Ch4 - Probability Theory - 4

(Joint, Marginal and Conditional Distributions)

Statistics For Business and Economics - I

Shaikh Tanvir Hossain

East West University, Dhaka

Last Updated August 21, 2025

Outline

Outline

- In this chapter we will see a short overview of concepts related to joint distribution. The idea of joint probability distribution is just an application of the idea of joint probability table. Then automatically you can think about marginal probabilities and conditional probabilities. Almost all ideas are conceptually same as the ideas we saw in Chapter-2, but now this will be for multiple random variables.
- So let's start... 🚶 🚶 🚶 .

1. **Joint, Marginal and Conditional Distributions**

2. **Covariance and Correlation**

3. **Independence of random variables**

Joint, Marginal and Conditional Distributions

- We already know objects like probability distriution, expectation and variance. So far we have seen only for a single variable cases, both for discrete and continuous random variables. We will now see how to extend these concepts for multiple variables, and how to use them to understand the relationship between two random variables. Important concepts are
  - *Joint Distribution*,
  - *Covariance* and *Correlation*.
  - *Marginal distribution* (related to Marginal Expectation and Marginal Variance)
  - *Conditional distribution* (related to Conditional Expectation and Conditional Variance)

- **Recap of Expectation and Variance Formulas**
- Recall for discrete random variable X with probability mass function $f_X(x)$, we have

$$\mathbb{E}(X) = \sum_x x \cdot f_X(x)$$

Suppose if we have X with following probability distribution,

Value of X	**Probability $f_X(x)$**
1	0.2
2	0.2
3	0.6

Then we can calculate expectation as follows,

$$\mathbb{E}(X) = 1 \cdot 0.2 + 2 \cdot 0.2 + 3 \cdot 0.6 = 2.4$$

- *Ques:* What is the intuition behind the Expectation formula? *Ans:* It gives you population mean without using the population.

- And for variance we have two formulas, the definition is

$$\mathbb{V}(X) = \mathbb{E}[X - \mathbb{E}(X)]^2$$

- We can directly apply this definition, and get

$$\mathbb{V}(X) = (1-2.4)^2 \cdot 0.2 + (2-2.4)^2 \cdot 0.2 + (3-2.4)^2 \cdot 0.6 = 0.64$$

- However there is a shortcut formula for variance (can you derive this?), which is

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

  where we can calculate $\mathbb{E}(X^2)$ as follows,

$$\mathbb{E}(X^2) = 1^2 \cdot 0.2 + 2^2 \cdot 0.2 + 3^2 \cdot 0.6 = 6.4$$

  Then we can calculate variance as follows, both will give you same result,

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 6.4 - (2.4)^2 = 0.64$$

- What is the intuition behind the Variance formula? *Ans:* It gives you population variance without using the population.

Joint, Marginal and Conditional Distributions

- Now we can start ....

Joint, Marginal and Conditional Distributions

- Suppose we have following data of 150 students at East West University (EWU) regarding their family income categories and whether they tried to go to abroad for higher studies or not. *For now assume this is the population data, so we have all 150 students in the population*

	Family Income Categories (X)				
	Difficult	Middle	Higher Middle	Rich	Total
Tried	18	13	22	24	77
Not Tried	22	25	16	10	73
Total	40	38	38	34	150

- From here we can easily calculate the joint probability table,

	Family Income Categories (X)				
	Difficult	Middle	Higher Middle	Rich	Total
Tried	0.12	0.08	0.15	0.16	0.51
Not Tried	0.15	0.17	0.10	0.07	0.49
Total	0.27	0.25	0.25	0.23	1

Joint, Marginal and Conditional Distributions

- Here we can X represents Family Income Categories, 1 for Difficult, 2 for Middle, 3 for Higher Middle and 4 for Rich and Y represents tried or not, 1 means the student tried 0 means the student didn't try
- Now we can write following table which is actually called the joint probability distribution of random variables X and Y,

	Family Income Categories (X)				
Tried/Not Tried (Y)	1	2	3	4	Total
1	0.12	0.08	0.15	0.16	0.51
0	0.15	0.17	0.10	0.07	0.49
Total	0.27	0.25	0.25	0.23	1

Joint, Marginal and Conditional Distributions

- From joint probability distribution we can derive different type of probabilities and probability distributions, also Expectation and Variance.
  - **Joint Probability** $\mathbb{P}(X = x, Y = y)$ **or Joint PMF** $f(x, y)$ **:**

    For example $\mathbb{P}(X = 1, Y = 0) = 0.15$ means if we randomly select a student from the *population of* 150, then there is a 15% chance that he/she is from Difficult income category and she didn't try to go abroad for higher studies. And all the joint probabilities will sum to 1, i.e. $\sum_x \sum_y \mathbb{P}(X = x, Y = y) = 1$ and the 8 joint probabilities together is called *joint probability distribution* of X and Y. We will often use $f(x, y)$ to denote the joint probability distribution.

	$f(x, y)$			
	$x = 1$	$x = 2$	$x = 3$	$x = 4$
$y = 1$	0.12	0.08	0.15	0.16
$y = 0$	0.15	0.17	0.10	0.07

Joint, Marginal and Conditional Distributions

- **Marginal Probability $\mathbb{P}(X = x)$ or Marginal PMF $f_X(x)$:**

  This is the probability of X taking a specific value, regardless of the value of Y. For example, $\mathbb{P}(X = 1) = 0.27$ means if we randomly select a student from the *population of* 150, then there is a 27% chance that he/she is from Difficult income category. Similarly, we can find $\mathbb{P}(X = 2) = 0.25$, $\mathbb{P}(X = 3) = 0.25$, and $\mathbb{P}(X = 4) = 0.23$. From here we can calculate the *marginal probability distribution of X* as follows.

  $$\mathbb{P}(X = 1) = 0.27, \quad \mathbb{P}(X = 2) = 0.25, \quad \mathbb{P}(X = 3) = 0.25, \quad \mathbb{P}(X = 4) = 0.23$$

  We will use $f_X(x)$ to denote the *marginal probability distribution of X*,

Departments (x)	**Probability $f_X(x)$**
1	0.27
2	0.25
3	0.25
4	0.23

- And using the marginal probability distribution, we can calculate Marginal Expectation $\mathbb{E}(X)$ and Marginal Variance $\mathbb{V}(X)$ (please do it as an exercise).

Joint, Marginal and Conditional Distributions

- **Marginal Probability $\mathbb{P}(Y = y)$ or Marginal PMF $f_Y(y)$:**

  This is the probability of Y taking a specific value, regardless of the value of X. For example, $\mathbb{P}(Y = 1) = 0.51$ means if we randomly select a student from the *population of* 150, then there is a 51% chance that he/she tried to go abroad for higher studies. Similarly, we can find $\mathbb{P}(Y = 0) = 0.49$. From here we can calculate the *marginal probability distribution of* Y as follows,

  $$\mathbb{P}(Y = 1) = 0.51, \quad \mathbb{P}(Y = 0) = 0.49$$

  We will use $f_Y(y)$ to denote the *marginal probability distribution of* Y,

Tried/Not Tried (y)	**Probability $f_Y(y)$**
1	0.51
0	0.49

- And using the marginal probability distribution, we can calculate Marginal Expectation $\mathbb{E}(Y)$ and Marginal Variance $\mathbb{V}(Y)$ (please do it as an exercise).

Joint, Marginal and Conditional Distributions

- **Conditional Probability $\mathbb{P}(Y = y \mid X = x)$ or conditional PMF $f_{Y|X}(y)$:**

  This is something new, this is the probability of Y taking a specific value given that X takes a specific value. For example, $\mathbb{P}(Y = 1 \mid X = 1)$ means if we randomly select a student from the *population of* 150 and we know she is from Difficult income category (so we are fixing only for Difficult income category), then what is the probability that he/she tried to go abroad for higher studies. The calculation of conditional probability is straightforward, we can use the joint probability and marginal probability as follows,

$$\mathbb{P}(Y = 1 \mid X = 1) = \frac{\mathbb{P}(X = 1, Y = 1)}{\mathbb{P}(X = 1)} = \frac{0.12}{0.27} \approx 0.4444$$

  Or using the $f(x, y)$ and $f_X(x)$ we can write it as (the symbol becomes complicated but the calculation is easy)

$$f_{Y|X=1}(y) = f_{Y|X=1}(1) = \frac{f(x, y)}{f_X(x)} = \frac{f(1, 1)}{f_X(1)} = \frac{0.12}{0.27} \approx 0.4$$

- In fact conditioning on $X = 1$, we can calculate both $Y = 1$ (which we did) and $Y = 0$ as follows, and then write the conditional distribution of Y given $X = 1$, in a table we can write as follows,

Tried/Not Tried (y)	**Probability $f_{Y|X=1}(y)$**
1	0.4
0	0.6

- Note this is conditional distribution of Y given $X = 1$, this is different from marginal distribution of Y which is $f_Y(y)$, which we calculated earlier. And conditional distribution is a distribution so this will sum to 1.

Joint, Marginal and Conditional Distributions

- Now we can also Conditional Expectation $\mathbb{E}(Y \mid X = 1)$ as follows,

$$\begin{aligned}\mathbb{E}(Y \mid X = 1) &= \sum_y y \cdot f_{Y|X=1}(y) \\ &= 1 \cdot 0.4 + 0 \cdot 0.6 = 0.4\end{aligned}$$

- And Conditional Variance $\mathbb{V}(Y \mid X = 1)$ as follows,

$$\begin{aligned}\mathbb{V}(Y \mid X = 1) &= \mathbb{E}[Y - \mathbb{E}(Y \mid X = 1)]^2 \\ &= (1 - 0.4)^2 \cdot 0.4 + (0 - 0.4)^2 \cdot 0.6 = 0.24\end{aligned}$$

- In this case you can think about conditional expectation as a population average of all Y values given $X = x$ (for example $X = 1$). Similat interpretation can be given for conditional variance.
- From this joint distribution we can calculate 4 conditional distribution of Y, given four possible values of X, i.e. $X = 1, 2, 3, 4$. This will give us 4 conditional mean and 4 conditional variance.
- Similatly we can also calculate two conditional distributions of X given $Y = 1$ and $Y = 0$, and then calculate conditional expectation and conditional variance.

Joint, Marginal and Conditional Distributions

- Here is an example olot for joint PMF of X and Y,
  X

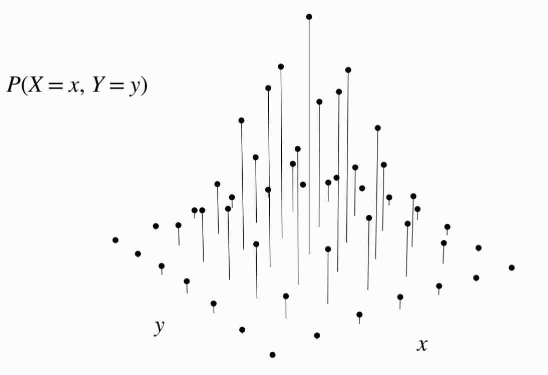


Figure 1: Figure above shows a sketch of what the joint PMF of two discrete random variables could look like. The height of a vertical bar at (x, y) represents the probability $\mathbb{P}(X = x, Y = y)$ or $f(x, y)$. For the joint PMF to be valid, the total height of the vertical bars must be 1 .

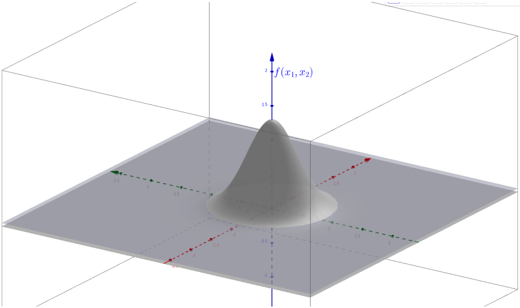
Joint, Marginal and Conditional Distributions

We only looked at discrete random variables, but we can also extend this to continuous random variables. For example, if X and Y are two continuous random variables, then we can define joint probability density function (PDF) $f(x, y)$ such that

$$\mathbb{P}(X \in A, Y \in B) = \iint_{A \times B} f(x, y)\, dx\, dy$$

The things become more complicated when we have continuous random variables, but the idea is similar. We can define marginal PDF $f_X(x)$ and $f_Y(y)$, and then we can define conditional PDF $f_{Y|X}(y \mid x)$ as follows,

Here is an example of *bi-variate Normal or jointly Normal*,



Joint, Marginal and Conditional Distributions

The functions looks a bit more scary, sorry,

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \times$$
$$e^{\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2-2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right\}}$$

Here we have two random variables, X and Y which are jointly normal. Now we have 5 parameters, μ_X, μ_Y, σ_X, σ_Y and ρ. here μ_X and μ_Y are the means of X and Y, σ_X and σ_Y are the standard deviations of X and Y, and ρ is the correlation between X and Y.

Joint, Marginal and Conditional Distributions

For continuous random variables we can also define marginal PDF $f_X(x)$ and $f_Y(y)$, and then we can define conditional PDF $f_{Y|X}(y \mid x)$ with integration, I won't go to details here but impotant is here everything will be a function of x and y. I give one example below

Joint PDF of X and Y is given by

$$f(x, y) = x + \frac{3}{2}y^2, \quad 0 < x < 1, \quad 0 < y < 1$$

In this case from this joint just by integrating we can find **marginal PDF of X and Y as follows**,

$$f_X(x) = x + \frac{1}{2}$$
$$f_Y(y) = \frac{3}{2}y^2$$

We can also calculate **conditional PDF of Y given X** as follows,

$$f_{Y|X}(y \mid X = x) = \frac{f(x, y)}{f_X(x)} = \frac{x + \frac{3}{2}y^2}{x + \frac{1}{2}}$$
$$= \frac{2x + 3y^2}{2x + 1}$$

Notice for each fixed x, this is a density function of y, so this is a conditional PDF of Y given $X = x$. For example, if $x = \frac{1}{2}$, then we can write the conditional PDF of Y given $X = \frac{1}{2}$ as follows,

$$f_{Y|X}(y \mid X = \frac{1}{2}) = \frac{2 \cdot \frac{1}{2} + 3y^2}{2 \cdot \frac{1}{2} + 1} = \frac{1 + 3y^2}{2}$$

If we use $f_{Y|X}(y \mid X = x) = \frac{2x+3y^2}{2x+1}$ and calculate expectation of Y given $X = x$, then we can write as follows,

$$\mathbb{E}(Y \mid X = x) = \frac{1}{2(2x+1)} \left(x + \frac{3}{4} \right)$$

Note that conditional expectation becomes a function of X. This is called conditional expectation function. How do we visualize this, there is a nice way to visualize this in scatter plot. We will come back to this later, however important is conditional expectation is a function of X, so we can write $\mathbb{E}(Y \mid X) = g(X)$, where $g(X)$ is a function of X.

Covariance and Correlation

- Before we end this section we will learn about two other quantities, which are very important in statistics, these are *Covariance* and *Correlation*. Probably you already know the sample covariance and sample correlation, but here we will learn about population covariance and population correlation. These two quantities will help us to understand the relationship between two random variables.
- Here is the formula or definition

Covariance and Correlation

Definition 3.1: (Covariance and Correlation)

The population covariance between two random variables X and Y is

$$\mathbb{C}\text{ov}(X,Y) = \mathbb{E}\left[(X - \mathbb{E}(X))\,(Y - \mathbb{E}(Y))\right] = \mathbb{E}\left[(X - \mu_X)\,(Y - \mu_Y)\right]$$

And the Correlation between two random variables X and Y is

$$\rho_{X,Y} = \text{Cor}(X,Y) = \frac{\mathbb{C}\text{ov}(X,Y)}{\left(\sqrt{\mathbb{V}(X)}\right)\left(\sqrt{\mathbb{V}(Y)}\right)} = \frac{\mathbb{C}\text{ov}(X,Y)}{\sigma_X \times \sigma_Y}$$

- where μ_X and μ_Y are the marginal Expected values of X and Y, and σ_X and σ_Y are the standard deviations of X and Y.
- **What does covariance mean?** If covariance is positive, then X and Y are *positively associated or related*, which roughly means if X increases, then Y also increases. If covariance is negative, then X and Y are *negatively associated / related*, which roughly means if X increases, then Y decreases. If covariance is close to 0, then there is almost no relationship between X and Y.
- Now **What does correlation mean?** Correlation is a normalized version of covariance, which means it gives a value between -1 and 1 (we will always have $-1 \leq \rho_{X,Y} \leq 1$). So it's a better measure of association than covariance, since we can understand the strength of association between X and Y from correlation.

Covariance and Correlation

- In particular if $\rho_{X,Y}$ is close to $+1$, then X and Y are *positively correlated*, which means if X increases, then Y also increases. If $\rho_{X,Y}$ is close to -1, then X and Y are perfectly negatively correlated, which means if X increases, then Y decreases. If $\rho_{X,Y} = 0$, then there is no linear relationship between X and Y.

Covariance and Correlation

- Let's calculate covariance and correlation for our example of EWU students. We can use the joint distribution of X and Y that we calculated earlier, and then we can calculate the covariance and correlation as follows,
- But before that there is also a shortcut formula for covariance, which is (this is easy to derive, please do it as an exercise)

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

- Here for $\mathbb{E}(XY)$, we need the joint distribution of X and Y, which we can calculate as follows,

$$\begin{aligned}
\mathbb{E}(XY) &= \sum_x \sum_y x \cdot y \cdot f(x, y) \\
&= 1 \cdot 1 \cdot 0.12 + 1 \cdot 0 \cdot 0.15 + 2 \cdot 1 \cdot 0.08 + 2 \cdot 0 \cdot 0.17 + 3 \cdot 1 \cdot 0.15 + 3 \cdot 0 \cdot 0.10 \\
&+ 4 \cdot 1 \cdot 0.16 + 4 \cdot 0 \cdot 0.07 \\
&= 0.12 + 0 + 0.16 + 0 + 0.45 + 0 + 0.64 + 0 \\
&= 2.95
\end{aligned}$$

- We need to calculate $\mathbb{E}(X)$ and $\mathbb{E}(Y)$, which we can calculate from the marginal distributions of X and Y that we calculated earlier,

$$\begin{aligned}
\mathbb{E}(X) &= 1 \cdot 0.27 + 2 \cdot 0.25 + 3 \cdot 0.25 + 4 \cdot 0.23 = 2.45 \\
\mathbb{E}(Y) &= 1 \cdot 0.51 + 0 \cdot 0.49 = 0.51
\end{aligned}$$

Covariance and Correlation

- Now we can calculate covariance as follows,

$$\begin{aligned}\text{Cov}(X, Y) &= \mathbb{E}(XY) - \mathbb{E}(X) \cdot \mathbb{E}(Y) \\ &= 2.95 - 2.45 \cdot 0.51 \\ &= 2.95 - 1.25 = 1.70\end{aligned}$$

- Now calculate the standard deviations of X and Y, which we can calculate from the marginal distributions of X and Y that we calculated earlier and then calculate correlation (do it as an exercise),
- Since Covariance is positive, we can say X and Y are positively associated, which means if a student is from higher income categories, then he/she is more likely to try to go abroad for higher studies.
- How strong is the relationship between Y and X? We can use the correlation between Y and X to measure this (please do it as an exercise).

1. Joint, Marginal and Conditional Distributions

2. Covariance and Correlation

3. Independence of random variables

Independence of discrete random variables

We earlier discussed independence of events. Armed with an understanding of joint, marginal, and conditional distributions, we can revisit the definition of independence for random variables.

Definition 3.2: (Independence of discrete random variables)

If X and Y are discrete random variables, then X and Y are *independent* if for all x and y,

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$$

for all x and y. Note that, this means

$$f(x, y) = f(x)f(y)$$

for all x and y

- The definition says that for independent r.v.s, the joint PMF factors into the product of the marginal PMFs.

Independence of discrete random variables

- Remember! in general, the marginal distributions do not determine the joint distribution.
- This means joint distribution comes first, and this is the entire reason why we wanted to study joint distributions in the first place!
- But in the case of independence, the marginal distributions are all we need to specify the joint distribution.
- Try to construct a joint PMF of X and Y after constructing Marginal PMFs of X and Y such that the two random variables are independent.
- Another way of looking at independence is that all the conditional PMFs are the same as the marginal PMF (why is that?), becuase if two random variables are independent conditional probabilities are same as marginal probabilities.
- In other words, starting with the marginal PMF of Y, no updating is necessary when we condition on $X = x$, regardless of what x is.

Independence of discrete random variables

Example 3.3: Show that, X and Y variables whose joint given in Example 3.3 are NOT independent (Note: Finding even one pair of values x and y such that $f(x, y) \neq f(x)f(y)$ is enough to rule out independence.)

Covariance and Independence

Example 3.4: Suppose that the random variable X has probability distribution

$$\mathbb{P}(-1) = 1/4 \quad \mathbb{P}(0) = 1/2 \quad \mathbb{P}(1) = 1/4$$

Let the random variable Y be defined as follows:

$$Y = X^2$$

Thus, knowledge of the value taken by X implies knowledge of the value taken by Y, and, therefore, these two random variables are certainly not independent. Whenever $X = 0$, then $Y = 0$, and if X is either -1 or 1 , then $Y = 1$. The joint probability distribution of X and Y is

$$\mathbb{P}(-1, 1) = 1/4 \quad \mathbb{P}(0, 0) = 1/2 \quad \mathbb{P}(1, 1) = 1/4$$

with the probability of any other combination of values being equal to 0 . It is then straightforward to verify that

$$E[X] = 0 \quad E[Y] = 1/2 \quad E[XY] = 0$$

The covariance between X and Y is 0 . Thus we see that random variables that are not independent can have a covariance equal to 0 .

References