Review of Probability Concepts

Appendix B

Prepared by Vera Tabakova, East Carolina University

Appendix B: Review of Probability Concepts

- B.1 Random Variables
- B.2 Probability Distributions
- B.3 Joint, Marginal and Conditional Probability Distributions
- B.4 Properties of Probability Distributions
- B.5 Some Important Probability Distributions

B.1 Random Variables

- A **random variable** is a variable whose value is unknown until it is observed.
- A **discrete random variable** can take only a limited, or countable, number of values.
- A **continuous random variable** can take any value on an interval.

B.2 Probability Distributions

- The probability of an event is its "limiting relative frequency," or the proportion of time it occurs in the long-run.
- The **probability density function (*pdf*)** for a discrete random variable indicates the probability of each possible value occurring.

$$f(x) = P(X = x)$$

$$f(x_1) + f(x_2) + \cdots + f(x_n) = 1$$

B.2 Probability Distributions

Table B.1 **Probabilities of a College Degree**

College Degree	x	$f(x)$
No	0	0.73
Yes	1	0.27

B.2 Probability Distributions

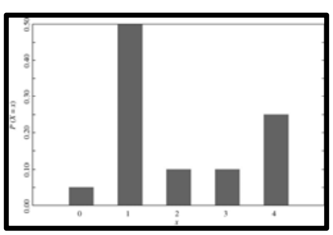

Figure B.1 College Employment Probabilities

B.2 Probability Distributions

- The **cumulative distribution function (*cdf*)** is an alternative way to represent probabilities. The *cdf* of the random variable *X*, denoted *F*(x), gives the probability that *X* is less than or equal to a specific value *x*.

$$F(x) = P(X \le x)$$

B.2 Probability Distributions

Table B.2 **A *pdf* and *cdf***

x	$f(x)$	$F(x)$
0	0.05	0.05
1	0.50	0.55
2	0.10	0.65
3	0.10	0.75
4	0.25	1.00

B.2 Probability Distributions

- For example, a **binomial random variable** *X* is the number of successes in *n* independent trials of identical experiments with probability of success *p*.

$$P(X = x) = f(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{(B.1)}$$

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \text{ where } n! = n\big((n-1)\big)(n-2)\cdots(2)(1)$$

B.2 Probability Distributions

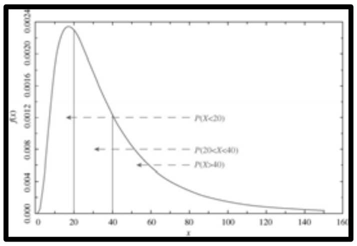

Figure B.2 PDF of a continuous random variable

Principles of Econometrics, 3rd Edition *Slide B-10*

B.2 Probability Distributions

$$P\left(20 \le X \le 40\right) = \int_{20}^{40} f\left(x\right)dx = .355$$

$$P\left(X \le x\right) = \int_{-\infty}^{x} f\left(t\right)dt = F\left(x\right)$$

$$P\left(20 \le X \le 40\right) = F(40) - F(20) = .649 - .294 = .355$$

Principles of Econometrics, 3rd Edition *Slide B-11*

B.3 Joint, Marginal and Conditional Probability Distributions

$$X = \begin{cases} 1 & \text{high school diploma or less} \\ 2 & \text{some college} \\ 3 & \text{four year college degree} \\ 4 & \text{advanced degree} \end{cases}$$

$$Y = \begin{cases} 0 & \text{if had no money earnings in 2002} \\ 1 & \text{if had positive money earnings in 2002} \end{cases}$$

Principles of Econometrics, 3rd Edition *Slide B-12*

B.3 Joint, Marginal and Conditional Probability Distributions

Table B.3 Joint Probability Function $f(x,y)$

	x			
y	1	2	3	4
0	0.19	0.06	0.04	0.02
1	0.28	0.19	0.14	0.08

$$\sum_x \sum_y f(x,y) = 1$$

B.3.1 Marginal Distributions

$$f_X(x) = \sum_y f(x,y) \quad \text{for each value } X \text{ can take}$$

$$f_Y(y) = \sum_x f(x,y) \quad \text{for each value } Y \text{ can take} \qquad \text{(B.2)}$$

$$f_Y(y) = \sum_{x=1}^{4} f(x,y) \quad y = 0,1$$

$$f_Y(1) = .19 + .06 + .04 + .02 = .31$$

B.3.1 Marginal Distributions

Table B.4 Marginal Distributions for X and Y

	x				
y	1	2	3	4	$f_Y(y)$
0	0.19	0.06	0.04	0.02	0.31
1	0.28	0.19	0.14	0.08	0.69
$f_X(x)$	0.47	0.25	0.18	0.10	1

B.3.2 Conditional Probability

$$f(y \mid x) = P(Y = y \mid X = x) = \frac{P(Y = y, X = x)}{P(X = x)} = \frac{f(x,y)}{f_X(x)} \quad \text{(B.3)}$$

y	$f(y \mid X = 3)$
0	.04/.18=.22
1	.14/.18=.78

Principles of Econometrics, 3rd Edition *Slide B-16*

B.3.2 Conditional Probability

- Two random variables are **statistically independent** if the conditional probability that $Y = y$ given that $X = x$, is the same as the unconditional probability that $Y = y$.

$$P(Y = y \mid X = x) = P(Y = y) \quad \text{(B.4)}$$

$$f(y \mid x) = \frac{f(x,y)}{f_X(x)} = f_Y(y) \quad \text{(B.5)}$$

$$f(x,y) = f_X(x) f_Y(y) \quad \text{(B.6)}$$

Principles of Econometrics, 3rd Edition *Slide B-17*

B.3.3 A Simple Experiment

Table B.5 **A Population**

1	2	3	4	4
2	3	3	4	4

Principles of Econometrics, 3rd Edition *Slide B-18*

B.3.3 A Simple Experiment

Table B.6 Probability Distribution of Y

Shaded	Y	$f(y)$
No	0	0.6
Yes	1	0.4

Principles of Econometrics, 3rd Edition *Slide B-19*

B.3.3 A Simple Experiment

Table B.7 Probability Distribution of X

X	$f(x)$
1	0.1
2	0.2
3	0.3
4	0.4

Principles of Econometrics, 3rd Edition *Slide B-20*

B.3.3 A Simple Experiment

Table B.8 Joint Probability Function $f(x, y)$

	x			
y	1	2	3	4
0	0	0.1	0.2	0.3
1	0.1	0.1	0.1	0.1

Principles of Econometrics, 3rd Edition *Slide B-21*

B.4 Properties of Probability Distributions

- **B.4.1 Mean, median and mode**

$$E[X] = x_1 P(X = x_1) + x_2 P(X = x_2) + \cdots + x_n P(X = x_n) \quad \text{(B.7)}$$

- For a discrete random variable the expected value is:

$$\begin{aligned} \mu = E[X] &= x_1 f(x_1) + x_2 f(x_2) + \cdots + x_n f(x_n) \\ &= \sum_{i=1}^{n} x_i f(x_i) = \sum_{x} x f(x) \end{aligned} \quad \text{(B.8)}$$

Principles of Econometrics, 3rd Edition *Slide B-22*

B.4.1 Mean, median and mode

- For a continuous random variable the expected value is:

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

The mean has a flaw as a measure of the center of a probability distribution in that it can be pulled by extreme values.

Principles of Econometrics, 3rd Edition *Slide B-23*

B.4.1 Mean, median and mode

- For a continuous distribution the **median** of *X* is the value *m* such that

$$P(X > m) = P(X < m) = .5$$

- In symmetric distributions, like the familiar "bell-shaped curve" of the normal distribution, the mean and median are equal.
- The **mode** is the value of *X* at which the *pdf* is highest.

Principles of Econometrics, 3rd Edition *Slide B-24*

B.4.2 Expected values of functions of a random variable

$$E[g(X)] = \sum_x g(x) f(x) \quad \text{(B.9)}$$

$$E[aX] = aE[X] \quad \text{(B.10)}$$

$$E[g(X)] = \sum g(x) f(x) = \sum axf(x) = a\sum xf(x) = aE(X)$$

Principles of Econometrics, 3rd Edition *Slide B-25*

B.4.2 Expected values of functions of a random variable

$$E[aX + b] = aE[X] + b \quad \text{(B.11)}$$

$$E[g_1(X) + g_2(X)] = E[g_1(X)] + E[g_2(X)] \quad \text{(B.12)}$$

- The **variance** of a discrete or continuous random variable *X* is the expected value of

$$g(X) = [X - E(X)]^2$$

Principles of Econometrics, 3rd Edition *Slide B-26*

B.4.2 Expected values of functions of a random variable

- The variance of a random variable is important in characterizing the scale of measurement, and the spread of the probability distribution.
- Algebraically, letting $E(X) = \mu$,

$$\text{var}(X) = \sigma^2 = E[X - \mu]^2 = E[X^2] - \mu^2 \quad \text{(B.13)}$$

Principles of Econometrics, 3rd Edition *Slide B-27*

B.4.2 Expected values of functions of a random variable

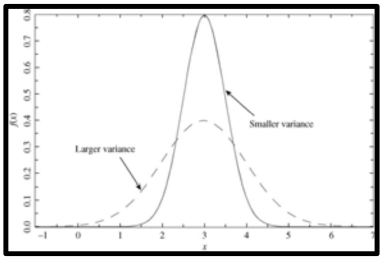

Figure B.3 Distributions with different variances

B.4.2 Expected values of functions of a random variable

$$\text{var}(aX+b) = a^2\,\text{var}(X) \qquad \text{(B.14)}$$

$$\text{var}(aX+b) = E\left[aX+b-E\left(aX+b\right)\right]^2 = E\left[aX+b-a\mu-b\right]^2$$

$$= E\left[a\left(X-\mu\right)\right]^2 = a^2E\left(X-\mu\right)^2 = a^2\,\text{var}\left(X\right)$$

B.4.2 Expected values of functions of a random variable

$$skewness = \frac{E\left[\left(X-\mu\right)^3\right]}{\sigma^3}$$

$$kurtosis = \frac{E\left[\left(X-\mu\right)^4\right]}{\sigma^4}$$

B.4.3 Expected values of several random variables

$$E[g(X,Y)] = \sum_x \sum_y g(x,y) f(x,y) \quad \text{(B.15)}$$

$$E(X+Y) = E(X) + E(Y) \quad \text{(B.16)}$$

$$E(X+Y) = \sum_x \sum_y (x+y) f(x,y) = \sum_x \sum_y xf(x,y) + \sum_x \sum_y yf(x,y)$$
$$= \sum_x x \sum_y f(x,y) + \sum_y y \sum_x f(x,y) = \sum_x xf(x) + \sum_y yf(y)$$
$$= E(X) + E(Y)$$

Principles of Econometrics, 3rd Edition — *Slide B-31*

B.4.3 Expected values of several random variables

$$E(aX + bY + c) = aE(X) + bE(Y) + c \quad \text{(B.17)}$$

$$E[XY] = E[g(X,Y)] = \sum_x \sum_y xyf(x,y) = \sum_x \sum_y xyf(x)f(y)$$
$$= \sum_x xf(x) \sum_y yf(y) = E(X)E(Y) \text{ if } X \text{ and } Y \text{ are independent.}$$

$$g(X,Y) = (X - \mu_X)(Y - \mu_Y) \quad \text{(B.18)}$$

Principles of Econometrics, 3rd Edition — *Slide B-32*

B.4.3 Expected values of several random variables

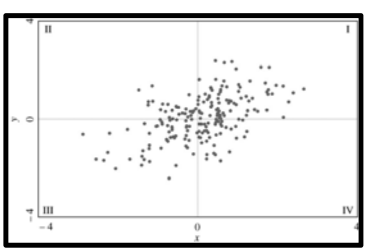

Figure B.4 Correlated data

Principles of Econometrics, 3rd Edition — *Slide B-33*

B.4.3 Expected values of several random variables

$$\text{cov}(X,Y) = \sigma_{XY} = E\left[(X-\mu_X)(Y-\mu_Y)\right] = E[XY] - \mu_X\mu_Y \quad \text{(B.19)}$$

$$\rho = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} \quad \text{(B.20)}$$

- If *X* and *Y* are independent random variables then the covariance and correlation between them are zero. The converse of this relationship is *not* true.

B.4.3 Expected values of several random variables

- If *a* and *b* are constants then:

$$\text{var}\left[aX + bY\right] = a^2\,\text{var}(X) + b^2\,\text{var}(Y) + 2ab\,\text{cov}(X,Y) \quad \text{(B.21)}$$

$$\text{var}\left[X + Y\right] = \text{var}(X) + \text{var}(Y) + 2\,\text{cov}(X,Y) \quad \text{(B.22)}$$

$$\text{var}\left[X - Y\right] = \text{var}(X) + \text{var}(Y) - 2\,\text{cov}(X,Y) \quad \text{(B.23)}$$

B.4.3 Expected values of several random variables

- If *X* and *Y* are independent then:

$$\text{var}\left[aX + bY\right] = a^2\,\text{var}(X) + b^2\,\text{var}(Y) \quad \text{(B.24)}$$

$$\text{var}\left(X \pm Y\right) = \text{var}(X) + \text{var}(Y) \quad \text{(B.25)}$$

$$\text{var}\left[X + Y + Z\right] = \text{var}\left[X\right] + \text{var}\left[Y\right] + \text{var}\left[Z\right]$$

B.4.4 The Simple Experiment Again

$$E(X)=\sum_{x=1}^{4} xf(x)=(1\times.1)+(2\times.2)+(3\times.3)+(4\times.4)=3=\mu_X$$

$$\begin{aligned}\sigma_X^2 &= E(X-\mu_X)^2\\ &=\left[(1-3)^2\times.1\right]+\left[(2-3)^2\times.2\right]+\left[(3-3)^2\times.3\right]+\left[(4-3)^2\times.4\right]\\ &=(4\times.1)+(1\times.2)+(0\times.3)+(1\times.4)\\ &=1\end{aligned}$$

Principles of Econometrics, 3rd Edition *Slide B-37*

B.5 Some Important Probability Distributions

- **B.5.1 The Normal Distribution**

- If X is a normally distributed random variable with mean μ and variance σ^2, it can be symbolized as $X \sim N(\mu,\sigma^2)$.

$$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right], \qquad -\infty < x < \infty \qquad \text{(B.26)}$$

Principles of Econometrics, 3rd Edition *Slide B-38*

B.5.1 The Normal Distribution

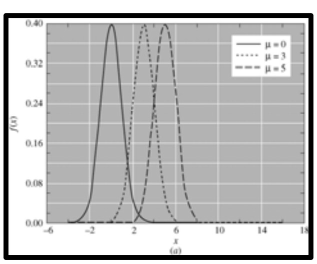

Figure B.5a Normal Probability Density Functions with Means μ and Variance 1

Principles of Econometrics, 3rd Edition *Slide B-39*

B.5.1 The Normal Distribution

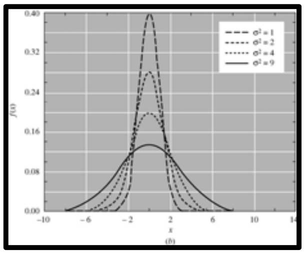

Figure B.5b Normal Probability Density Functions with Mean 0 and Variance σ^2

B.5.1 The Normal Distribution

- A **standard normal random variable** is one that has a normal probability density function with mean 0 and variance 1.

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1) \qquad \text{(B.27)}$$

- The *cdf* for the standardized normal variable *Z* is $\Phi(z) = P(Z \le z)$.

B.5.1 The Normal Distribution

$$P[X \le a] = P\left[\frac{X-\mu}{\sigma} \le \frac{a-\mu}{\sigma}\right] = P\left[Z \le \frac{a-\mu}{\sigma}\right] = \Phi\left(\frac{a-\mu}{\sigma}\right) \qquad \text{(B.28)}$$

$$P[X > a] = P\left[\frac{X-\mu}{\sigma} > \frac{a-\mu}{\sigma}\right] = P\left[Z > \frac{a-\mu}{\sigma}\right] = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right) \qquad \text{(B.29)}$$

$$P[a \le X \le b] = P\left[\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right] = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \qquad \text{(B.30)}$$

B.5.1 The Normal Distribution

- A weighted sum of normal random variables has a normal distribution.

$X_1 \sim N\left(\mu_1, \sigma_1^2\right)$

$X_2 \sim N\left(\mu_2, \sigma_2^2\right)$

$$Y = a_1 X_1 + a_2 X_2 \sim N\left(\mu_Y = a_1\mu_1 + a_2\mu_2, \sigma_Y^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2a_1a_2\sigma_{12}\right) \quad \text{(B.27)}$$

Principles of Econometrics, 3rd Edition *Slide B-43*

B.5.2 The Chi-square Distribution

$$V = Z_1^2 + Z_2^2 + \ldots + Z_m^2 \sim \chi^2_{(m)} \quad \text{(B32)}$$

$$E[V] = E\left[\chi^2_{(m)}\right] = m$$

$$\text{var}[V] = \text{var}\left[\chi^2_{(m)}\right] = 2m \quad \text{(B.33)}$$

Principles of Econometrics, 3rd Edition *Slide B-44*

B.5.2 The Chi-square Distribution

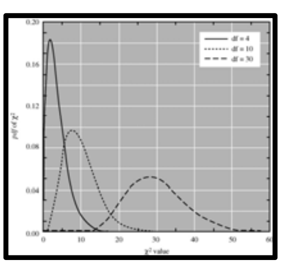

Figure B.6 The chi-square distribution

Principles of Econometrics, 3rd Edition *Slide B-45*

B.5.3 The *t*-Distribution

- A "*t*" random variable (no upper case) is formed by dividing a standard normal random variable $Z \sim N(0,1)$ by the square root of an *independent* chi-square random variable, $V \sim \chi^2_{(m)}$, that has been divided by its degrees of freedom *m*.

$$t = \frac{Z}{\sqrt{V/m}} \sim t_{(m)} \qquad \text{(B.34)}$$

Principles of Econometrics, 3rd Edition *Slide B-46*

B.5.3 The *t*-Distribution

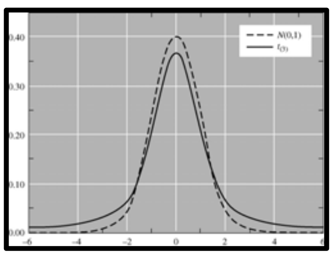

Figure B.7 The standard normal and $t_{(3)}$ probability density functions

Principles of Econometrics, 3rd Edition *Slide B-47*

B.5.4 The *F*-Distribution

- An *F* random variable is formed by the ratio of two independent chi-square random variables that have been divided by their degrees of freedom.

$$F = \frac{V_1/m_1}{V_2/m_2} \sim F_{(m_1,m_2)} \qquad \text{(B.35)}$$

Principles of Econometrics, 3rd Edition *Slide B-48*

B.5.4 The *F*-Distribution

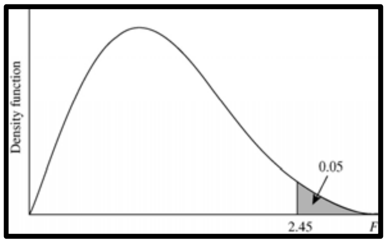

Figure B.8 The probability density function of an *F* random variable

Keywords

- binary variable
- binomial random variable
- *cdf*
- chi-square distribution
- conditional *pdf*
- conditional probability
- continuous random variable
- correlation
- covariance
- cumulative distribution function
- degrees of freedom
- discrete random variable
- expected value
- experiment
- *F*-distribution
- joint probability density function
- marginal distribution
- mean
- median
- mode
- normal distribution
- *pdf*
- probability
- probability density function
- random variable
- standard deviation
- standard normal distribution
- statistical independence
- variance