Econometrics I

José Luis Montiel Olea

Lectures 3 and 4

MV Distributions, Independence, and Conditional Probability

Overview

- ⋆ Define random vectors along with multivariate c.d.f.s
  (mean, covariance matrix, moment generating function)
- ⋆ Present some useful characterizations of independence
  (general definition relegated to the appendix of the notes)
- ⋆ Introduce conditional probability and conditional expectation
  (general definition also in the appendix of the notes)

1. Random Vectors and MV distributions

Introduction

- ⋆ So far, we have been working with real-valued random variables:

$$X : \Omega \to \mathbb{R}$$

- ⋆ Consequently, we have learned to think about statements like:

$$P_X(X \leq x),$$

- ⋆ where x is some real number.

Motivation

- ⋆ Econ data usually involve more than one random variable (think about cross-sectional or time series data)

- ⋆ Thus, we will work with $X_s : \Omega \to \mathbb{R}, \quad s \in \{1, 2 \ldots S\}$

- ⋆ We will introduce the following statements:

  1. Joint Probability Statements.

$$\mathbb{P}_X \Big[X_1 \leq x_1, \ldots X_S \leq x_s \Big]$$

  2. Conditional Probability Statements.

$$\mathbb{P}_X \Big[X_1 \leq x_1 \Big| X_2 \leq x_2 \Big]$$

$\mathbb{R}^s$-valued random variable

⋆ The $\mathbb{R}^S$-valued mapping defined over $(\Omega, \mathcal{F})$

$$\mathbf{X}(\omega) \equiv \Big(X_1(\omega), \ldots, X_S(\omega) \Big)'$$

is a random vector if for all $A \in B(\mathbb{R}^S)$

$$\mathbf{X}^{-1}(A) \in \mathcal{F}.$$

⋆ The definition is analogous to real-valued case

Multivariate Cumulative Distribution Functions

⋆ The c.d.f. of the $\mathbb{R}^S$ valued random vector $\mathbf{X}(\omega)$ is a function

$$F_X : \mathbb{R}^S \to [0, 1]$$

defined as

$$F_X(x_1, \ldots, x_S) \equiv \mathbb{P}\{\omega \in \Omega \mid X_i(\omega) \leq x_i \text{ for all } i = 1, \ldots, S.\}$$

⋆ Thus, the c.d.f. tell us how often each X_i is below x_i.

We classify random vectors according to their c.d.f.s
(discrete and continuous)

Absolutely Continuous Random Vector

⋆ An $\mathbb{R}^S$-valued random vector is absolutely continuous if:

$$F(x_1, x_2, \ldots x_S) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_S} f(z_1, \ldots z_S) dz_1 \ldots dz_S$$

for some nonnegative function $f : \mathbb{R}^S \to \mathbb{R}^+$.

⋆ $f(x_1, \ldots, x_n) = \partial^n F(x_1, \ldots, x_n) / \partial x_1 \ldots \partial x_n$ is the p.d.f. of $\mathbf{X}$.

Marginal Distributions of $\mathbf{X}$

$$F_s : \mathbb{R} \to [0, 1]$$

$$F_s(x) \equiv \mathbb{P}\Big[\mathbf{X}^{-1}\Big(\mathbb{R} \times \ldots (-\infty, x) \ldots \times \mathbb{R}\Big)\Big]$$

From joint to marginals

How to recover a marginal p.d.f. from a joint p.d.f?

Just integrate variables out.

Moments of Random Vector

Let $g : \mathbb{R}^S \to \mathbb{R}^m$. Write

$$g(\mathbf{x}) = (g_1(\mathbf{x}), g_2(\mathbf{x}), \ldots g_m(\mathbf{x}))'$$

and let

$$\begin{aligned}
\mathbb{E}_F[g(\mathbf{X})] &= \Big(\mathbb{E}_F[g_1(\mathbf{X})], \mathbb{E}_F[g_2(\mathbf{X})], \ldots, \mathbb{E}_F[g_m(\mathbf{X})]\Big)' \\
&\equiv \left(\int_{\mathbb{R}^S} g_1(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}, \ldots, \int_{\mathbb{R}^S} g_m(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}\right)'
\end{aligned}$$

Mean, Variance, Covariance

$$\mu \equiv \mathbb{E}[\mathbf{X}],\ \Sigma \equiv \mathbb{E}[(\mathbf{X} - \mu)(\mathbf{X} - \mu)']$$

$$\mathrm{Cov}(X, Y) \equiv \mathbb{E}[(X - \mu_x)(Y - \mu_y)].$$

Moment Generating Function of **X**

The moment generating function of $m_{\mathbf{X}} : \mathbb{R}^S \to \mathbb{R}$ is given by:

$$m_{\mathbf{X}}(t) \equiv \mathbb{E}_F[\exp(t'\mathbf{X})] \quad t \in \mathbb{R}^S$$

Remarks about the m.g.f.

- $\star$ Vectors with the same m.g.f. have the same joint distribution
- $\star$ Vectors with the same distribution $\forall$ linear combinations have the same joint distribution
  (see Cramer-Wold Theorem in the notes and problem 2)

Examples of Bivariate Vectors

(Bivariate Normal and Bivariate Bernoulli)

Bivariate Normal

- ⋆ Let $\mu \in \mathbb{R}^2$ and let Σ be a p.s.d. matrix of dimension 2×2.
- ⋆ $\mathbf{X} \sim \mathcal{N}_2(\mu, \Sigma)$, if:

$$\mathbb{E}_F[\exp(t'\mathbf{X})] = \exp\left(t'\mu + \frac{1}{2}t'\Sigma t\right).$$

Some Properties of the Bivariate Normal

1. $\mu \in \mathbb{R}^2, A \in \mathbb{R}^{2\times 2}$.

$$\mathbf{Z} \sim \mathcal{N}_2(0, \mathbb{I}_2) \implies \mu + A\mathbf{Z} \sim \mathcal{N}_2(\mu, AA').$$

2. $\mathbf{X} \sim \mathcal{N}(\mu, \Sigma) \iff c'\mathbf{X} \sim \mathcal{N}(c'\mu, c'\Sigma c)$ for all $c \in \mathbb{R}^2$.

3. $\mathbf{X} \sim \mathcal{N}(\mu, \Sigma)$, Σ invertible. The p.d.f. of $\mathbf{X}$ is:

$$f(\mathbf{x}) = \frac{1}{(\det 2\pi\Sigma)^{1/2}} \exp\Big(-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\Big).$$

Bivariate Bernoulli

$$\text{Supp} = \Big\{(0,0),\ (0,1),\ (1,0),\ (1,1)\Big\}$$

		Y = 0	Y = 1
X	0	p_1	p_2
	1	p_3	p_4

$$p_1 + p_2 + p_3 + p_4 = 1$$

(what are the marginal distributions?)

Remark: Joints are not 'identified' by marginals

$$X \sim \text{Bernoulli}(p_x), \quad Y \sim \text{Bernoulli}(p_y)$$

		Y: 0	Y: 1
X	0	p_1	p_2
	1	p_3	p_4

$$\begin{aligned} p_2 + p_4 &= p_y \\ p_3 + p_4 &= p_x \\ p_1 + p_2 + p_3 + p_4 &= 1 \end{aligned}$$

Solve for p_1, p_2, p_3, p_4.

Best Linear Predictor (Practice Problem)

- ⋆ Let X, Y be real-valued random variables.
- ⋆ Assume $\mu = (\mu_x, \mu_y)'$ and Σ are known.
- ⋆ "Predict" Y using using a linear function of $(X - \mu_x)$:

$$\alpha + \beta(X - \mu_x)$$

- ⋆ The best linear predictor minimizes expected squared error

$$\min_{\alpha,\beta} \mathbb{E}[(Y - \alpha - \beta(X - \mu_x))^2]$$

- ⋆ Show that $\alpha^* = \mu_y$, $\beta^* = \mathrm{Cov}(X, Y)/V(X)$.

2. Independence

(In)dependence

⋆ Important issue in the multivariate world:

How to summarize dependence or lack of dependence between random variables?

⋆ Say $X_1, \dots, X_n$ are independent if for any $A_1, \dots, A_n$

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \dots \mathbb{P}(X_n \in A_n).$$

⋆ i.e., joint distribution equals the product of the marginals.

Are X and Y independent?

		Y = 0	Y = 1
X	0	.3	.2
	1	.5	0

Useful Characterizations

⋆ Joint c.d.f is the product of the marginal c.d.f.s

$$F(X_1, \ldots, X_n) = F(X_1) \ldots F(X_n).$$

⋆ Joint p.d.f. is the product of marginal p.d.f.s

$$f(x_1, \ldots, x_n) = f(x_1) \ldots f(x_n).$$

⋆ Expectation of "products" is the "product" of expectations

$$\mathbb{E}[g_1(X_1), \ldots, g_n(X_n)] = \mathbb{E}[g_1(X_1)] \ldots \mathbb{E}[g_n(X_n)].$$

⋆ Joint m.g.f. is the product of the marginal m.g.f.s

$$\mathbb{E}[\exp(\mathbf{t}'\mathbf{X})] = \mathbb{E}[\exp(t_1 X_1)] \ldots \mathbb{E}[\exp(t_n X_n)]$$

Independence implies zero Covariance

$$\mathrm{Cov}(X, Y) \equiv \mathbb{E}[(X - \mu_x)(Y - \mu_y)]$$

$$\implies$$

$$\mathrm{Cov}(X, Y) \equiv \mathbb{E}[XY] - \mu_x \mu_y.$$

Therefore (X, Y) independent $\implies \mathrm{Cov}(X, Y) = 0$.

Does zero covariance implies independence?

- ⋆ In general, the answer is no. Consider:

		Y = 0	Y = 1
X	-1	0	3/9
	0	3/9	0
	$+1$	0	3/9

- ⋆ But in some cases like multivariate normals, the answer is yes. (I will ask you to work this out in this week's problem set)

3. Conditional Probability and Conditional Expectation

Definition of the conditional probability function

⋆ $P(Y \in A|x)$: Conditional probability of $Y \in A$ given x.

⋆ Defined as the *function* such that

$$\int_B P(Y \in A|x) f_X(x) dx = P(Y \in A, X \in B).$$

⋆ When X, Y have joint p.d.f. f then

$$P(Y \in A|x) = \int_A \frac{f(x,y)}{f_X(x)} dy$$

⋆ The p.d.f. of $Y|X$ is defined as:

$$f(y|x) \equiv \frac{f(x,y)}{f_X(x)}.$$

In our example

		Y	
		0	1
X	0	p_1	p_2
	1	p_3	p_4

$$P(Y = 1|X = 1) = \frac{p_4}{p_3 + p_4}$$

$$P(Y = 1|X = 0) = \frac{p_2}{p_1 + p_2}$$

Conditional Expectation

⋆ If (X, Y) have joint p.d.f. f(x,y):

$$\mathbb{E}[g(Y)|x] \equiv \int g(y)\frac{f(x,y)}{f_X(x)}dy.$$

⋆ Law of Iterated Expectations $\mathbb{E}[\mathbb{E}[g(Y)|x]] = \mathbb{E}[g(Y)]$

Example

			Y	
			0	1
X	0		p_1	p_2
	1		p_3	p_4

$$\mathbb{E}[Y|X=1] \equiv (0)\frac{p_3}{p_3+p_4} + (1)\frac{p_4}{p_3+p_4}$$

Bivariate Normal

In the problem set I will ask you show that if (X, Y) are bivariate normal:

$$Y|X \sim \mathcal{N}_1(\underbrace{\alpha^* + \beta^*(X - \mu_x)}_{\text{Best Linear Pred}}, \mathrm{Var}((\underbrace{Y - \alpha^* - \beta^*(X - \mu_x)}_{\text{Approximation Error}})))$$

Signal and Noise

In the problem set I also ask you to consider the model

$$\underbrace{X}_{\text{Noisy Measure}} = \underbrace{\theta}_{\text{signal}} + \underbrace{\epsilon}_{\text{noise}},$$

$$\theta \sim \mathcal{N}(\mu, \sigma_\theta^2), \quad \epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2), \quad \theta \perp \epsilon$$

and to work-out the distribution of $\theta | X$.

4. Sums of Random Variables

Let's go back to the example

		Y: 0	Y: 1
X	0	p_1	p_2
	1	p_3	p_4

For any $t_1, t_2 \in \mathbb{R}$ define

$$W = t_1 X + t_2 Y$$

Distribution of $t_1 X + t_2 Y$

⋆ What is the distribution of W?

$$\text{Supp} = \{0, t_1, t_2, t_1 + t_2\}$$

$$\mathbb{P}_Z(W = w)?$$

⋆ Note that:

$$\mathbb{P}_Z[W = t_1 + t_2] = p_4, \quad \mathbb{P}_Z[W = t_2] = p_2, \quad \mathbb{P}_Z[W = t_1] = p_3$$

$$\mathbb{P}_Z[W = 0] = p_1$$

Sums of Independent Random Variables

- ⋆ The distribution of $X_1 + X_2$ need not be easy to obtain
- ⋆ If X_1 and X_2 are independent and have m.g.f.s, it is

$$\mathbb{E}[\exp(t(X_1 + X_2))] = \mathbb{E}[\exp(tX_1)]\mathbb{E}[\exp(tX_2)]$$

- ⋆ Also, if X_1 and X_2 are independent and have p.d.f.s f and g;

$$Z = X + Y \text{ has p.d.f. } \int f(z-y)g(y)dy$$