Short Guides to Microeconometrics
Fall 2023

Kurt Schmidheiny
Universität Basel

Elements of Probability Theory

Contents

1 Random Variables and Distributions **2**
1.1 Univariate Random Variables and Distributions 2
1.2 Bivariate Random Variables and Distributions 3

2 Moments **5**
2.1 Expected Value or Mean 5
2.2 Variance and Standard Deviation 6
2.3 Higher order Moments 7
2.4 Covariance and Correlation 7
2.5 Conditional Expectation and Variance 8

3 Random Vectors and Random Matrices **9**

4 Important Distributions **10**
4.1 Univariate Normal Distribution 10
4.2 Bivariate Normal Distribution 10
4.3 Multivariate Normal Distribution 11

1 Random Variables and Distributions

A *random variable* is a variable whose values are determined by a probability distribution. This is a casual way of defining random variables which is sufficient for our level of analysis. For more advanced probability theory, a random variable will be defined as a real-valued function over some probability space.

In section 1 to 3, a random variable is denoted by capital letters, e.g. X, whereas its realizations are denoted by small letters, e.g. x.

1.1 Univariate Random Variables and Distributions

A *univariate discrete random variable* is a variable that takes a countable number K of real numbers with certain probabilities. The probability that the random variable X takes the value x_k among the K possible realizations is given by the *probability distribution*

$$P(X = x_k) = P(x_k) = p_k$$

with $k = 1, 2, ..., K$. K may be ∞ in some cases. This can also be written as

$$P(x_k) = \begin{cases} p_1 & \text{if} \quad X = x_1 \\ p_2 & \text{if} \quad X = x_2 \\ \vdots & \\ p_K & \text{if} \quad X = x_K \end{cases}$$

Note that

$$\sum_{k=1}^{K} p_k = 1.$$

A *univariate continuous random variable* is a variable that takes a continuum of values in the real line. The distribution of a continuous random variable X can be characterized by a *density function* or *probability* *density function (pdf)* $f(x)$. The nonnegative function $f(x)$ is such that

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x)dx.$$

defines the probability that X takes a value in the interval $[x_1, x_2]$. Note that there is no chance that X takes exactly the value x, $P(X = x) = 0$. The probability that X takes any value on the real line is

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

The distribution of a univariate random variable X is alternatively described by the *cumulative distribution function (cdf)*

$$F(x) = P(X < x).$$

The cdf of a discrete random variable X is

$$F(x) = \sum_{x_k \leq x} P(X = x_k) = \sum_{x_k \leq x} p_k.$$

and of a continuous random variable X

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

$F(x)$ has the following properties:

- $F(x)$ is monotonically nondecreasing
- $F(-\infty) = 0$ and $F(\infty) = 1$.
- $F(x)$ is continuous to the left

1.2 Bivariate Random Variables and Distributions

A *bivariate continuous random variable* is a variable that takes a continuum of values in the plane. The distribution of a bivariate continuous random variable (X, Y) can be characterized by a *joint density function* or *joint probability density function*, $f(x, y)$. The nonnegative function $f(x, y)$ is such that

$$P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f(x, y)dydx$$

defines the probability that X and Y take values in the interval $[x_1, x_2]$ and $[y_1, y_2]$, respectively. Note that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dydx = 1.$$

The *marginal density function* or *marginal probability density function* is given by

$$f(x) = \int_{-\infty}^{\infty} f(x, y)dy$$

such that

$$P(x_1 \leq X \leq x_2) = P(x_1 \leq X \leq x_2, -\infty \leq Y \leq \infty) = \int_{x_1}^{x_2} f(x)dx.$$

The *conditional density function* or *conditional probability density function* with respect to the event $\{Y = y\}$ is given by

$$f(y|x) = \frac{f(x, y)}{f(x)}$$

provided that $f(x) > 0$. Note that

$$\int_{-\infty}^{\infty} f(y|x)dy = 1.$$

Two random variables X and Y are called *independent*, if and only if

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

If X and Y are *independent*, then:

- $f(y|x) = f(y)$
- $P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = P(x_1 \leq X \leq x_2) \cdot P(y_1 \leq Y \leq y_2)$

More generally, if a finite set of n continuous random variables $X_1, X_2, X_3, ..., X_n$ are mutually independent, then

$$f(x_1, x_2, x_3, ..., x_n) = f(x_1) \cdot f(x_2) \cdot f(x_3) \cdot ... \cdot f(x_n).$$

2 Moments

2.1 Expected Value or Mean

The *expected value* or *mean* of a *discrete* random variable with probability distribution $P(x_k)$ and $k = 1, 2, ..., K$ is defined as

$$\mathrm{E}[X] = \sum_{k=1}^{K} x_k P(x_k)$$

if the series converges absolutely.

The *expected value* or *mean* of a *continuous* univariate random variable with density function $f(x)$ is defined as

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x) dx$$

if the integral exists.

For a random variable Z which is a continuous function ϕ of a discrete random variable X, we have:

$$\mathrm{E}[Z] = \mathrm{E}[\phi(X)] = \sum_{k=1}^{K} \phi(x_k) P(x_k)$$

For a random variable Z which is a continuous function ϕ of the continuous random variables X and Y, we have:

$$\mathrm{E}[Z] = \mathrm{E}[\phi(X)] = \int_{-\infty}^{\infty} \phi(x) f(x) dx$$

$$\mathrm{E}[Z] = \mathrm{E}[\phi(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(x,y) f(x,y) dx\, dy$$

The following rules hold in general, i.e. for discrete, continuous and mixed types of random variables:

- $\mathrm{E}[\alpha] = \alpha$
- $\mathrm{E}[\alpha X + \beta Y] = \alpha \mathrm{E}[X] + \beta \mathrm{E}[Y]$
- $\mathrm{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$
- $\mathrm{E}[X\,Y] = \mathrm{E}[X]\,\mathrm{E}[Y]$ if X and Y are independent

where $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}$ are constants.

2.2 Variance and Standard Deviation

The variance of a univariate random variable X is defined as

$$\mathrm{V}[X] = \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = E[X^2] - (\mathrm{E}[X])^2$$

The variance has the following properties:

- $\mathrm{V}[X] \geq 0$
- $\mathrm{V}[X] = 0$ if and only if $X = \mathrm{E}[X]$

The following rules hold in general, i.e. for discrete, continuous and mixed types of random variables:

- $\mathrm{V}[\alpha X + \beta] = \alpha^2 \mathrm{V}[X]$
- $\mathrm{V}[X + Y] = \mathrm{V}[X] + \mathrm{V}[Y] + 2\mathrm{Cov}[X, Y]$
- $\mathrm{V}[X - Y] = \mathrm{V}[X] + \mathrm{V}[Y] - 2\mathrm{Cov}[X, Y]$
- $\mathrm{V}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{V}[X_i]$ if X_i and X_j independent for all $i \neq j$

where $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}$ are constants.

Instead of the variance, one often considers the *standard deviation*

$$\sigma_X = \sqrt{\mathrm{V}X}.$$

2.3 Higher order Moments

The j-th moment around zero is defined as

$$E\left[(X - \mathrm{E}[X])^j\right].$$

2.4 Covariance and Correlation

The *Covariance* between two random variables X and Y is defined as:

$$\begin{aligned}\mathrm{Cov}[X,Y] &= \mathrm{E}\big[(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])\big] \\ &= \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] \\ &= \mathrm{E}\big[(X-\mathrm{E}[X])Y\big] = \mathrm{E}\big[X(Y-\mathrm{E}[Y])\big]\end{aligned}$$

The following rules hold in general, i.e. for discrete, continuous and mixed types of random variables:

- $\mathrm{Cov}[\alpha X + \gamma, \beta Y + \mu] = \alpha\beta \mathrm{Cov}[X,Y]$
- $\mathrm{Cov}[X_1 + X_2, Y_1 + Y_2]$
  $= \mathrm{Cov}[X_1,Y_1] + \mathrm{Cov}[X_1,Y_2] + \mathrm{Cov}[X_2,Y_1] + \mathrm{Cov}[X_2,Y_2]$
- $\mathrm{Cov}[X,Y] = 0 \quad$ if X and Y are independent

where $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}$, $\gamma \in \mathbb{R}$ and $\mu \in \mathbb{R}$ are constants.

The *correlation coefficient* between two random variables X and Y is defined as:

$$\rho_{X,Y} = \frac{\mathrm{Cov}[X,Y]}{\sigma_X \sigma_Y}$$

where σ_X and σ_Y denote the corresponding standard deviations. The correlation coefficient has the following property:

- $-1 \leq \rho_{X,Y} \leq 1$

The following rule holds:

- $\rho_{\alpha X+\gamma, \beta Y+\mu} = \rho_{X,Y}$
- $\rho_{X,Y} = 0 \quad$ if X and Y are independent

where $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}$ are constants.

We say that

- X and Y are *uncorrelated* if $\rho = 0$
- X and Y are *positively correlated* if $\rho > 0$
- X and Y are *negatively correlated* if $\rho < 0$

2.5 Conditional Expectation and Variance

Let (X,Y) be a bivariate discrete random variable and $P(y_k|X)$ the conditional probability of $Y = y_k$ given X. Then the *conditional expected value* or *conditional mean* of Y given X is

$$\mathrm{E}[Y|X] = \mathrm{E}_{Y|X}[Y] = \sum_{k=1}^{K} y_k P(y_k|X).$$

Let (X,Y) be a bivariate continuous random variable and $f(y|x)$ the conditional density of Y given X. Then the *conditional expected value* or *conditional mean* of Y given X is

$$\mathrm{E}[Y|X] = \mathrm{E}_{Y|X}[Y] = \int_{-\infty}^{\infty} y f(y|X) dy.$$

The *law of iterated means* or *law of iterated expecations* holds in general, i.e. for discrete, continuous or mixed random variables:

$$\mathrm{E}_X\big[\mathrm{E}[Y|X]\big] = \mathrm{E}[Y].$$

The conditional variance of Y given X is given by

$$\mathrm{V}[Y|X] = \mathrm{E}\left[(Y - \mathrm{E}[Y|X])^2 \middle| X\right] = \mathrm{E}\left[Y^2 \middle| X\right] - (\mathrm{E}[Y|X])^2.$$

The *law of total variance* is

$$\mathrm{V}[Y] = \mathrm{E}_X\big[V[Y|X]\big] + \mathrm{V}_X\big[\mathrm{E}[Y|X]\big].$$

3 Random Vectors and Random Matrices

In this section we denote matrices (random or non-random) by bold capital letters, e.g. $\mathbf{X}$ and vectors by small letters, e.g. $\mathbf{x}$.

Let $\mathbf{x} = (x_1, \ldots, x_n)'$ be a $(n \times 1)$-dimensional vector such that each element x_i is a random variable. Let $\mathbf{X}$ be a $(n \times k)$-dimensional matrix such that each element x_{ij} is a random variable. Let $\mathbf{a} = (a_1, \ldots, a_n)'$ be a $n \times 1$-dimensional vector of constants and $\mathbf{A}$ a $(m \times n)$ matrix of constants.

The expectation of a random vector, $\mathrm{E}[\mathbf{x}]$, and of a random matrix, $\mathrm{E}[\mathbf{X}]$, summarize the expected values of its elements, respectively:

$$\mathrm{E}[\mathbf{x}] = \begin{pmatrix} \mathrm{E}[x_1] \\ \mathrm{E}[x_2] \\ \vdots \\ \mathrm{E}[x_n] \end{pmatrix} \quad \text{and} \quad \mathrm{E}[\mathbf{X}] = \begin{pmatrix} \mathrm{E}[x_{11}] & \mathrm{E}[x_{12}] & \ldots & \mathrm{E}[x_{1k}] \\ \mathrm{E}[x_{21}] & \mathrm{E}[x_{22}] & \ldots & \mathrm{E}[x_{2k}] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}[x_{n1}] & \mathrm{E}[x_{n2}] & \ldots & \mathrm{E}[x_{nk}] \end{pmatrix}.$$

The following rules hold:

- $\mathrm{E}[\mathbf{a}'\mathbf{x}] = \mathbf{a}'\mathrm{E}[\mathbf{x}]$
- $\mathrm{E}[\mathbf{A}\mathbf{x}] = \mathbf{A}\mathrm{E}[\mathbf{x}]$
- $\mathrm{E}[\mathbf{A}\mathbf{X}] = \mathbf{A}\mathrm{E}[\mathbf{X}]$
- $\mathrm{E}[tr(\mathbf{X})] = tr(\mathrm{E}[\mathbf{X}])$ for $\mathbf{X}$ a quadratic matrix

The variance-Covariance matrix of a random vector, $\mathrm{V}(\mathbf{x})$, summarizes all variances and Covariances of its elements:

$$\mathrm{V}[\mathbf{x}] = \mathrm{E}\big[(\mathbf{x} - \mathrm{E}[\mathbf{x}])\,(\mathbf{x} - \mathrm{E}[\mathbf{x}])'\big] = \mathrm{E}[\mathbf{x}\mathbf{x}'] - (\mathrm{E}[\mathbf{x}])([\mathrm{E}[\mathbf{x}])'$$
$$= \begin{pmatrix} \mathrm{V}[x_1] & \mathrm{Cov}[x_1, x_2] & \ldots & \mathrm{Cov}[x_1, x_n] \\ \mathrm{Cov}[x_2, x_1] & \mathrm{V}[x_2] & \ldots & \mathrm{Cov}[x_2, x_n] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}[x_n, x_1] & \mathrm{Cov}[x_n, x_2] & \ldots & \mathrm{V}[x_n] \end{pmatrix}.$$

The following rules hold:

- $\mathrm{V}[\mathbf{a}'\mathbf{x}] = \mathbf{a}'\,\mathrm{V}[\mathbf{x}]\,\mathbf{a}$
- $\mathrm{V}[\mathbf{A}\mathbf{x}] = \mathbf{A}\,\mathrm{V}[\mathbf{x}]\,\mathbf{A}'$

where the $(m \times n)$ dimensional matrix $\mathbf{A}$ with $m \leq n$ has full row rank.

If the variance-Covariance matrix $\mathrm{V}[\mathbf{x}]$ is positive definite (p.d.) then all random elements and all linear combinations of its random elements have strictly positive variance:

$$V[\mathbf{a}'\mathbf{x}] = \mathbf{a}'\,V[\mathbf{x}]\,\mathbf{a} > 0 \text{ for all } \mathbf{a} \neq 0.$$

4 Important Distributions

4.1 Univariate Normal Distribution

The density of the *univariate normal distribution* is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

The normal distribution is characterized by the two parameters μ and σ. The mean of the normal distribution is $\mathrm{E}[X] = \mu$ and the variance $\mathrm{V}[X] = \sigma^2$. We write $X \sim \mathrm{N}(\mu, \sigma^2)$.

The univariate normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$ is called the *standard normal distribution* $\mathrm{N}(0, 1)$.

4.2 Bivariate Normal Distribution

The density of the *bivariate normal distribution* is

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$$
$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)\right]\right\}.$$

If (X, Y) follows a bivariate normal distribution, then:

- The marginal densities $f(x)$ and $f(y)$ are univariate normal.
- The conditional densities $f(x|y)$ and $f(y|x)$ are univariate normal.
- $\mathrm{E}[X] = \mu_X$, $\mathrm{V}[X] = \sigma_X^2$, $\mathrm{E}[Y] = \mu_Y$, $\mathrm{V}[Y] = \sigma_Y^2$.
- The correlation coefficient between X and Y is $\rho_{X,Y} = \rho$.
- $\mathrm{E}[Y|X] = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X)$ and $\mathrm{V}[Y|X] = \sigma_Y^2(1-\rho^2)$.

The above properties characterize the normal distribution. It is the only distribution with all these properties.

Further important properties:

- If (X, Y) follows a bivariate normal distribution, then $aX + bY$ is also normally distributed:

  $$aX + bY \sim \mathrm{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y).$$

  The reverse implication is not true.

- If X and Y are bivariate normally distributed with $\mathrm{Cov}[X, Y] = 0$, then X and Y are independent.

4.3 Multivariate Normal Distribution

Let $\mathbf{x} = (x_1, \ldots, x_n)'$ be a n-dimensional vector such that each element x_i is a random variable. In addition let $\mathrm{E}[\mathbf{x}] = \mu = (\mu_1, \ldots, \mu_n)$ and $\mathrm{V}[\mathbf{x}] = \mathbf{\Sigma}$ with

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \ldots & \sigma_{nn} \end{pmatrix}$$

where $\sigma_{ij} = \mathrm{Cov}[x_i, x_j]$.

A n-dimensional random variable $\mathbf{x}$ is *multivariate normally distributed* with mean μ and variance-Covariance matrix $\mathbf{\Sigma}$, $\mathbf{x} \sim \mathrm{N}(\mu, \mathbf{\Sigma})$ if its density is:

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

Let $\mathbf{x} \sim \mathrm{N}(\mu, \mathbf{\Sigma})$ and $\mathbf{A}$ a $(m \times n)$ matrix with $m \leq n$ and m linearly independent rows then we have

$$\mathbf{A}\mathbf{x} \sim \mathrm{N}(\mathbf{A}\mu, \mathbf{A}\mathbf{\Sigma}\mathbf{A}').$$

References

Amemiya, Takeshi (1994), *Introduction to Statistics and Econometrics*, Cambridge: Harvard University Press.

Hayashi, Fumio (2000), *Econometrics*, Princeton: Princeton University Press. Appendix A.

Stock, James H. and Mark W. Watson (2020), *Introduction to Econometrics*, 4th Global ed., Pearson. Chapter 2.