Econometrics 120A: Discussion Section

Week 5

Natalia Madrid & Lapo Bini

Department of Economics

Chapter 4: Two Random Variables

Joint Probability Distribution

The Joint Probability Distribution represents the probability of two events, X and Y, occurring simultaneously.

Discrete Bivariate Random Variables:

- ⇨ The joint probability mass function (PMF) is given by:

$$P(X = x, Y = y) = p(x, y)$$

- ⇨ Properties:

$$\sum_x \sum_y P(X = x, Y = y) = 1$$

Continuous Bivariate Random Variables:

- ⇨ The joint probability density function (PDF) is represented as:

$$P(a < X < b, c < Y < d) = \int_a^b \int_c^d p(x, y)\, dy\, dx$$

- ⇨ Properties:

$$\iint_{-\infty}^{\infty} p(x, y)\, dy\, dx = 1$$

Marginal Distribution

A marginal distribution captures the probabilities associated with one variable alone, without considering the dependency on other variables.

Discrete Bivariate Random Variables:

- ⇨ The marginal distribution of Y is obtained by summing the joint probabilities over all values of X.

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

Continuous Bivariate Random Variables:

- ⇨ The marginal distribution of Y is obtained by summing the joint probabilities over all values of X.

$$P(Y = y) = \int_{-\infty}^{\infty} p(x, y) dx$$

Independence

Independence means knowing the value of one variable provides no information about the other.

For two independent random variables X and Y :

- ⇨ The **joint probability distribution** is equal to the product of the marginals:

$$P(X, Y) = P(X) \cdot P(Y)$$

- ⇨ The **variance** of their sum is the sum of their variances (covariance is 0):

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

- ⇨ The **expected value** of the product of X and Y can be calculated simply as the product of their individual expected values:

$$E[XY] = E[X] \cdot E[Y]$$

- ⇨ The **conditional expectation** of one variable given the other is simply the expectation of that variable:

$$E[X|Y] = E[X] \quad \text{and} \quad E[Y|X] = E[Y]$$

Conditional probability distribution

A conditional probability distribution describes the probability of a random variable X given that another random variable Y has taken on specific values. It is denoted as $P(Y|X)$ and is defined as $P(Y = y|X = x)$

Bayes' Rule allows us to update our beliefs about X given new evidence Y.

$$P(X = x \mid Y = y) = \frac{P(Y = y, X = x) \cdot P(X = x)}{P(Y = y)}$$

If Y is a random variable and X is another random variable, the **expected value** of Y given X is defined as:

$$E[Y|X] = \sum_y y \cdot P(Y = y|X) \quad \text{(for discrete random variables)}$$

$$E[Y|X] = \int_{-\infty}^{\infty} y \cdot p(y|x)\, dy \quad \text{(for continuous random variables)}$$

Covariance and independence

Covariance measures the degree to which two random variables change together. Specifically, it is defined as:

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

Independence Implies Zero Covariance: If X and Y are independent random variables, then their covariance is zero:

- ⇨ Due to independence, we have:

$$E[XY] = E[X] \cdot E[Y]$$

- ⇨ Therefore, the covariance becomes:

However, the converse is not necessarily true: **Zero Covariance Does Not Imply Independence**

Properties expectation and variance

The **expectation** of a linear combination of variables is the linear combination of their expectations.

$$E[aX + bY] = aE[X] + bE[Y]$$

For two variables, their combined **variance** accounts for individual variances and their covariance.

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$$

Kind of PS3 Questions 1-2

Suppose the following table represents the joint distribution of two variables: **Wage (W)** and **Education Level (E)** in a small economy. Here, *W* represents an individual's income category (low, medium, high), and *E* represents their education level (high school, bachelor's, master's).

W \E	High School	Bachelor's	Master's
Low	0.10	0.05	0.02
Medium	0.15	0.20	0.08
High	0.05	0.10	0.25

(a) Are **Wage** and **Education Level** independent? Justify your answer by calculating the marginal distributions and checking if the product of marginals equals the joint probabilities.

(b) Calculate the conditional distribution of **Wage** given **Education Level = Bachelor's**. Is this the same as the marginal distribution for **Wage**? Is this consistent with your answer in the previous question?

Kind of PS3 Questions 1-2

W \E	High School	Bachelor's	Master's
Low	0.10	0.05	0.02
Medium	0.15	0.20	0.08
High	0.05	0.10	0.25

(a) Are **Wage** and **Education Level** independent? Justify your answer by calculating the marginal distributions and checking if the product of marginals equals the joint probabilities.

Kind of PS3 Questions 1-2

W \E	High School	Bachelor's	Master's
Low	0.10	0.05	0.02
Medium	0.15	0.20	0.08
High	0.05	0.10	0.25

(b) Calculate the conditional distribution of **Wage** given **Education Level = Bachelor's**. Is this the same as the marginal distribution for **Wage**? Is this consistent with your answer in the previous question?

Kind of PS3 Question 3

Suppose we have the following joint distribution for two economic variables: *Employment Status* (X) and *Industry* (Y). The variable X represents employment status, where:

$X = 0$ denotes "Unemployed"

$X = 1$ denotes "Employed in a Traditional Job"

$X = 2$ denotes "Self-Employed"

The variable Y represents industry type, where:

$Y = 0$ denotes "Agriculture"

$Y = 1$ denotes "Finance"

The joint probabilities for these variables are given in the table below:

$X \setminus Y$	0 (Agriculture)	1 (Finance)
0 (Unemployed)	0.05	0.15
1 (Employed)	0.20	0.25
2 (Self-employed)	0.10	0.25

Kind of PS3 Question 3

The joint probabilities for these variables are given in the table below:

$X \setminus Y$	0 (Agriculture)	1 (Finance)
0 (Unemployed)	0.05	0.15
1 (Employed)	0.20	0.25
2 (Self-employed)	0.10	0.25

Using this table, answer the following questions:

(a) Verify that this is indeed a probability distribution.

(b) What are the marginal distributions for X and Y?

(c) What are the means of X and Y?

(d) What is the conditional distribution of X given $Y = 0$? Also for X given $Y = 1$.

(e) What do you notice about this result? Does this show that X and Y are independent?

(f) How does the conditional mean of X depend on Y?

(g) Does the result in (f) indicate if the random variables are independent? Be clear.

Kind of PS3 Question 3

Marginal Distribution

The joint probabilities for these variables are given in the table below:

$X \setminus Y$	0 (Agriculture)	1 (Finance)
0 (Unemployed)	0.05	0.15
1 (Employed)	0.20	0.25
2 (Self-employed)	0.10	0.25

(a) Verify that this is indeed a probability distribution.

Marginal Distribution

The joint probabilities for these variables are given in the table below:

$X \setminus Y$	0 (Agriculture)	1 (Finance)
0 (Unemployed)	0.05	0.15
1 (Employed)	0.20	0.25
2 (Self-employed)	0.10	0.25

(b) What are the marginal distributions for X and Y?

Kind of PS3 Question 3

The joint probabilities for these variables are given in the table below:

$X \setminus Y$	0 (Agriculture)	1 (Finance)
0 (Unemployed)	0.05	0.15
1 (Employed)	0.20	0.25
2 (Self-employed)	0.10	0.25

(c) What are the means of X and Y?

Kind of PS3 Question 3

The joint probabilities for these variables are given in the table below:

$X \setminus Y$	0 (Agriculture)	1 (Finance)
0 (Unemployed)	0.05	0.15
1 (Employed)	0.20	0.25
2 (Self-employed)	0.10	0.25

(d) What is the conditional distribution of X given $Y = 0$? Also for X given $Y = 1$.

Kind of PS3 Question 3

The joint probabilities for these variables are given in the table below:

$X \setminus Y$	0 (Agriculture)	1 (Finance)
0 (Unemployed)	0.05	0.15
1 (Employed)	0.20	0.25
2 (Self-employed)	0.10	0.25

(e) What do you notice about this result? Does this show that X and Y are independent?

Kind of PS3 Question 3

The joint probabilities for these variables are given in the table below:

$X \setminus Y$	0 (Agriculture)	1 (Finance)
0 (Unemployed)	0.05	0.15
1 (Employed)	0.20	0.25
2 (Self-employed)	0.10	0.25

(f) How does the conditional mean of X depend on Y?

Kind of PS3 Question 3

The joint probabilities for these variables are given in the table below:

$X \setminus Y$	0 (Agriculture)	1 (Finance)
0 (Unemployed)	0.05	0.15
1 (Employed)	0.20	0.25
2 (Self-employed)	0.10	0.25

(g) Does the result in (f) indicate if the random variables are independent? Be clear.

Kind of PS3 Question 7

A random variable X has mean $E[X] = 20$ and variance of 25; a second random variable Y has mean $E[Y] = 30$ and variance of 55.

(a) What is the mean and variance of $X + Y$ assuming they are independent?

(b) What is the mean and variance of $X + Y$ if they gave correlation $\rho = 0.5$?

Chapter 5: Random Sampling

Where Are We?

Big Picture

⇨ We started from a Bernoulli Random Variable

$$X_i = \begin{cases} 0 & \text{(not cured)} & p \\ 1 & \text{(cured)} & (1-p) \end{cases}$$

⇨ Standard Assumption: $X_1, \ldots, X_n$ independent.

⇨ We define the random variable $S_n = \sum_{i=1}^{n} X_i$ with distribution $S_n \sim Binomial(n, p)$

⇨ Now we are interested in the distribution of $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$

Moments of Sample Mean

- We can easily compute the expected value of the sample mean:

  $E[\bar{X}_n] =$

- We can compute the variance of the sample mean as well:

  $Var(\bar{X}_n) =$

Asymptotic VS Finite Sample

Central Limit Theorem

⇨ If $\{X_1, \ldots, X_n\}$ are a VSRS with mean μ and variance σ^2 then

$$\bar{X}_n \overset{a}{\sim} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

⇨ Example: if we consider the Bernoulli random variables introduced before we have $\bar{X}_n \overset{a}{\sim} \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$

⇨ Given $Y \sim \mathcal{N}(\mu, \sigma^2)$, if we define $Z = aY + b$ by properties of expectation and variance we obtain:

$$Z = aY + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

⇨ $n\bar{X}_n \overset{a}{\sim} \mathcal{N}(p, p(1-p))$ but we know that $n\bar{X}_n = S_n \sim Binomial(n, p)$. Is there something wrong?

Asymptotic VS Finite Sample

Central Limit Theorem

- **Finite Sample:** When dealing with a finite sample, we cannot rely on the Central Limit Theorem (CLT) to perfectly approximate the distribution of the sample mean as normal, since the CLT only applies asymptotically (as $n \to \infty$).
- **Asymptotic:** As n becomes very large, the CLT approximation improves, and the sample mean $\bar{X}_n$ behaves increasingly like a normal distribution. However, this is an approximation that holds in the limit, meaning it is not exact for finite samples.
- For the **Bernoulli random variables** with mean p and variance $p(1-p)$, the sample mean $\bar{X}_n$ is approximated as follows according to the CLT:

$$\bar{X}_n \approx \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$$

  However, for a finite sample size, the actual distribution of $\bar{X}_n$ is binomial (not normal), and the normal distribution is only an approximation, which may not hold well for small n.