Lecture 0: Review of Probability Theory On Random Vectors

Econ 205A: Econometric Methods I

Ruoyao Shi

Fall 2018

Contents

1 Random Vectors

Notation and Convention

- Uppercase letters such as X, Y, Z denote random variables; lowercase letters such as x, y_1, z_2 denote particular values (numbers) the random variables take.
- Bold-faced uppercase letters such as $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ denote random vectors; bold-faced lowercase letters such as $\mathbf{x}, \mathbf{y}, \mathbf{z}$ denote particular values (constant vectors) the random vectors take.
- $\equiv$ means the left hand side is defined as the right hand side.
- Extended real number: $\bar{\mathbb{R}} \equiv \mathbb{R} \cup \{+\infty, -\infty\}$.

Random Vectors

- **Definition 1** *A k-dimensional **random vector** $\mathbf{X}$ is a function defined on a probability space $\{S, \mathcal{F}, P\}$ into $\mathbb{R}^k$. The random vector / function $\mathbf{X}(\cdot)$ assigns one and only one value $x \in \mathbb{R}^k$ to each $s \in S$, and for any $x_1, \dots, x_k \in \mathbb{R}$,*

$$A \equiv \{s \in S : X_1(s) \leq x_1, \dots, X_k(s) \leq x_k\} \in \mathcal{F}.$$

- Each element / component / coordinate of a random variable $\mathbf{X}$ is a random variable.

- $\mathbf{X} \in \mathbb{R}^k$: true or not? $\mathbf{x} \in \mathbb{R}^k$: true or not?

- **Definition 2** *The **(joint) cumulative distribution function (cdf)** F of a random vector $\mathbf{X}$ is a function from $\mathbb{R}^k$ to $\mathbb{R}$ such that for any $x_1, \dots, x_k \in \mathbb{R}$,*

$$F(x_1, \dots, x_k) \equiv P(X_1 \leq x_1, \dots, X_k \leq x_k).$$

- Similar to the cdf of a random variable, the cdf of a random vector has the following properties:

  (i) $F(x_1, \dots, x_k)$ is monotonically non-decreasing in each of its arguments:

  $$F(x_1, \dots, x_k) \geq F(y_1, \dots, y_k) \text{ for any } x_j \geq y_j,\ j = 1, \dots, k.$$

  (ii) $F(x_1, \dots, x_k)$ is right continuous in each of its arguments:

  $$\lim_{h \downarrow 0} F(x_1, \dots, x_i + h, \dots, x_k) = F(x_1, \dots, x_k).$$

  (iii) Since cdf are probabilities,

  $$\lim_{\min x_j \to -\infty} F(x_1, \dots, x_k) = 0,$$
  $$\lim_{\min x_j \to +\infty} F(x_1, \dots, x_k) = 1.$$

- If the support of $\mathbf{X}$ is a rectangular $[\underline{x}_1, \bar{x}_1] \times \cdots \times [\underline{x}_k, \bar{x}_k]$, then $F(x_1, \dots, \underline{x}_j, \dots, x_k) = 0$ (for any $j = 1, \dots, k$) and $F(\bar{x}_1, \dots, \bar{x}_k) = 1$.

- **Definition 3** *Let* $\mathbf{X}$ *be a random vector and* $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *be its (*k_1*- and* k_2*-dimensional) subvectors. Let* F *be the joint cdf of* $\mathbf{X}$ *and* F_1 *and* F_2 *be the joint cdf of* $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *respectively.* F_1 *and* F_2 *are called* ***marginal distribution functions*** *of* F.

- $F_1(x_1, \ldots, x_{k_1}) = F(x_1, \ldots, x_{k_1}, +\infty, \ldots, +\infty)$, and $F_2(x_{k_1+1}, \ldots, x_k) = F(+\infty, \ldots, +\infty, x_{k_1+1}, \ldots, x_k)$.

- **Definition 4** *Let* $(\mathbf{X}_1', \mathbf{X}_2')'$ *be a discrete random vector. The function* $f(\mathbf{x}_1, \mathbf{x}_2)$ *from* $\mathbb{R}^k$ *to* $\mathbb{R}$ *defined by* $f(\mathbf{x}_1, \mathbf{x}_2) \equiv P(\mathbf{X}_1 = \mathbf{x}_1, \mathbf{X}_2 = \mathbf{x}_2)$ *is called the* ***joint probability mass function*** *or* ***joint pmf*** *of* $(\mathbf{X}_1', \mathbf{X}_2')'$. *Sometimes, we use the subscripted* $f_{\mathbf{X}_1,\mathbf{X}_2}(\mathbf{x}_1, \mathbf{x}_2)$ *to emphasize (and to distinguish from others) that the pmf is of the random vector* $(\mathbf{X}_1', \mathbf{X}_2')'$.

- **Theorem 1** *Let* $(\mathbf{X}_1', \mathbf{X}_2')'$ *be a discrete random vector with joint pmf* $f_{\mathbf{X}_1,\mathbf{X}_2}(\mathbf{x}_1, \mathbf{x}_2)$. *Then the* ***marginal pmfs*** *of* $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *are given by*

$$f_{\mathbf{X}_1}(\mathbf{x}) \equiv P(\mathbf{X}_1 = \mathbf{x}) = \sum_{\mathbf{x}_2 \in \mathbb{R}^{k_2}} f_{\mathbf{X}_1,\mathbf{X}_2}(\mathbf{x}, \mathbf{x}_2),$$
$$f_{\mathbf{X}_2}(\mathbf{x}) \equiv P(\mathbf{X}_2 = \mathbf{x}) = \sum_{\mathbf{x}_1 \in \mathbb{R}^{k_1}} f_{\mathbf{X}_1,\mathbf{X}_2}(\mathbf{x}_1, \mathbf{x}).$$

- **Definition 5** *A function* $f(\mathbf{x}_1, \mathbf{x}_2)$ *from* $\mathbb{R}^k$ *to* $\mathbb{R}$ *is called a* ***joint probability density function*** *or* ***joint pdf*** *of the continuous random vector* $(\mathbf{X}_1', \mathbf{X}_2')'$ *if, for every* $A \in \mathcal{F}$,

$$P((\mathbf{X}_1', \mathbf{X}_2')' \in A) = \int\int_A f(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2.$$

- By this definition, the **marginal pdfs** are given by

$$f_{\mathbf{X}_1}(\mathbf{x}) = \int_{\mathbf{x}_2 \in \mathbb{R}^{k_2}} f_{\mathbf{X}_1,\mathbf{X}_2}(\mathbf{x}, \mathbf{x}_2) d\mathbf{x}_2,$$
$$f_{\mathbf{X}_2}(\mathbf{x}) = \int_{\mathbf{x}_1 \in \mathbb{R}^{k_1}} f_{\mathbf{X}_1,\mathbf{X}_2}(\mathbf{x}_1, \mathbf{x}) d\mathbf{x}_1.$$

Expectation and Moment Generating Function

- **Definition 6** *Let* $\mathbf{X}$ *be a discrete random vector and let* $Y \equiv g(\mathbf{X})$ *where* g *is a function from* $\mathbb{R}^k$ *to* $\mathbb{R}$. *If*

$$\sum_{\mathbf{x} \in \mathbb{R}^k} |g(\mathbf{x})| f_{\mathbf{X}}(\mathbf{x}) < \infty,$$

*then the **expectation*** $\mathbb{E}(Y)$ *exitst and*

$$\mathbb{E}(Y) \equiv \sum_{\mathbf{x} \in \mathbb{R}^k} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}).$$

Similarly, if $\mathbf{X}$ *be a continuous random vector, then* $\mathbb{E}(Y)$ *exists if*

$$\int_{\mathbb{R}^k} |g(\mathbf{x})| f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} < \infty,$$

and

$$\mathbb{E}(Y) \equiv \int_{\mathbb{R}^k} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}.$$

- $\mathbb{E}$ is a linear operator. That is, Let $Y_1 = g_1(\mathbf{X})$ and $Y_2 = g_2(\mathbf{X})$ be random variables whose expectation exist. Then for any real numbers a_1 and a_2,

$$\mathbb{E}(a_1 Y_1 + a_2 Y_2) = a_1 \mathbb{E}(Y_1) + a_2 \mathbb{E}(Y_2).$$

- The expectation of a random vector exists if and only if the expectations of all its components do, and it equals the vector consisting of the components' expectations.

- **Definition 7** *Let* $\mathbf{X}$ *be a random vector. If* $\mathbb{E}\left(e^{\mathbf{t}'\mathbf{X}}\right)$ *exists for* $\mathbf{t}$ *such that* $|t_j| < h_j$ *where* $h_j > 0$ *for* $j = 1, \ldots, k$. *Then the following function with respect to* $\mathbf{t}$*:*

$$M_{\mathbf{X}}(\mathbf{t}) \equiv \mathbb{E}\left(e^{\mathbf{t}'\mathbf{X}}\right)$$

*is called the **moment generating function (mgf)** of* $\mathbf{X}$.

- The mgf, if exists, uniquely determines the distribution of the random vector.

- The mgf of a subvector is easy to find from that of the longer vector:

$$M_{\mathbf{X}_1}(\mathbf{t}_1) = M_{\mathbf{X}_1,\mathbf{X}_2}(\mathbf{t}_1, 0).$$

- Take the bivariate random vector as example, the moments can be generated from the mgf as

$$\mathbb{E}(X_1^{n_1} X_2^{n_2}) = \frac{\partial^{n_1+n_2} M_X(t_1, t_2)}{\partial t_1^{n_1} t_2^{n_2}}\bigg|_{t_1=t_2=0.}$$

Conditional Distributions

- **Definition 8** *Let $(\mathbf{X}_1', \mathbf{X}_2')'$ be a discrete random vector with joint pmf $f(\mathbf{x}_1, \mathbf{x}_2)$ and marginal pmfs $f_{\mathbf{X}_1}(\mathbf{x}_1)$ and $f_{\mathbf{X}_2}(\mathbf{x}_2)$. For any $\mathbf{x}_1$ such that $f_{\mathbf{X}_1}(\mathbf{x}_1) > 0$, the* ***conditional pmf of*** $\mathbf{X}_2$ ***given*** $\mathbf{X}_1 = \mathbf{x}_1$ *is the function of $\mathbf{x}_2$ denoted by $f(\mathbf{x}_2|\mathbf{x}_1)$, and defined by*

$$f(\mathbf{x}_2|\mathbf{x}_1) \equiv P(\mathbf{X}_2 = \mathbf{x}_2|\mathbf{X}_1 = \mathbf{x}_1) = \frac{f(\mathbf{x}_1, \mathbf{x}_2)}{f(\mathbf{x}_1)}.$$

  The conditional pmf of $\mathbf{X}_1$ given $\mathbf{X}_2 = \mathbf{x}_2$ is defined symmetrically.

- The conditional pmf is indeed a pmf, in that it is non-negative and

$$\sum_{\mathbf{x}_2 \in \mathbb{R}^{k_2}} f(\mathbf{x}_2|\mathbf{x}_1) = \mathbf{x}_1) = \sum_{\mathbf{x}_2 \in \mathbb{R}^{k_2}} \frac{f(\mathbf{x}_1, \mathbf{x}_2)}{f(\mathbf{x}_1)} = \frac{\sum_{\mathbf{x}_2 \in \mathbb{R}^{k_2}} f(\mathbf{x}_1, \mathbf{x}_2)}{f(\mathbf{x}_1)} = \frac{f(\mathbf{x}_1)}{f(\mathbf{x}_1)} = 1.$$

- **Definition 9** *Let $(\mathbf{X}_1', \mathbf{X}_2')'$ be a continuous random vector with joint pdf $f(\mathbf{x}_1, \mathbf{x}_2)$ and marginal pdfs $f_{\mathbf{X}_1}(\mathbf{x}_1)$ and $f_{\mathbf{X}_2}(\mathbf{x}_2)$. For any $\mathbf{x}_1$ such that $f_{\mathbf{X}_1}(\mathbf{x}_1) > 0$, the* ***conditional pdf of*** $\mathbf{X}_2$ ***given*** $\mathbf{X}_1 = \mathbf{x}_1$ *is the function of $\mathbf{x}_2$ denoted by $f(\mathbf{x}_2|\mathbf{x}_1)$, and defined by*

$$f(\mathbf{x}_2|\mathbf{x}_1) \equiv \frac{f(\mathbf{x}_1, \mathbf{x}_2)}{f_{\mathbf{X}_1}(\mathbf{x}_1)}.$$

  The conditional pdf of $\mathbf{X}_1$ given $\mathbf{X}_2 = \mathbf{x}_2$ is defined symmetrically.

- You may verify that $f(\mathbf{x}_2|\mathbf{x}_1)$ is indeed a pdf since it integrates to one and is non-negative.

- **Definition 10** *If $g(\mathbf{x}_2)$ is a function of $\mathbf{X}_2$, then the* ***conditional expectation of*** $g(\mathbf{X}_2)$ ***given*** $\mathbf{X}_1 = \mathbf{x}_1$ *is denoted by* $\mathbb{E}(g(\mathbf{X}_2)|\mathbf{X}_1 = \mathbf{x}_1)$ *and is given by*

$$\mathbb{E}(g(\mathbf{X}_2)|\mathbf{X}_1 = \mathbf{x}_1) \equiv \sum_{\mathbf{x}_2 \in \mathbb{R}^{k_2}} g(\mathbf{x}_2) f(\mathbf{x}_2|\mathbf{x}_1),$$

or

$$\mathbb{E}(g(\mathbf{X}_2)|\mathbf{X}_1 = \mathbf{x}_1) \equiv \int_{\mathbf{x}_2 \in \mathbb{R}^{k_2}} g(\mathbf{x}_2) f(\mathbf{x}_2|\mathbf{x}_1) d\mathbf{x}_2.$$

- **Theorem 2** ***(Law of iterated expectations)***

$$\mathbb{E}[g(\mathbf{X}_1, \mathbf{X}_2)] = \mathbb{E}_{\mathbf{X}_1}\{\mathbb{E}[g(\mathbf{X}_1, \mathbf{X}_2)|\mathbf{X}_1]\} = \mathbb{E}_{\mathbf{X}_2}\{\mathbb{E}[g(\mathbf{X}_1, \mathbf{X}_2)|\mathbf{X}_2]\}.$$

- **Theorem 3** ***(Law of iterated variances)***

$$var[g(\mathbf{X}_1, \mathbf{X}_2)] = \mathbb{E}_{\mathbf{X}_1}\{var[g(\mathbf{X}_1, \mathbf{X}_2)|\mathbf{X}_1]\} + var_{\mathbf{X}_1}\{\mathbb{E}[g(\mathbf{X}_1, \mathbf{X}_2)|\mathbf{X}_1]\}.$$

Proof. *Law of iterated variances can be proved using the definition of the variance and the law of iterated expectation. For succinct notation, let $Y = g(\mathbf{X}_1, \mathbf{X}_2)$.*

$$\begin{aligned}
var(Y) &= \mathbb{E}(Y^2) - [\mathbb{E}(Y)]^2 \\
&= \mathbb{E}_{\mathbf{X}_1}[\mathbb{E}(Y^2|\mathbf{X}_1)] - \{\mathbb{E}_{\mathbf{X}_1}[\mathbb{E}(Y|\mathbf{X}_1)]\}^2 \\
&= \mathbb{E}_{\mathbf{X}_1}\{var(Y|\mathbf{X}_1) + [\mathbb{E}(Y|\mathbf{X}_1)]^2\} - \{\mathbb{E}_{\mathbf{X}_1}[\mathbb{E}(Y|\mathbf{X}_1)]\}^2 \\
&= \mathbb{E}_{\mathbf{X}_1}[var(Y|\mathbf{X}_1)] + \mathbb{E}_{\mathbf{X}_1}\{[\mathbb{E}(Y|\mathbf{X}_1)]^2\} - \{\mathbb{E}_{\mathbf{X}_1}[\mathbb{E}(Y|\mathbf{X}_1)]\}^2 \\
&= \mathbb{E}_{\mathbf{X}_1}[var(Y|\mathbf{X}_1)] + var_{\mathbf{X}_1}[\mathbb{E}(Y|\mathbf{X}_1)].
\end{aligned}$$

- This theorem implies that $var[\mathbb{E}(Y|\mathbf{X}_1)] \leq var(Y)$; that is, the variance of a random variable becomes (weakly) smaller when controlling some random variables.

Independent Random Variables

- For ease of discussion, let's first focus on bivariate random vectors (X_1, X_2) with joint pmf or pdf $f(x_1, x_2)$ and marginals $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$.

- **Definition 11** ***(Independence)*** *X_1 and X_2 are* ***independent*** *if and only if, for every $x_1 \in \mathbb{R}$ and $x_2 \in \mathbb{R}$,*

$$f(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2).$$

  Random variables that are not independent are said to be ***dependent****.*

- **Theorem 4** *Let X_1 and X_2 have supports S_1 and S_2, respectively. Then X_1 and X_2 are independent if and only if*

$$f(x_1, x_2) = g(x_1)h(x_2),$$

  where $g(x_1) > 0$ for $x_1 \in S_1$ and zero elsewhere, and $h(x_2) > 0$ for $x_2 \in S_2$ and zero elsewhere.

- Note that in the above theorem, the support S_1 of X_1 cannot depend on the value of X_2, otherwise X_1 and X_2 are not independent. Similar is true vice versa.

- **Theorem 5** *Let $F(x_1, x_2)$, $F_{X_1}(x_1)$ and $F_{X_2}(x_2)$ be the joint and marginal cdfs of X_1 and X_2. Then X_1 and X_2 are independent if and only if, for any $x_1 \in \mathbb{R}$ and $x_2 \in \mathbb{R}$,*

$$F(x_1, x_2) = F_{X_1}(x_1) F_{X_2}(x_2).$$

- **Theorem 6** *A corollary of the above theorem is: X_1 and X_2 are independent if and only if for any constants $a < b$ and $c < d$,*

$$P(a < X_1 \leq b, c < X_2 \leq d) = P(a < X_1 \leq b) P(c < X_2 \leq d).$$

- **Theorem 7** *Suppose the moment generating function $M(t_1, t_2)$ exists for (X_1, X_2). Then X_1 and X_2 are independent if and only if*

$$M(t_1, t_2) = M(t_1, 0)M(0, t_2) \equiv M_{X_1}(t_1) M_{X_2}(t_2).$$

- These theorems provide equivalent definitions for independent random variables. The following theorems are consequences of independence (the converse is not necessarily true).

- **Theorem 8** *If X_1 and X_2 are independent random variables, and $\mathbb{E}[u(X_1)]$ and $\mathbb{E}[v(X_2)]$ exist for some functions u and v. Then*

$$\mathbb{E}[u(X_1)v(X_2)] = \mathbb{E}[u(X_1)]\mathbb{E}[v(X_2)].$$

- **Theorem 9** *If X_1 and X_2 are independent random variables, and if the conditional pmf or pdf exits. Then*

$$f_{X_1|X_2}(x_1|x_2) = f_{X_1}(x_1) \text{ and } f_{X_2|X_1}(x_2|x_1) = f_{X_2}(x_2).$$

- If X_1 and X_2 are independent random variables, then for any subsets A and B of $\mathbb{R}$ such that they are in the respective σ-algebra,

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B);$$

  that is, $\{X \in A\}$ and $\{Y \in B\}$ are independent events.

- All of these theorems can be generalized to more than two random variables with little modification to accommodate the dimension. But conceptually, independence of more than two random variables requires more than the independence of two random variables.

- For more than two random variables $X_1, \ldots, X_k$ ($k > 2$), they are said to be **pairwise independent** if any pair of random variables from them are independent.

- A set of random variables $X_1, \ldots, X_k$ ($k > 2$) are said to be **mutually independent** if for any subsets $X_{k_1}, \ldots, X_{k_n}$ of the random variables and any constants $a_{k_1}, \ldots, a_{k_n}$ ($n \leq k$), the events $\{X_{k_1} \leq a_{k_1}\}, \ldots, \{X_{k_n} \leq a_{k_n}\}$ are mutually independent events.

- If $X_1, \ldots, X_k$ are mutually independent, and let $Z \equiv X_1 + \cdots + X_k$. Then

$$M_Z(t) = M_{X_1}(t) \cdot \cdots \cdot M_{X_k}(t).$$

- Recall that the mgf of $Z \equiv aX + b$ (a, b are constants) is $M_Z(t) = e^{bt}M_X(at)$. Generalizing to more random variables: let $X_1, \ldots, X_k$ are mutually independent random variables with mgfs, and let $a_1, \ldots, a_k$ and $b_1, \ldots, b_k$ be constants. Then the mgf of the sum $Z \equiv (a_1X_1 + b_1) + \cdots + (a_kX_k + b_k)$ is

$$M_Z(t) = e^{t\sum_i b_i} M_{X_1}(a_1t) \cdot \cdots \cdot M_{X_k}(a_kt).$$

2 Bivariate Transformation

Bivariate Transformation

- First review how to find the pdf of a monotone function of a univariate continuous random variable. Let X have pdf $f_X(x)$ and let $Y \equiv g(X)$ where g is a monotone function. Let the sample spaces be

$$\mathcal{X} \equiv \{x : f_X(x) > 0\} \text{ and } \mathcal{Y} \equiv \{y : y = g(x) \text{ for some } x \in \mathcal{X}\}.$$

  Suppose that $f_X(x)$ is continuous on $\mathcal{X}$ and that $g^{-1}(y)$ has a continuous derivative on $\mathcal{Y}$. Then the pdf of Y is given by

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| = f_X(g^{-1}(y)) \left| \frac{1}{dy/dx} \right|,$$

  for $y \in \mathcal{Y}$ and zero otherwise.

- **Theorem 10** ***(Pdf of bivariate transformation)*** *Let (X, Y) be a bivariate random vector with joint pdf $f_{X,Y}(x, y)$, and let $U \equiv g_1(X, Y)$ and $V \equiv g_2(X, Y)$ for some functions g_1 and g_2. Let*

$$\mathcal{A} \equiv \{(x, y) : f_{X,Y}(x, y) > 0\},$$

  and

$$\mathcal{B} \equiv \{(u, v) : u = g_1(x, y), v = g_2(x, y) \textit{ for some } (x, y) \in \mathcal{A}\}.$$

  Suppose g_1 and g_2 are one-to-one and onto transformations, then there are unique inverse transformations h_1 and h_2 such that $x = h_1(u, v)$ and $y = h_2(u, v)$. The ***Jacobian of the transformation****, which is the determinant of the partial derivative matrix, is defined as*

$$J \equiv \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u},$$

  where $\frac{\partial x}{\partial u} = \frac{\partial h_1(u,v)}{\partial u}$, $\frac{\partial x}{\partial v} = \frac{\partial h_1(u,v)}{\partial v}$, $\frac{\partial y}{\partial u} = \frac{\partial h_2(u,v)}{\partial u}$ and $\frac{\partial y}{\partial v} = \frac{\partial h_2(u,v)}{\partial v}$. Assume that J is not identically zero on $\mathcal{B}$. Then the joint pdf of (U, V) is given by

$$f_{U,V}(u, v) = f_{X,Y}(h_1(u, v), h_2(u, v))|J|,$$

  for $(u, v) \in \mathcal{B}$ and zero otherwise.

- This formula can be generalized to more than two random variables with little modification.

- **Example 1** ***(Sum and difference of normal variables)*** *Let X and Y be independent, standard normal random variables, and let $U = X + Y$ and $V = X - Y$. The joint pdf of (X, Y) is*

$$f_{X,Y}(x, y) = \frac{1}{2\pi} \exp\left(-\frac{x^2}{2}\right) \exp\left(-\frac{y^2}{2}\right),$$

for $(x, y) \in \mathbb{R}^2$. Note that $\mathcal{A} = \mathcal{B} = \mathbb{R}^2$. It is easy to solve out (x, y) given (u, v):

$$x \equiv h_1(u, v) = \frac{u+v}{2} \text{ and } y \equiv h_2(u, v) = \frac{u-v}{2}.$$

Then the Jacobian:

$$J \equiv \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}.$$

Using the formula in Theorem 10, we get

$$\begin{aligned} f_{U,V}(u, v) &= f_{X,Y}(h_1(u, v), h_2(u, v))|J| \\ &= f_{X,Y}\left(\frac{u+v}{2}, \frac{u-v}{2}\right)/2 \\ &= \frac{1}{2\pi} \exp\left(-\frac{(u+v)^2/2^2}{2}\right) \exp\left(-\frac{(u-v)^2/2^2}{2}\right)/2 \\ &= \frac{1}{2\sqrt{\pi}} \exp\left(-\frac{u^2}{4}\right) \frac{1}{2\sqrt{\pi}} \exp\left(-\frac{v^2}{4}\right). \end{aligned}$$

This implies that U and V are independent, and they both have distributions $\mathcal{N}(0, 2)$, which confirms our intuition from basic statistics.

- **Example 2** ***(Convolution)*** *In econometrics, we sometimes need to find the pdf of the sum $Z \equiv X + Y$ (convolution). Let $U = Z = X + Y$ and $V = Y$ (or let $U = X$ and $V = Z = X + Y$), using the bivariate transformation formula and the relationship between the joint and marginal pdf, it is easy to verify that*

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(z - y, y)dy = \int_{-\infty}^{\infty} f_{X,Y}(x, z - x)dx.$$

3 Exercises

1. Prove Theorem 2.

2. Prove the results in Example 2.

3. Suppose that (X_1, X_2) have the joint pdf

$$f_{X_1,X_2}(x_1, x_2) = e^{-(x_1+x_2)},$$

   for $x_1 > 0, x_2 > 0$ and zero otherwise. Let $Y_1 \equiv X_1 + X_2$ and $Y_2 \equiv X_1/(X_1 + X_2)$. Answer the following questions:

   (a) What is the joint pdf of (Y_1, Y_2)?

   (b) Are Y_1 and Y_2 independent? Justify your answer.

   (c) What is the probability that $Y_1 \leq 1$ given that $Y_2 \leq 0.5$?

   (d) What is the cdf of Y_1?

   (e) What is the moment generating function (mgf) of (Y_1, Y_2)?