Moments and Conditional Expectations

Chi-Chao Hung *

February 19, 2025

Contents

1 Expectation and Moments **1**
1.1 Expectation 1
1.2 Moments 2
1.3 Variance 2

2 Multivariate Moments **3**
2.1 Covariance 3
2.2 Useful Properties 4

3 Conditional Expectations **5**
3.1 Law of Iterated Expectation 6
3.2 Conditional Variance 8

1 Expectation and Moments

1.1 Expectation

The *expectation* (or expected value) of a random variable X, denoted by $\mathrm{E}[X]$, is a measure of the central tendency of its probability distribution.

Definition 1.1 (Expectation). Let $f_X(x)$ be the pmf or pdf of X.

For a discrete random variable, the expectation is defined as:

$$\mathrm{E}[X] = \sum_x x \cdot P(X = x).$$

For a continuous random variable, the expectation is given by:

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx$$

*This is the TA note for reviewing materials in the previous semester. Some of the proofs are omitted. Please refer to textbooks for more details. All errors are mine. If you find any errors or have any suggestions, please contact me via email at r13323021@ntu.edu.tw.

Remark 1. The subscript X in E_X emphasizes that the expectation is calculated by integrating with respect to the probability density function $f_X(x)$. We omit the subscript as long as the context is clear.

Remark 2. For some random variables the expectation may not exist. In this course, we use the following notation to denote the existence of the expectation of X.[1]

$$\mathrm{E}[|X|] < \infty.$$

1.2 Moments

Definition 1.2 (Moment). The kth moment of a random variable X is defined as:

$$\mathrm{E}[X^k].$$

Definition 1.3 (Central Moment). The kth **central moment** is the expected value of $(X - \mathrm{E}[X])^k$:

$$\mathrm{E}[(X - \mathrm{E}[X])^k].$$

Moments are used to describe the shape of a distribution.

1.3 Variance

Variance measures the spread or dispersion of a random variable around its expectation.

Definition 1.4 (Variance).

$$\mathrm{Var}(X) = \mathrm{E}[(X - \mathrm{E}[X])^2].$$

The variance of a random variable X can also be expressed as:

$$\mathrm{Var}(X) = \mathrm{E}[X^2] - \big(\mathrm{E}[X]\big)^2.$$

The **standard deviation** of a random variable X, denoted by σ_X, is the square root of its variance:

$$\sigma_X = \sqrt{\mathrm{Var}(X)}.$$

Exercise 1. Let X be a random variable following an uniform distribution, $U[0, \theta]$. Find $Var(X)$.

[1] This notation might not make sense at first glance. It means that the pdf $f_X(x)$ is Lebesgue integrable, which is out of the scope of this course.

Answer 1. Since $X \sim U[0, \theta]$, its probability density function is

$$f_X(x) = \frac{1}{\theta} \quad \text{for } 0 \leq x \leq \theta.$$

First, we compute the expected value:

$$\mathrm{E}[X] = \int_0^\theta x \cdot \frac{1}{\theta}\, dx = \frac{1}{\theta} \cdot \frac{\theta^2}{2} = \frac{\theta}{2}.$$

Next, we compute the second moment:

$$\mathrm{E}[X^2] = \int_0^\theta x^2 \cdot \frac{1}{\theta}\, dx = \frac{1}{\theta} \cdot \frac{\theta^3}{3} = \frac{\theta^2}{3}.$$

Thus, the variance is given by

$$\mathrm{Var}(X) = \mathrm{E}[X^2] - \big(\mathrm{E}[X]\big)^2 = \frac{\theta^2}{3} - \left(\frac{\theta}{2}\right)^2 = \frac{\theta^2}{3} - \frac{\theta^2}{4} = \frac{\theta^2}{12}.$$

2 Multivariate Moments

For two random variables X and Y, the expectation of their product is given by:

$$\mathrm{E}[XY] = \int_y \int_x xy \cdot f_{XY}(x, y)\, dx\, dy,$$

where $f_{XY}(x, y)$ is the joint probability density function of X and Y.

Theorem 2.1. If X and Y are independent, then the following property holds:

$$\mathrm{E}[XY] = \mathrm{E}[X]\ \mathrm{E}[Y].$$

See Theorem 4.3 in Hansen (2022b) for the proof.

2.1 Covariance

Definition 2.1. Covariance measures the linear relationship between two random variables X and Y. It is defined as:

$$\mathrm{Cov}(X, Y) = \mathrm{E}_{XY}\left[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])\right].$$

It could be shown that:

$$\mathrm{Cov}(X, Y) = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y].$$

Definition 2.2. The correlation coefficient is defined as

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}.$$

Sometimes we use the notation

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

2.2 Useful Properties

Property 1.

$$\mathrm{E}[aX + bY] = a\,\mathrm{E}[X] + b\,\mathrm{E}[Y].$$

Expectation is a linear operator.

Property 2.

$$\mathrm{Var}(aX + bY) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X, Y).$$

Hint: Think of $(a + b)^2 = a^2 + b^2 + 2ab$

Property 3.

$$\mathrm{Cov}(aX + b,\, cY + d) = ac\,\mathrm{Cov}(X, Y).$$

Adding constant to random variables does not change their variances or covariance.

Property 4.

$$\begin{aligned}&\mathrm{Cov}(aX + bY,\, cZ + dW)\\ &= ac\,\mathrm{Cov}(X, Z) + ad\,\mathrm{Cov}(X, W) + bc\,\mathrm{Cov}(Y, Z) + bd\,\mathrm{Cov}(Y, W).\end{aligned}$$

Hint: Think of $(a + b)(c + d) = ac + ad + bc + bd$

Exercise 2. X and Y follows a joint normal distribution,

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix},\right)$$

Find the moment $\mathrm{E}[XY]$.

Answer 2. By definition,

$$\mathrm{E}[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy \cdot f_{XY}(x, y)\,dx\,dy,$$

where $f_{XY}(x, y)$ is the density function of the bivariate normal distibution.

However, recall that

$$\mathrm{Cov}(X, Y) = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y].$$

In this case, we have

$$\mathrm{E}[X] = 0, \mathrm{E}[Y] = 1, \mathrm{Var}(X) = 1, \mathrm{Var}(Y) = 2, \text{ and } \mathrm{Cov}(X, Y) = 1.$$

So $\mathrm{E}[XY] = 1 + (0)(1) = 1$.

3 Conditional Expectations

Definition 3.1. The conditional expectation of Y given $X = x$, denoted by $\mathrm{E}[Y \mid X = x]$, is the expected value of Y when $X = x$.

$$\mathrm{E}[Y \mid X = x] = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y \mid x)\, dy,$$

where $f_{Y|X}(y \mid x)$ is the conditional density of Y given X.

Example 3.1. The following table presents the joint probability mass function (PMF) of two random variables, Age (Y) and Gender (X):

Table 1: Distribution of Age and Gender

	Gender (X)	
Age (Y)	Male (0)	Female (1)
18	0.2	0.2
19	0.1	0.2
20	0.1	0.2

To find the CEF of Y given X, we compute the conditional probabilities:

Case 1: $X = 0$

$$P(Y = 18 \mid X = 0) = \frac{0.2}{0.4} = 0.5, \quad P(Y = 19 \mid X = 0) = \frac{0.1}{0.4} = 0.25,$$
$$P(Y = 20 \mid X = 0) = \frac{0.1}{0.4} = 0.25.$$

Case 2: $X = 1$

$$P(Y = 18 \mid X = 1) = \frac{0.2}{0.6} = \frac{1}{3}, P(Y = 19 \mid X = 1) = \frac{0.2}{0.6} = \frac{1}{3},$$
$$P(Y = 20 \mid X = 1) = \frac{0.2}{0.6} = \frac{1}{3}.$$

The conditonal expectation function of Y is

$$\mathrm{E}[Y \mid X = x] = \begin{cases} 0.5 \times 18 + 0.25 \times 19 + 0.25 \times 20 = 18.75 & \text{if } x = 0 \\ 0.33 \times 18 + 0.33 \times 19 + 0.33 \times 20 = 19 & \text{if } x = 1 \end{cases}$$

Remark 3. $\mathrm{E}[Y \mid X = x]$ is a function of x, known as the **conditional expectation function** (CEF). We often use the CEF as a **predictor** of Y. For any specific value of x, the prediction $\mathrm{E}[Y \mid X = x]$ is just a constant (e.g., 18.75 or 19).

Substituting the random X into the CEF, we write $\mathrm{E}[Y \mid X]$. This quantity is also random because it depends on X. In example 3.1, $\mathrm{E}[Y \mid X]$ is a random variable that takes the value 18.75 with probability 0.4 and 19 with probability 0.6. Understanding the properties of this predictor is a key focus of this course.

3.1 Law of Iterated Expectation

Theorem 3.1 (Simple Law of Iterated Expectation). The law of iterated expectation states that given $\mathrm{E}[|Y|] < \infty$,

$$\mathrm{E}_Y[Y] = \mathrm{E}_X[\mathrm{E}_{Y|X}[Y \mid X]].$$

Remark 4. This property shows that the overall expectation of Y can be obtained by

- Step 1: Taking the conditional mean given X.
- Step 2: Averaging it over the distribution of X.

Example 3.2. Following Example 3.1, the mean wage $\mathrm{E}[Y]$ can be calculated using the conditional means $\mathrm{E}[Y \mid X = 0]$ and $\mathrm{E}[Y \mid X = 1]$.

$$\begin{aligned} &\mathrm{E}[Y] \\ &= \Pr(X = 0)\ \mathrm{E}[Y \mid X = 0] + \Pr(X = 1)\ \mathrm{E}[Y \mid X = 1] \\ &= 0.4 \times 18.75 + 0.6 \times 19 = 18.9. \end{aligned}$$

The law can be generalized to two or more random variables:

Theorem 3.2 (Law of Iterated Expectation).

$$\mathrm{E}_{Y|X}[Y \mid X] = \mathrm{E}_{Y,Z|X}\Big[\mathrm{E}_{Y|X,Z}[Y \mid X, Z] \mid X\Big].$$

This result is also phrased as *"The smaller information set wins."* [2]

Table 2: Mean Monthly Wage(1000 NTD)

	North	South
Male(0)	25	22
Female(1)	30	26

Exercise 3. The table presents the mean of wages($\mathrm{E}[Y \mid X, Z]$) conditional on Gender (X) and Region of residence (Z):

There are only two regions in the country, North and South. We aslo know that conditional on being male the probability of living in the North is 0.6, while conditional on being female the probability of living in the North is 0.5. Find $\mathrm{E}\left[\mathrm{E}\left[Y \mid X, Z\right] \mid X = 0\right]$.

Answer 3.

$$
\begin{aligned}
&\mathrm{E}\left[\mathrm{E}\left[Y \mid X, Z\right] \mid X = 0\right] \\
&= Pr(Z = N \mid X = 0) \times \mathrm{E}\left[Y \mid 0, N\right] \\
&+ Pr(Z = S \mid X = 0) \times \mathrm{E}\left[Y \mid 0, S\right] \\
&= 0.6 \times 25 + 0.4 \times 22 = 23.8
\end{aligned}
$$

Which is exactly $\mathrm{E}\left[Y \mid X = 0\right]$.

Another useful result is the Conditioning Theorem.[3]

Theorem 3.3 (Conditioning Theorem). Given $E|Y| < \infty$,

$$E[g(X)Y|X] = g(X)E[Y|X].$$

Remark 5. The conditioning theorem said that once we condtion on X, any function of X could be seen as *nonrandom* and be taken out of the expectation.

Exercise 4. Let X and Y be two random variables. $X \sim Ber(p)$. Calculate $E[XY]$. (Intuitively, this is calculating the mean of Y while treating Y of those with $X = 0$ as 0.)

[2] See Theorem 2.2 in Hansen (2022a) for the proof.
[3] See Theorem 2.3 in Hansen (2022a) for the proof.

Answer 4.

$$\begin{aligned}\mathrm{E}[XY] &= \mathrm{E}[\mathrm{E}[XY \mid X]] = \mathrm{E}[X\,\mathrm{E}[Y \mid X]] \\ &= (1-p) * 0 * \mathrm{E}[Y|X=0] + p * 1 * \mathrm{E}[Y \mid X=1] \\ &= p\,\mathrm{E}[Y|X=1].\end{aligned}$$

It is the conditional mean scaled by the probability of X=1.

3.2 Conditional Variance

Definition 3.2.

$$\mathrm{Var}(Y \mid X=x) = \mathrm{E}\left[(Y - \mathrm{E}[Y \mid X])^2 \mid X=x\right]$$

Again,

$$\mathrm{Var}(Y \mid X) = \mathrm{E}(Y^2 \mid X) - \mathrm{E}(Y \mid X)^2$$

Theorem 3.4 (Law of Total Varaince)**.** The variance decomposition formula expresses the total variance of a random variable X in terms of its conditional variance and the variance of its conditional expectation:

$$\mathrm{Var}(Y) \;=\; \underbrace{\mathrm{Var}\big(\mathrm{E}[Y \mid X]\big)}_{\text{Explained by } X} \;+\; \underbrace{\mathrm{E}\big[\mathrm{Var}(Y \mid X)\big]}_{\text{Unexplained}}$$

This decomposition is useful for understanding the sources of variability in X.

Proof. Let us starts from decomposing Y. Let

$$e \;:=\; Y \;-\; \mathrm{E}[Y \mid X] \quad \text{and} \quad h(X) \;:=\; \mathrm{E}[Y \mid X].$$

so $Y = h(X) + e$.

Here we list some useful lemmas.

1. $\mathrm{E}[\,e \mid X\,] = 0$.
2. $\mathrm{E}[\,e\,] = 0$.
3. $\mathrm{E}[\,e\,h(X)\,] = \mathrm{E}\big[h(X)\;\mathrm{E}[\,e \mid X\,]\big] = 0$.

Take the variance of Y. The total variance can be decompose into three parts.

$$\mathrm{Var}(Y) = \mathrm{Var}\big(h(X)\big) \;+\; \mathrm{Var}(e) \;+\; 2\;\mathrm{Cov}\big(h(X),\,e\big).$$

Using the lemmas:

$$\mathrm{Cov}\big(h(X), e\big) = \mathrm{E}\big[h(X)\,e\big] - \mathrm{E}[\,h(X)\,]\;\mathrm{E}[\,e\,] = 0 - (\mathrm{E}[\,h(X)\,]\cdot 0) = 0.$$

Hence the third part vanishes.

$$\mathrm{Var}(Y) = \mathrm{Var}\big(h(X)\big) + \mathrm{Var}(e).$$

$\mathrm{Var}\big(h(X)\big)$ is the same as $\mathrm{Var}\big(\mathrm{E}[Y \mid X]\big)$, which represents the *variation in* Y *explained by* X. Next, observe that

$$\mathrm{Var}(e) = \mathrm{E}[\,e^2\,] = \mathrm{E}\big[\mathrm{E}(e^2 \mid X)\big] = \mathrm{E}\big[E\left[(Y - E[Y \mid X])^2 \mid X\right]\big] = \mathrm{E}\big[\mathrm{Var}(Y \mid X)\big],$$

the *unexplained variation* of Y once X is known.

Putting it all together,

$$\mathrm{Var}(Y) = \mathrm{Var}\big(\mathrm{E}[Y \mid X]\big) + \mathrm{E}\big[\mathrm{Var}(Y \mid X)\big].$$

□

References

Hansen, Bruce E., *Econometrics*, Princeton: Princeton University Press, 2022.

_ , *Probability & Statistics for Economists*, Princeton Oxford: Princeton University Press, 2022.