Review of Probability: Discrete Distribution

Dr. Patrick Toche

Textbook:

James H. Stock and Mark W. Watson, *Introduction to Econometrics*, 4th Edition, Pearson.

Other references:

Joshua D. Angrist and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, 1st Edition, Princeton University Press.
Jeffrey M. Wooldridge, *Introductory Econometrics: A Modern Approach*, 7th Edition, Cengage Learning.

The textbook comes with online resources and study guides. Other references will be given from time to time.

Discrete Probability Distribution

X and Y are discrete random variables with the following joint distribution:

		Value of Y				
		14	22	30	40	65
	1	0.02	0.05	0.10	0.03	0.01
Value of X	5	0.17	0.15	0.05	0.02	0.01
	8	0.02	0.03	0.15	0.10	0.09

That is, $\Pr[X = 1, Y = 14]$, and so forth.

Discrete Probability Distribution

1. Calculate the probability distribution, mean, and variance of Y.

		Value of Y					
		14	22	30	40	65	Total
	1	0.02	0.05	0.10	0.03	0.01	0.21
Value of X	5	0.17	0.15	0.05	0.02	0.01	0.40
	8	0.02	0.03	0.15	0.10	0.09	0.39
Total		0.21	0.23	0.30	0.15	0.11	1.00

The probability distribution can be obtained by adding the values in the table. For instance, the joint probability distribution of $Y = 14$ is $\Pr[Y = 14] = 0.02 + 0.17 + 0.02 = 0.21$.

Discrete Probability Distribution

1. Calculate the probability distribution, mean, and variance of Y.

		Value of Y					
		14	22	30	40	65	**Total**
	1	0.02	0.05	0.10	0.03	0.01	0.21
Value of X	5	0.17	0.15	0.05	0.02	0.01	0.40
	8	0.02	0.03	0.15	0.10	0.09	0.39
Total		0.21	0.23	0.30	0.15	0.11	1.00

The probability distribution can be obtained by adding the values in the table. For instance, the joint probability distribution of $Y = 14$ is $\Pr[Y = 14] = 0.02 + 0.17 + 0.02 = 0.21$.

Discrete Probability Distribution

1. Calculate the probability distribution, mean, and variance of Y.

$$
\begin{aligned}
\mathrm{E}[Y] &= 14 \cdot 0.21 + 22 \cdot 0.23 + 30 \cdot 0.30 + 40 \cdot 0.15 + 65 \cdot 11 = 30.15 \\
\mathrm{E}[Y^2] &= 14^2 \cdot 0.21 + 22^2 \cdot 0.23 + 30^2 \cdot 0.30 + 40^2 \cdot 0.15 + 65^2 \cdot 11 = 1127.23 \\
\mathrm{var}[Y] &= \mathrm{E}[Y^2] - (\mathrm{E}[Y])^2 = 1127.23 - (30.15)^2 = 218.21 \\
\sigma_Y &= \sqrt{218.21} = 14.77
\end{aligned}
$$

(Note: equalities are typically approximate)

Discrete Probability Distribution

1. Calculate the probability distribution, mean, and variance of Y.

$$
\begin{aligned}
\mathrm{E}[Y] &= 14 \cdot 0.21 + 22 \cdot 0.23 + 30 \cdot 0.30 + 40 \cdot 0.15 + 65 \cdot 11 = 30.15 \\
\mathrm{E}[Y^2] &= 14^2 \cdot 0.21 + 22^2 \cdot 0.23 + 30^2 \cdot 0.30 + 40^2 \cdot 0.15 + 65^2 \cdot 11 = 1127.23 \\
\mathrm{var}[Y] &= \mathrm{E}[Y^2] - (\mathrm{E}[Y])^2 = 1127.23 - (30.15)^2 = 218.21 \\
\sigma_Y &= \sqrt{218.21} = 14.77
\end{aligned}
$$

(Note: equalities are typically approximate)

Discrete Probability Distribution

2. Calculate the probability distribution, mean, and variance of Y given $X = 8$.

$$\begin{aligned}
\mathrm{E}[Y|X=8] &= 14 \cdot \frac{0.02}{0.39} + 22 \cdot \frac{0.03}{0.39} + 30 \cdot \frac{0.15}{0.39} + 40 \cdot \frac{0.10}{0.39} + 65 \cdot \frac{0.09}{0.39} = 39.21 \\
\mathrm{E}[Y^2|X=8] &= 14^2 \cdot \frac{0.02}{0.39} + 22^2 \cdot \frac{0.03}{0.39} + 30^2 \cdot \frac{0.15}{0.39} + 40^2 \cdot \frac{0.10}{0.39} + 65^2 \cdot \frac{0.09}{0.39} = 1778.70 \\
\mathrm{var}[Y|X=8] &= \mathrm{E}[Y^2|X=8] - (\mathrm{E}[Y|X=8])^2 = 1778.70 - (39.21)^2 = 241.65 \\
\sigma_{Y|X=8} &= \sqrt{241.65} = 15.54
\end{aligned}$$

Discrete Probability Distribution

2. Calculate the probability distribution, mean, and variance of Y given $X = 8$.

$$\begin{aligned}
\mathrm{E}[Y|X=8] &= 14 \cdot \frac{0.02}{0.39} + 22 \cdot \frac{0.03}{0.39} + 30 \cdot \frac{0.15}{0.39} + 40 \cdot \frac{0.10}{0.39} + 65 \cdot \frac{0.09}{0.39} = 39.21 \\
\mathrm{E}[Y^2|X=8] &= 14^2 \cdot \frac{0.02}{0.39} + 22^2 \cdot \frac{0.03}{0.39} + 30^2 \cdot \frac{0.15}{0.39} + 40^2 \cdot \frac{0.10}{0.39} + 65^2 \cdot \frac{0.09}{0.39} = 1778.70 \\
\mathrm{var}[Y|X=8] &= \mathrm{E}[Y^2|X=8] - (\mathrm{E}[Y|X=8])^2 = 1778.70 - (39.21)^2 = 241.65 \\
\sigma_{Y|X=8} &= \sqrt{241.65} = 15.54
\end{aligned}$$

Discrete Probability Distribution

3. Calculate the covariance and correlation between X and Y.

$$
\begin{aligned}
\mathrm{E}[XY] = {} & 14 \cdot (0.02 \cdot 1 + 0.17 \cdot 5 + 0.02 \cdot 8) \\
& + 22 \cdot (0.05 \cdot 1 + 0.15 \cdot 5 + 0.03 \cdot 8) \\
& + 30 \cdot (0.10 \cdot 1 + 0.05 \cdot 5 + 0.15 \cdot 8) \\
& + 40 \cdot (0.03 \cdot 1 + 0.02 \cdot 5 + 0.10 \cdot 8) \\
& + 65 \cdot (0.01 \cdot 1 + 0.01 \cdot 5 + 0.09 \cdot 8) \\
= {} & 171.7
\end{aligned}
$$

This is when you appreciate the power of computing machines!

$$\mathrm{cov}[XY] = \mathrm{E}[XY] - \mathrm{E}[X] \times \mathrm{E}[Y] = 171.70 - 5.33 \times 30.15 = 11.00$$

$$\mathrm{corr}[XY] = \frac{\mathrm{cov}[XY]}{\mathrm{var}[X] \times \mathrm{var}[Y]} = \frac{11.00}{2.60 \times 14.77} = 0.286$$

Discrete Probability Distribution

3. Calculate the covariance and correlation between X and Y.

$$\begin{aligned}
\mathrm{E}[XY] = 14 &\cdot (0.02 \cdot 1 + 0.17 \cdot 5 + 0.02 \cdot 8) \\
+\ 22 &\cdot (0.05 \cdot 1 + 0.15 \cdot 5 + 0.03 \cdot 8) \\
+\ 30 &\cdot (0.10 \cdot 1 + 0.05 \cdot 5 + 0.15 \cdot 8) \\
+\ 40 &\cdot (0.03 \cdot 1 + 0.02 \cdot 5 + 0.10 \cdot 8) \\
+\ 65 &\cdot (0.01 \cdot 1 + 0.01 \cdot 5 + 0.09 \cdot 8) \\
= 171.7 &
\end{aligned}$$

This is when you appreciate the power of computing machines!

$$\mathrm{cov}[XY] = \mathrm{E}[XY] - \mathrm{E}[X] \times \mathrm{E}[Y] = 171.70 - 5.33 \times 30.15 = 11.00$$

$$\mathrm{corr}[XY] = \frac{\mathrm{cov}[XY]}{\mathrm{var}[X] \times \mathrm{var}[Y]} = \frac{11.00}{2.60 \times 14.77} = 0.286$$

3. Calculate the covariance and correlation between X and Y.

$$
\begin{aligned}
\mathrm{E}[XY] = {} & 14 \cdot (0.02 \cdot 1 + 0.17 \cdot 5 + 0.02 \cdot 8) \\
& + 22 \cdot (0.05 \cdot 1 + 0.15 \cdot 5 + 0.03 \cdot 8) \\
& + 30 \cdot (0.10 \cdot 1 + 0.05 \cdot 5 + 0.15 \cdot 8) \\
& + 40 \cdot (0.03 \cdot 1 + 0.02 \cdot 5 + 0.10 \cdot 8) \\
& + 65 \cdot (0.01 \cdot 1 + 0.01 \cdot 5 + 0.09 \cdot 8) \\
= {} & 171.7
\end{aligned}
$$

This is when you appreciate the power of computing machines!

$$\mathrm{cov}[XY] = \mathrm{E}[XY] - \mathrm{E}[X] \times \mathrm{E}[Y] = 171.70 - 5.33 \times 30.15 = 11.00$$

$$\mathrm{corr}[XY] = \frac{\mathrm{cov}[XY]}{\mathrm{var}[X] \times \mathrm{var}[Y]} = \frac{11.00}{2.60 \times 14.77} = 0.286$$