_2 p_{12}^{X,Y}) + (x_2 p_{21}^{X,Y} + y_1 p_{21}^{X,Y}) + (x_2 p_{22}^{X,Y} + y_2 p_{22}^{X,Y})$$

$$= (x_1 p_{11}^{X,Y} + x_1 p_{12}^{X,Y} + x_2 p_{21}^{X,Y} + x_2 p_{22}^{X,Y}) + (y_1 p_{11}^{X,Y} + y_2 p_{12}^{X,Y} + y_1 p_{21}^{X,Y} + y_2 p_{22}^{X,Y})$$

$$= \sum_{i=1}^{2} \sum_{j=1}^{2} x_i p_{ij}^{X,Y} + \sum_{i=1}^{2} \sum_{j=1}^{2} y_j p_{ij}^{X,Y}.$$
Equation (4): To understand $\sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij}p_{ij}^{X,Y} = \sum_{i=1}^{n} x_i \cdot (\sum_{j=1}^{m} p_{ij}^{X,Y})$, consider the case of $n = m = 2$. Then,

$$
\sum_{i=1}^{2} \sum_{j=1}^{2} x_{ij}p_{ij}^{X,Y} = x_{11}p_{11}^{X,Y} + x_{12}p_{12}^{X,Y} + x_{21}p_{21}^{X,Y} + x_{22}p_{22}^{X,Y}
$$

$$
= x_1(p_{11}^{X,Y} + p_{12}^{X,Y}) + x_2(p_{21}^{X,Y} + p_{22}^{X,Y})
$$

$$
= \sum_{i=1}^{2} x_i(p_{i1}^{X,Y} + p_{i2}^{X,Y})
$$

$$
= \sum_{i=1}^{2} x_i(\sum_{j=1}^{2} p_{ij}^{X,Y}).
$$

Similarly, we may show that $\sum_{i=1}^{2} \sum_{j=1}^{2} y_{ij}p_{ij}^{X,Y} = \sum_{j=1}^{2} y_j \cdot (\sum_{i=1}^{2} p_{ij}^{X,Y})$.

2. If $c$ is a constant, then $\text{Cov}(X, c) = 0$.

**Proof:** According to the definition of covariance,

$$
\text{Cov}(X, c) = E[(X - E(X))(c - E(c))].
$$

Since the expectation of a constant is itself, i.e., $E(c) = c$,

$$
\text{Cov}(X, c) = E[(X - E(X))(c - c)]
$$

$$
= E[(X - E(X)) \cdot 0]
$$

$$
= E[0]
$$

$$
= \sum_{i=1}^{n} 0 \times p_i^{X}
$$

$$
= \sum_{i=1}^{n} 0
$$

$$
= 0 + 0 + ... + 0
$$

$$
= 0
$$

3. $\text{Cov}(X, X) = \text{Var}(X)$.

**Proof:** According to the definition of covariance, we can expand $\text{Cov}(X, X)$ as follows:

$$
\text{Cov}(X, X) = E[(X - E(X))(X - E(X))]
$$

$$
= \sum_{i=1}^{n} [x_i - E(X)][x_i - E(X)] \cdot P(X = x_i), \text{ where } E(X) = \sum_{i=1}^{n} x_i p_i^{X}
$$

$$
= \sum_{i=1}^{n} [x_i - E(X)][x_i - E(X)] \cdot p_i^{X}
$$

$$
= \sum_{i=1}^{n} (x_i - E(X))^2 \cdot p_i^{X}
$$

$$
= E[(X - E(X))^2] \text{ (by def. of the expected value)}
$$

$$
= \text{Var}(X).
$$
4. \( \text{Cov}(X, Y) = \text{Cov}(Y, X). \)

**Proof:** According to the definition of covariance, we can expand \( \text{Cov}(X, Y) \) as follows:

\[
\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]
= \sum_{i=1}^{n} \sum_{j=1}^{m} [x_i - E(X)][y_j - E(Y)] \cdot p_{ij}^{XY},
\]
where \( E(X) = \sum_{i=1}^{n} x_i p_i^X \) and \( E(Y) = \sum_{j=1}^{m} y_j p_j^Y \)

\[
= \sum_{j=1}^{m} \sum_{i=1}^{n} [y_j - E(Y)][x_i - E(X)] \cdot p_{ij}^{XY}
= E[(Y - E(Y))(X - E(X))] \quad \text{(by def. of the expected value)}
= \text{Cov}(Y, X). \quad \text{(by def. of the covariance)}
\]

5. \( \text{Cov}(a_1 + b_1 X, a_2 + b_2 Y) = b_1 b_2 \text{Cov}(X, Y), \) where \( a_1, a_2, b_1, \) and \( b_2 \) are some constants.

**Proof:** Using \( E(a_1 + b_1 X) = a_1 + b_1 E(X) \) and \( E(a_2 + b_2 Y) = a_2 + b_2 E(Y) \), we can expand \( \text{Cov}(a_1 + b_1 X, a_2 + b_2 Y) \) as follows:

\[
\text{Cov}(X, Y) = E[(a_1 + b_1 X - E(a_1 + b_1 X))(a_2 + b_2 Y - E(a_2 + b_2 Y))]
= E[(a_1 + b_1 X - (a_1 + b_1 E(X)))(a_2 + b_2 Y - (a_2 + b_2 E(Y))]
= E[(a_1 - a_1 + b_1 X - b_1 E(X))(a_2 - a_2 + b_2 Y - b_2 E(Y)]
= E[(b_1 X - b_1 E(X))(b_2 Y - b_2 E(Y)]
= E[b_1 (X - E(X)) \cdot b_2 (Y - E(Y)]
= E[b_1 b_2 (X - E(X))(Y - E(Y)]
= \sum_{i=1}^{n} \sum_{j=1}^{m} b_1 b_2 (x_i - E(X))(y_j - E(Y)) \cdot p_{ij}^{XY}
= b_1 b_2 \sum_{i=1}^{n} \sum_{j=1}^{m} [x_i - E(X)][y_j - E(Y)] \cdot p_{ij}^{XY} \quad \text{(by using (1))}
= b_1 b_2 \text{Cov}(X, Y).
\]

6. If \( X \) and \( Y \) are independent, then \( \text{Cov}(X, Y) = 0. \)

**Proof:** If \( X \) and \( Y \) are independent, by definition of stochastic independence, \( P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j) = p_i^X p_j^Y \) for any \( i = 1, ..., n \) and \( j = 1, ..., m. \) Then, we may
expand \( \text{Cov}(X, Y) \) as follows.

\[
\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]
\]

\[
= \sum_{i=1}^{n} \sum_{j=1}^{m} [x_i - E(X)][y_j - E(Y)] \cdot P(X = x_i, Y = y_j)
\]

\[
= \sum_{i=1}^{n} \sum_{j=1}^{m} [x_i - E(X)][y_j - E(Y)]p_i^X p_j^Y
\]

because \( X \) and \( Y \) are independent

\[
= \sum_{i=1}^{n} \sum_{j=1}^{m} \{(x_i - E(X))p_i^X \}{(y_j - E(Y))p_j^Y}
\]

(5)

because we can move \([x_i - E(X)]p_i^X\) outside of \( \sum_{j=1}^{m} \)

\[
= \left\{ \sum_{j=1}^{m} [y_j - E(Y)]p_j^Y \right\}\left\{ \sum_{i=1}^{n} [x_i - E(X)]p_i^X \right\}
\]

(6)

because we can move \( \sum_{j=1}^{m} [y_j - E(Y)]p_j^Y \) outside of \( \sum_{i=1}^{n} \)

\[
= \left\{ \sum_{i=1}^{n} x_i p_i^X - \sum_{i=1}^{n} E(X)p_i^X \right\}\cdot\left\{ \sum_{j=1}^{m} y_j p_j^Y - \sum_{j=1}^{m} E(Y)p_j^Y \right\}
\]

by definition of \( E(X) \) and \( E(Y) \)

\[
= \left\{ E(X) - \sum_{i=1}^{n} E(X)p_i^X \right\}\cdot\left\{ E(Y) - \sum_{j=1}^{m} E(Y)p_j^Y \right\}
\]

because we can move \( E(X) \) and \( E(Y) \) outside of \( \sum_{i=1}^{n} \) and \( \sum_{j=1}^{m} \), respectively

\[
= \{E(X) - E(X) \cdot 1\} \cdot \{E(Y) - E(Y) \cdot 1\}
\]

\[
= 0 \cdot 0 = 0.
\]

Equation (6): This is similar to equation (4). Please consider the case of \( n = m = 2 \) and convince yourself that (6) holds.

7. \( \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \).

\textbf{Proof:} By the definition of variance,

\[
\text{Var}(X + Y) = E[(X + Y - E(X + Y))^2].
\]
Then,

$$Var(X + Y) = E[(X + Y - E(X + Y))^2]$$
$$= E[((X - E(X)) + (Y - E(Y)))^2]$$
$$= E[(X - E(X))^2 + (Y - E(Y))^2 + 2(X - E(X))(Y - E(Y))]$$

because for any $a$ and $b$, $(a + b)^2 = a^2 + b^2 + 2ab$
$$= E[(X - E(X))^2] + E[(Y - E(Y))^2] + 2E[(X - E(X))(Y - E(Y))]$$  \[\text{(by using (2))}\]
$$= Var(X) + Var(Y) + 2Cov(X, Y)$$

by definition of variance and covariance

8. **Var**(X - Y) = **Var**((X + Y) - 2Cov(X, Y).

**Proof:** The proof of Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y) is similar to the proof of Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y). First, we may show that E(X - Y) = E(X) - E(Y). Then,

Var(X - Y) = E[(X - Y - E(X - Y))^2]
$$= E[((X - E(X)) - (Y - E(Y)))^2]$$
$$= E[(X - E(X))^2 + (Y - E(Y))^2 - 2(X - E(X))(Y - E(Y))]$$
$$= E[(X - E(X))^2] + E[(Y - E(Y))^2] - 2E[(X - E(X))(Y - E(Y))]$$  \[\text{(by using (2))}\]
$$= Var(X) + Var(Y) - 2Cov(X, Y)$$

9. Define $W = (X - E(X))/\sqrt{Var(X)}$ and $Z = (Y - E(Y))/\sqrt{Var(Y)}$. Show that Cov(W, Z) = Corr(X, Z).

**Proof:** Expanding Cov(W, Z), we have

$$Cov(W, Z) = E[(W - E(W))(Z - E(Z))]$$
$$= E[WZ] \text{ (because } E[W] = E[Z] = 0)$$
$$= E \left\{ \frac{X - E(X)}{\sqrt{Var(X)}} \cdot \frac{Y - E(Y)}{\sqrt{Var(Y)}} \right\}$$

by definition of $W$ and $Z$

$$= E \left\{ \frac{1}{\sqrt{Var(X)}} \cdot \frac{1}{\sqrt{Var(Y)}} \cdot [X - E(X)]E[Y - E(Y)] \right\}$$

$$= \frac{1}{\sqrt{Var(X)}} \cdot \frac{1}{\sqrt{Var(Y)}} \cdot E \left\{ [X - E(X)]E[Y - E(Y)] \right\}$$  \[\text{(by using (1))}\]

because both $\frac{1}{\sqrt{Var(X)}}$ and $\frac{1}{\sqrt{Var(Y)}}$ are constant

$$= \frac{Cov(X, Y)}{\sqrt{Var(X)} \sqrt{Var(Y)}}$$  \[\text{(by definition of covariance)}\]

$$= Corr(X, Y) \quad \text{(by definition of correlation coefficient)}$$
10. Let $b$ be a constant. Show that $E[(X - b)^2] = E(X^2) - 2bE(X) + b^2$. What is the value of $b$ that gives the minimum value of $E[(X - b)^2]$?

**Answer:** Because $(X - b)^2 = X^2 - 2bX + b^2$, we have


Noting that $E[X^2] - 2bE(X) + b^2$ is a quadratic convex function of $b$, we may find the minimum by differentiating $E[(X - b)^2]$ with respect to $b$ and set $\frac{\partial}{\partial b}E[(X - b)^2] = 0$, i.e.,

$$\frac{\partial}{\partial b}E[(X - b)^2] = -2E(X) + 2b = 0,$$

and, therefore, setting the value of $b$ equal to

$$b = E(X)$$

minimizes $E[(X - b)^2]$.

11. Let $\{x_i : i = 1, \ldots, n\}$ and $\{y_i : i = 1, \ldots, n\}$ be two sequences. Define the averages

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i,$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i.$$

(a) $\sum_{i=1}^{n} (x_i - \bar{x}) = 0$.

**Proof:**

$$\sum_{i=1}^{n} (x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x}$$

$$= \sum_{i=1}^{n} x_i - n\bar{x}$$

because $\sum_{i=1}^{n} \bar{x} = \bar{x} + \bar{x} + \ldots + \bar{x} = n\bar{x}$

$$= n \frac{1}{n} \sum_{i=1}^{n} x_i - n\bar{x}$$

because $\sum_{i=1}^{n} x_i = \frac{n}{n} \sum_{i=1}^{n} x_i = n\frac{\sum_{i=1}^{n} x_i}{n}$

$$= n\bar{x} - n\bar{x}$$

because $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

$$= 0.$$

(b) $\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i (x_i - \bar{x})$. 

8
Proof: We use the result of 2.(a) above.

\[ \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} (x_i - \bar{x}) (x_i - \bar{x}) \]

\[ = \sum_{i=1}^{n} x_i (x_i - \bar{x}) - \sum_{i=1}^{n} \bar{x} (x_i - \bar{x}) \]

\[ = \sum_{i=1}^{n} x_i (x_i - \bar{x}) - \bar{x} \sum_{i=1}^{n} (x_i - \bar{x}) \]

because \( \bar{x} \) is constant and does not depend on \( i \)'s

\[ = \sum_{i=1}^{n} x_i (x_i - \bar{x}) \]

because \( \sum_{i=1}^{n} (x_i - \bar{x}) = 0 \). as shown above

\[ = \sum_{i=1}^{n} x_i (x_i - \bar{x}) . \]

(c) \( \sum_{i=1}^{n} (x_i - \bar{x}) (y_i - \bar{y}) = \sum_{i=1}^{n} y_i (x_i - \bar{x}) = \sum_{i=1}^{n} x_i (y_i - \bar{y}) . \)

Proof: The proof is similar to the proof of 2.(b) above.

\[ \sum_{i=1}^{n} (x_i - \bar{x}) (y_i - \bar{y}) = \sum_{i=1}^{n} (x_i - \bar{x}) y_i - \sum_{i=1}^{n} (x_i - \bar{x}) \bar{y} \]

\[ = \sum_{i=1}^{n} (x_i - \bar{x}) y_i - \bar{y} \sum_{i=1}^{n} (x_i - \bar{x}) \]

\[ = \sum_{i=1}^{n} (x_i - \bar{x}) y_i - \bar{y} \cdot 0 \]

\[ = \sum_{i=1}^{n} y_i (x_i - \bar{x}) . \]

Also,

\[ \sum_{i=1}^{n} (x_i - \bar{x}) (y_i - \bar{y}) = \sum_{i=1}^{n} x_i (y_i - \bar{y}) - \sum_{i=1}^{n} \bar{x} (y_i - \bar{y}) \]

\[ = \sum_{i=1}^{n} x_i (y_i - \bar{y}) - \bar{x} \sum_{i=1}^{n} (y_i - \bar{y}) \]

\[ = \sum_{i=1}^{n} x_i (y_i - \bar{y}) - \bar{x} \cdot 0 \]

\[ = \sum_{i=1}^{n} x_i (y_i - \bar{y}) . \]

Conditional Mean and Conditional Variance

Let \( X \) and \( Y \) be two discrete random variables. The set of possible values for \( X \) is \( \{x_1, \ldots, x_n\} \); and the set of possible values for \( Y \) is \( \{y_1, \ldots, y_m\} \). We may define the conditional probability
function of $Y$ given $X$ as

$$p_{ij}^{Y|X} = P(Y = y_j|X = x_i) = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)} = \frac{p_{ij}^{X,Y}}{p_{i}^{X}},$$

where $p_{ij}^{X,Y} = P(X = x_i, Y = y_j)$ and $p_{i}^{X} = P(X = x_i)$.

The conditional mean of $Y$ given $X = x_i$ is given by

$$E_Y[Y|X = x_i] = \sum_{j=1}^{m} y_j P(Y = y_j|X = x_i) = \sum_{j=1}^{m} y_j p_{ij}^{Y|X},$$

where the symbol $E_Y$ indicates that the expectation is taken treating $Y$ as a random variable. The conditional variance of $Y$ given $X = x_i$ is given by

$$\text{Var}(Y|X = x_i) = E[(Y - E[Y|X = x_i])^2] = \sum_{j=1}^{m} (y_j - E[Y|X = x_i])^2 p_{ij}^{Y|X}.$$

The conditional mean of $Y$ given $X$ can be written as $E_Y[Y|X]$ without specifying a value of $X$. Then, $E_Y[Y|X]$ is a random variable because the value of $E_Y[Y|X]$ depends on a realization of $X$. The following shows that the unconditional mean of $Y$ is equal to the expected value of $E_Y[Y|X]$ where the expectation is taken with respect to $X$.


**Proof:** Because $E_Y[Y|X = x_i] = \sum_{j=1}^{m} y_j p_{ij}^{Y|X}$, we have

$$E_X[E_Y[Y|X]] = \sum_{i=1}^{n} E_Y[Y|X = x_i] p_{i}^{X}$$

$$= \sum_{i=1}^{n} \left( \sum_{j=1}^{m} y_j p_{ij}^{Y|X} \right) p_{i}^{X}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} y_j p_{ij}^{Y|X} p_{i}^{X}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} y_j p_{ij}^{X,Y}$$

$$= \sum_{j=1}^{m} y_j \sum_{i=1}^{n} p_{ij}^{X,Y}$$

$$= \sum_{j=1}^{m} y_j p_{j}^{Y} = E_Y[Y].$$

2. Let $g(Y)$ be some known function of $Y$. Show that $E_Y[g(Y)] = E_X[E_Y[g(Y)|X]]$. 

10
Proof:

\[ E_X[E_Y[g(Y)|X]] = \sum_{i=1}^{n} E_Y[g(Y)|X = x_i] p_i^X \]

\[ = \sum_{i=1}^{n} \left( \sum_{j=1}^{m} g(y_j) p_{ij}^{Y|X} \right) p_i^X \]

\[ = \sum_{i=1}^{n} \sum_{j=1}^{m} g(y_j) \frac{p_{ij}^{X,Y}}{p_i^X} p_i^X \]

\[ = \sum_{i=1}^{n} \sum_{j=1}^{m} g(y_j) p_{ij}^{X,Y} \]

\[ = \sum_{j=1}^{m} g(y_j) \sum_{i=1}^{n} p_{ij}^{X,Y} \]

\[ = \sum_{j=1}^{m} g(y_j) p_j^Y = E_Y[g(Y)]. \]

3. Let \( g(Y) \) and \( h(X) \) be some known functions of \( Y \) and \( X \), respectively. Show that \( E[g(Y)h(X)] = E_X[h(X)E_Y[g(Y)|X]] \).

Proof:

\[ E_X[h(X)E_Y[g(Y)|X]] = \sum_{i=1}^{n} h(x_i) E_Y[g(Y)|X = x_i] p_i^X \]

\[ = \sum_{i=1}^{n} h(x_i) \left( \sum_{j=1}^{m} g(y_j) p_{ij}^{Y|X} \right) p_i^X \]

\[ = \sum_{i=1}^{n} h(x_i) \sum_{j=1}^{m} g(y_j) \frac{p_{ij}^{X,Y}}{p_i^X} p_i^X \]

\[ = \sum_{i=1}^{n} \sum_{j=1}^{m} g(y_j) h(x_i) p_{ij}^{X,Y} \]

\[ = \sum_{i=1}^{n} \sum_{j=1}^{m} g(y_j) h(x_i) p_{ij}^{X,Y} \]

\[ = E[g(Y)h(X)] \]

4. Show that, if \( E[Y|X] = E_Y[Y] \), then \( \text{Cov}(X, Y) = 0 \).

Proof:

\[ \text{Cov}(X, Y) = E[(X - E_X(X))(Y - E_Y(Y))] \quad \text{(by definition of Covariance)} \]

\[ = E_X\{[E_Y|X][(X - E_X(X))(Y - E_Y(Y))]|X\} \quad \text{(by Law of Iterated Expectation)} \]

\[ = E_X\{(X - E_X(X))E_Y|X[Y - E_Y(Y)|X]\} \quad \text{(X is “known” once conditioned on X)} \]

\[ = E_X\{(X - E_X(X))[E_Y|X(Y|X) - E_Y(Y)]\} \quad \text{(}E_Y(Y)\text{ is a constant)} \]

\[ = E_X\{(X - E_X(X))[E_Y(Y) - E_Y(Y)]\} \quad \text{\(E[Y|X] = E_Y[Y]\)} \]

\[ = E_X[(X - E_X(X)) \times 0] = 0 \]

Alternative Proof (Please compare this proof with the above proof): Let \( E_X(X) = \)
\( \frac{1}{n} x_i p_i^X \) and \( E_Y(Y) = \frac{1}{m} y_j p_j^Y \). Define \( p_{ji}^{Y|X} = \Pr(Y = y_j | X = x_i) \).

\[
\text{Cov}(X, Y) = E_{(X,Y)}[(X - E_X(X))(Y - E_Y(Y))] \quad \text{(by definition of Covariance)}
\]

\[
= \frac{1}{n} \frac{1}{m} \sum_{i=1}^{n} \sum_{j=1}^{m} (x_i - E_X(X))(y_j - E_Y(Y)) p_{ij}^{X,Y}
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{1}{m} \sum_{j=1}^{m} (x_i - E_X(X))(y_j - E_Y(Y)) \frac{p_{ij}^{X,Y}}{p_i^X} \right\} p_i^X \quad \text{(by Law of Iterated Expectation)}
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \left\{ (x_i - E_X(X)) \left\{ \frac{1}{m} \sum_{j=1}^{m} y_j p_{ji}^{Y|X} - E_Y(Y) \frac{1}{m} \sum_{j=1}^{m} p_{ji}^{Y|X} \right\} \right\} p_i^X
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \left\{ (x_i - E_X(X))[E_Y(Y) - E_Y(Y)] \right\} p_i^X \quad \text{\( \text{E}[Y|X] = E_Y[Y] \)}
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \left\{ (x_i - E_X(X)) \times 0 \right\} p_i^X = 0
\]