Introduction to Econometrics

Recite 1 : Review of Probability Theory

Zixin Zhou

Business School,Nanjing University

Random Variables and Probability Distribution

Expected Values, Mean, and Variance

Two Random Variables

The Normal, Chi-Squared, Student t, and F Distributions

A Fundamental Axiom of Econometrics

Any economy can be viewed as a **stochastic process** governed by some probability law.

Economic phenomenon, as often summarized in form of data, can be reviewed as a **realization** of this stochastic data generating process.

Probabilities and the Sample Space

Random Phenomena, Outcomes and Probabilities

- The mutually exclusive potential results of a random process are called the ***outcomes***.
- The **probability** of an outcome is the proportion of the time that the outcome occurs in the long run.

The Sample Space and Random Event

- The set of all possible outcomes is called ***the sample space***.
- An ***event*** is a subset of the sample space, that is, an event is a set of one or more outcomes.

Random Variables

Definition

A r.v. X, is discrete if its range(the set of values it can take) is finite $(X \in \{x_1, x_2, ...x_k\})$ or countably infinite$(X \in \{x_1, x_2, ...\})$

- eg: the number of computer crashes before deadline

Definition

A r.v. X, is continuous if it can contain all real numbers in a interval. There are an uncountably infinite number of possible realizations.

- eg: commuting times from home to school

Distributional Functions of a Discrete R.V.

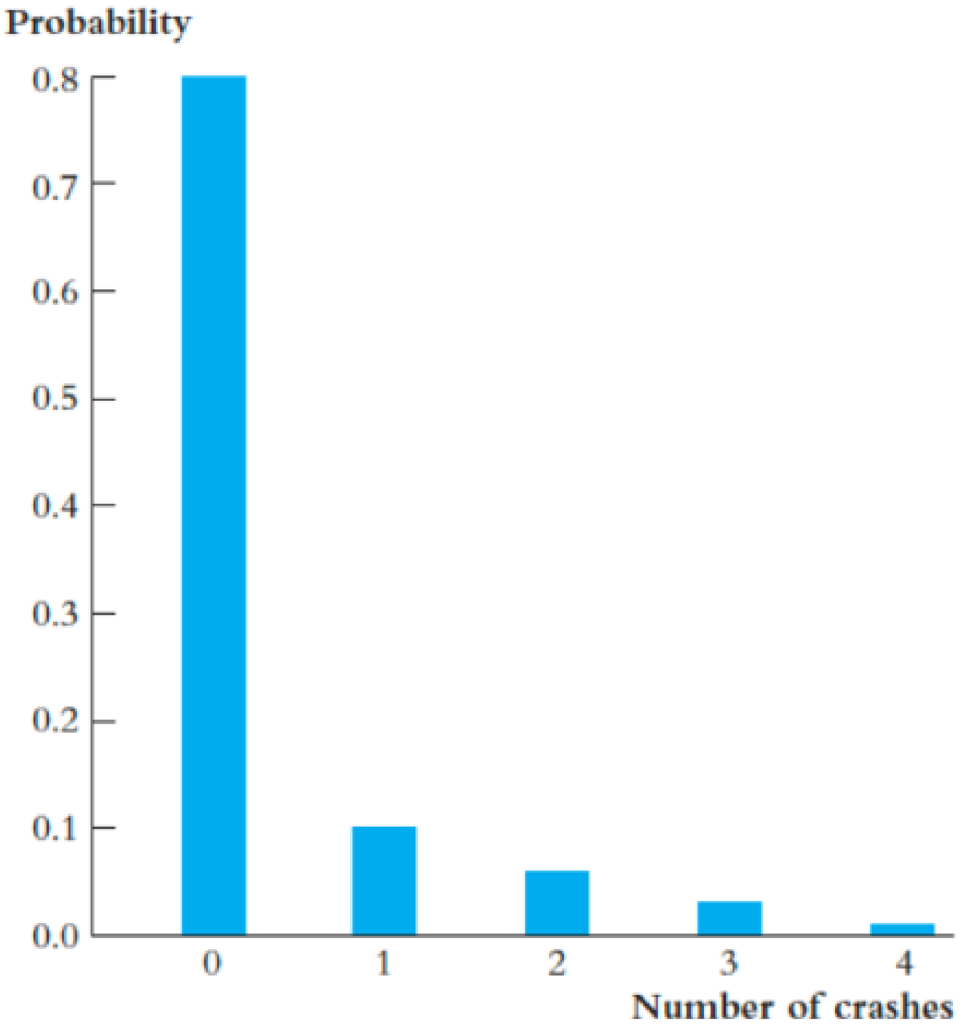
Probability mass function

Probability mass function (p.m.f.) describes the distribution of r.v. when it is discrete:

$$f_X(x_k) = P(X = x_k) = p_x,\ k = 1, 2, \ldots, n$$

FIGURE 2.1 Probability Distribution of the Number of Computer Crashes

The height of each bar is the probability that the computer crashes the indicated number of times. The height of the first bar is 0.8, so the probability of 0 computer crashes is 80%. The height of the second bar is 0.1, so the probability of 1 computer crash is 10%, and so forth for the other bars.



Distributional Functions of a Discrete R.V.

c.d.f of a discrete r.v

the c.d.f of a discrete r.v. is denoted as

$$F_X(x) = P(X \leq x) = \sum_{X_k \leq x} f_X(x_k)$$

TABLE 2.1 Probability of Your Computer Crashing *M* Times

	Outcome (number of crashes)				
	0	1	2	3	4
Probability distribution	0.80	0.10	0.06	0.03	0.01
Cumulative probability distribution	0.80	0.90	0.96	0.99	1.00

Distributional Functions of a Continuous R.V.

Probability density function

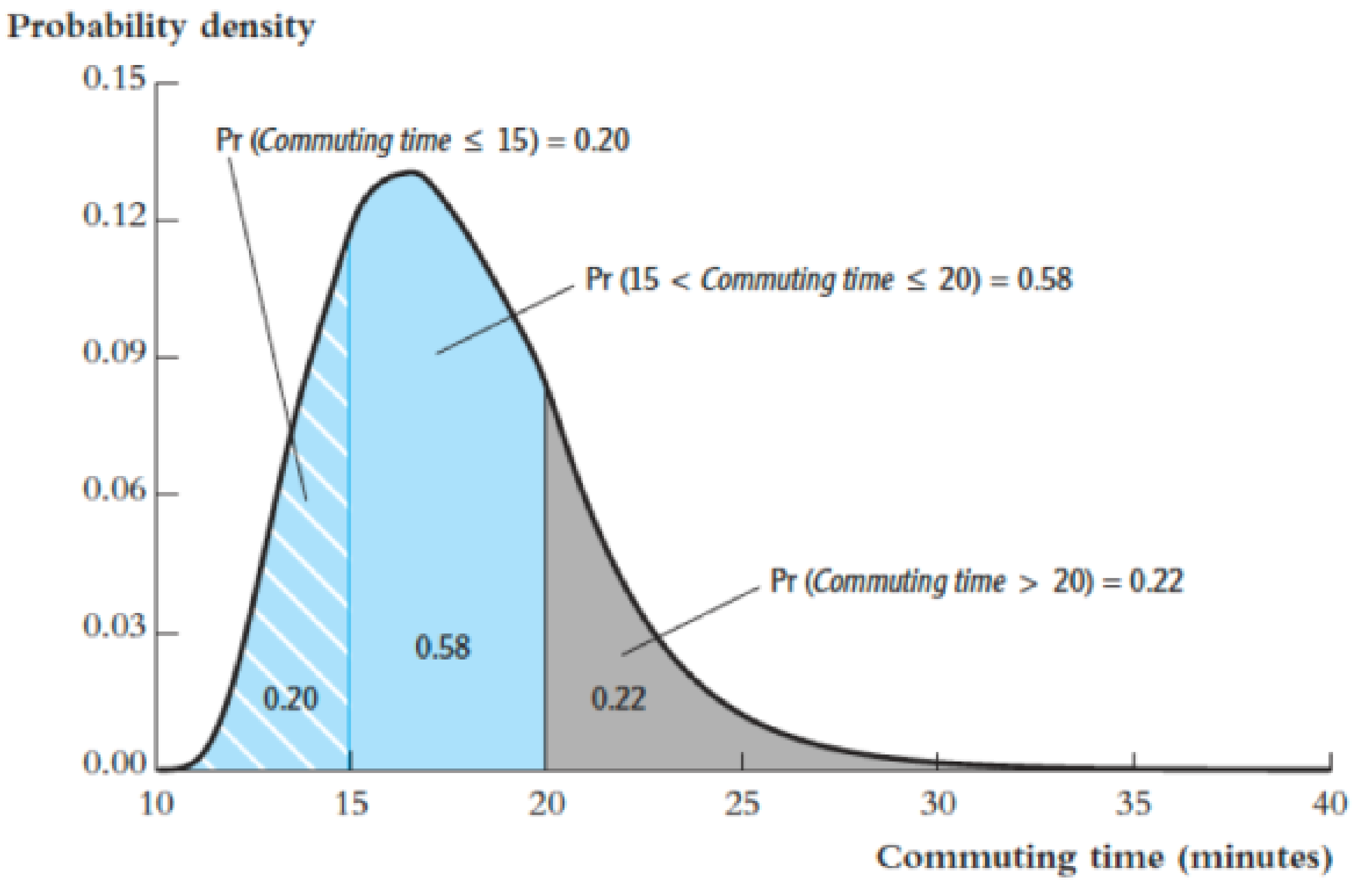
The probability density function or p.d.f., for a continuous random variable X is the function that satisfies for any interval, B

$$P(X \in B) = \int_B f_X(x)\,dx$$

Distributional Functions of a Continuous R.V.

Specifically, for a subset of the real line(a, b):

$P(a < X < b) = \int_a^b f_X(x)\,dx$, thus the probability of a region is the area under the p.d.f. for that region.



(b) Probability density function of commuting time

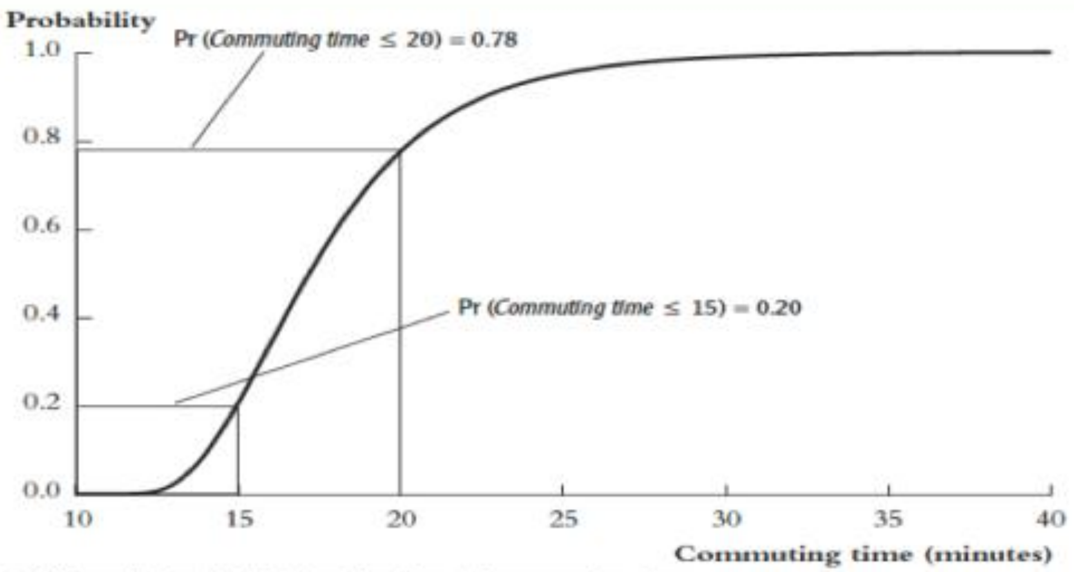
Distributional Functions of a Continuous R.V.

Cumulative probability distribution

just as it is for a discrete random variable, except using p.d.f to calculate the probability of x,

$$F(X) = P(X \leq x) = \int_{-\infty}^{x} f_X(x)\,dx$$

FIGURE 2.2 Cumulative Distribution and Probability Density Functions of Commuting Time



(a) Cumulative distribution function of commuting time

Random Variables and Probability Distribution

Expected Values, Mean, and Variance

Two Random Variables

The Normal, Chi-Squared, Student t, and F Distributions

Properties of Distributions

Probability distributions describe the uncertainty about r.v.s. The cdf/pmf/pdf give us all the information about the distribution of some r.v., but we are quite often interested in some feature of the distribution rather than the entire distribution.

- What is the difference between these two density curves? How might we summarize this difference?

There are two simple indictors:

1. **Central tendency**: where the center of the distribution is.
   - Mean/expectation (均值或期望)
2. **Spread**: how spread out the distribution is around the center.
   - Variance/standard deviation（方差或标准差）

The Expected Value of a Random Variable

The expected value of a random variable X, denoted $E(X)$ or μ_x, is the long-run average value of the random variable over many repeated trials or occurrences. it is a natural measure of central tendency.

For a *discrete* r.v., $X \in \{x_1, x_2, \ldots, x_k\}$

$$\mu_X = E[X] = \sum_{j=1}^{k} x_j p_j$$

it is computed as a *weighted average* of the value of r.v., where the weights are the probability of each value occurring.

For a *continuous* r.v., X, use the integral

$$\mu_X = E[X] = \int_{-\infty}^{+\infty} x f_X(x)\, dx$$

Properties of Expectation

Additivity: expectation of sums are sums of expectations

$$E[X+Y]=E[X]+E[Y]$$

Homogeneity: Suppose that a and b are constants. Then

$$E[aX+b]=aE[X]+b$$

Law of the Unconscious Statistician, or LOTUS, if g(x) is a function of a discrete random variable, then

$$E[g(X)]=\begin{cases}\sum_x g(x)f_X(x) & \mathit{when\, x\, is\, discrete}\\ \int g(x)f_X(x)\,dx & \mathit{when\, x\, is\, continuous}\end{cases}$$

The Variance of a Random Variable

- Besides some sense of where the middle of the distribution is, we also want to know how spread out the distribution is around that middle.

Definition

The **Variance** of a random variable X, denoted $var(X)$or σ_X^2

$$\sigma_X^2 = Var(X) = E[(X - \mu_X)^2]$$

The **Standard Deviation** of X, denoted σ_X, is just the square root of the variance.

$$\sigma_X = \sqrt{Var(X)}$$

Properties of Variance

- If a and b are constants, then we have the following properties:
  1. $V(b) = 0$
  2. $V(aX + b) = a^2 V(X)$
  3. $V(X) = E[X^2] - (E[X])^2$

Example

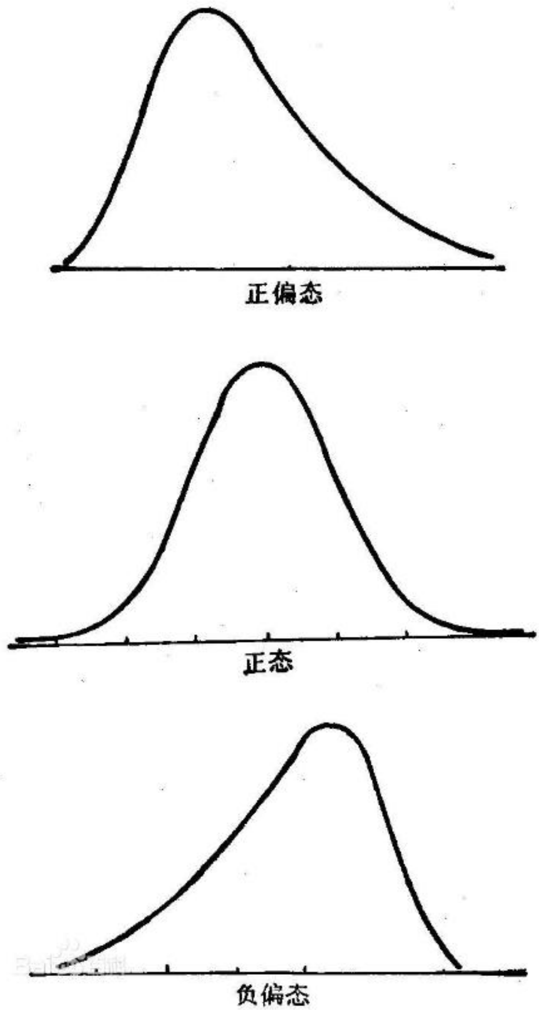
Bernoulli Distribution:

EX= p,

DX=p(1-p)

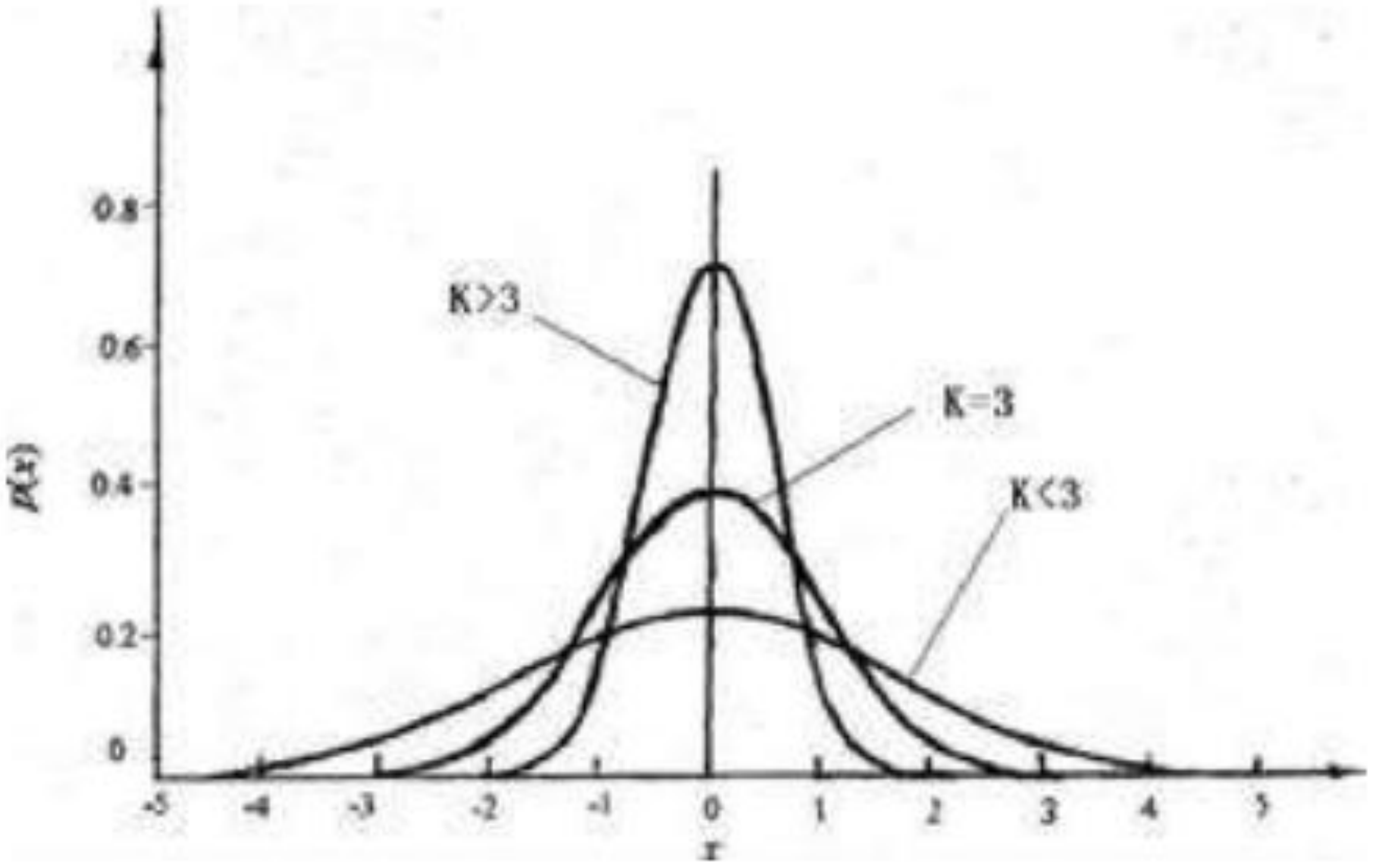
Other Measures of the Shape of a Distribution

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



Other Measures of the Shape of a Distribution

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



Random Variables and Probability Distribution

Expected Values, Mean, and Variance

Two Random Variables

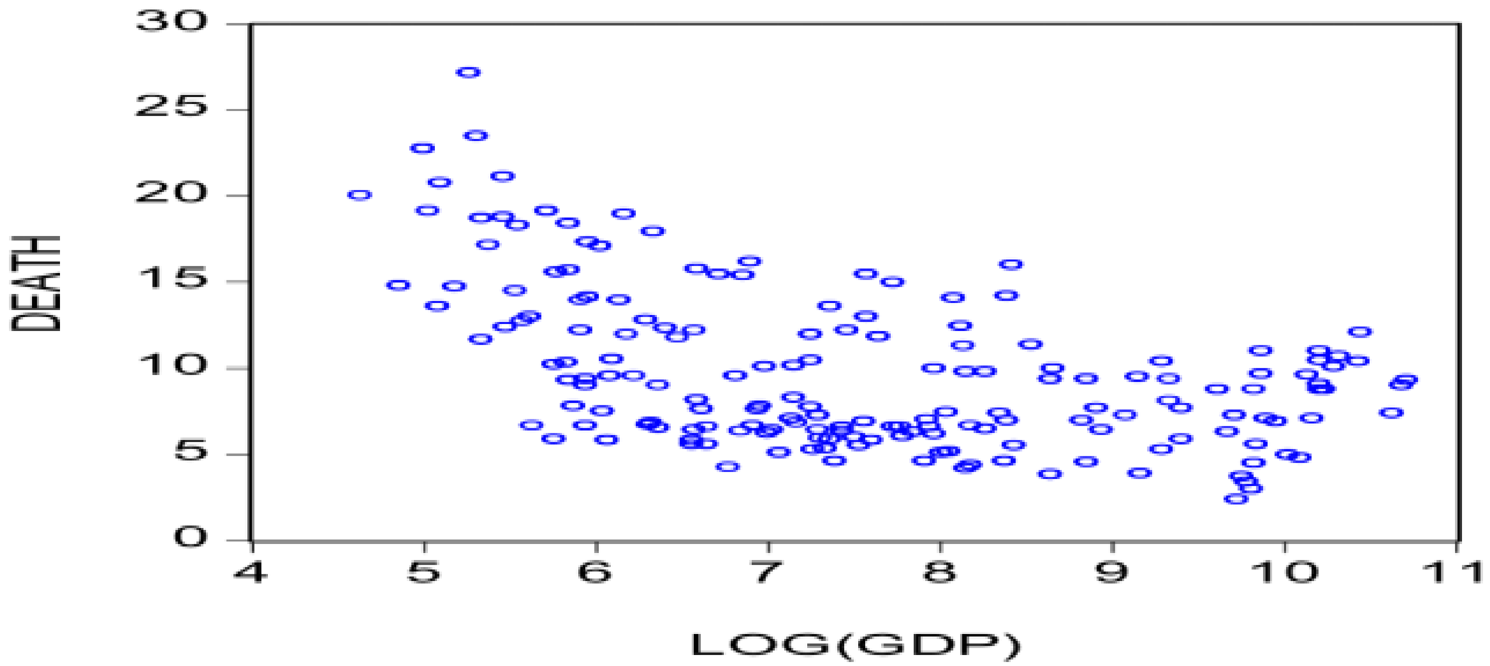
The Normal, Chi-Squared, Student t, and F Distributions

Why multiple random variables?

We are going to want to know what the relationships are between variables. "The objective of science is the discovery of the relations" —Lord Kelvin

In most cases,we often want to explore the relationship between two variables in one study.

- eg. Mortality and GDP growth



Joint Probability Distribution

Consider two *discrete* random variables X and Y with a joint probability distribution, then the joint probability mass function of (X, Y) describes the probability of any pair of values:

$$f_{X,Y}(x, y) = P(X = x, Y = y) = p_{xy}$$

TABLE 2.2 Joint Distribution of Weather Conditions and Commuting Times

	Rain ($X = 0$)	No Rain ($X = 1$)	Total
Long commute ($Y = 0$)	0.15	0.07	0.22
Short commute ($Y = 1$)	0.15	0.63	0.78
Total	0.30	0.70	1.00

Marginal Probability Distribution

The marginal distribution: often need to figure out the distribution of just one of the r.v.s.

$$f_Y(y) = P(Y = y) = \sum_x f_{X,Y}(x, y)$$

Intuition: sum over the probability that $Y = y$ for all possible values of x.

TABLE 2.2 Joint Distribution of Weather Conditions and Commuting Times

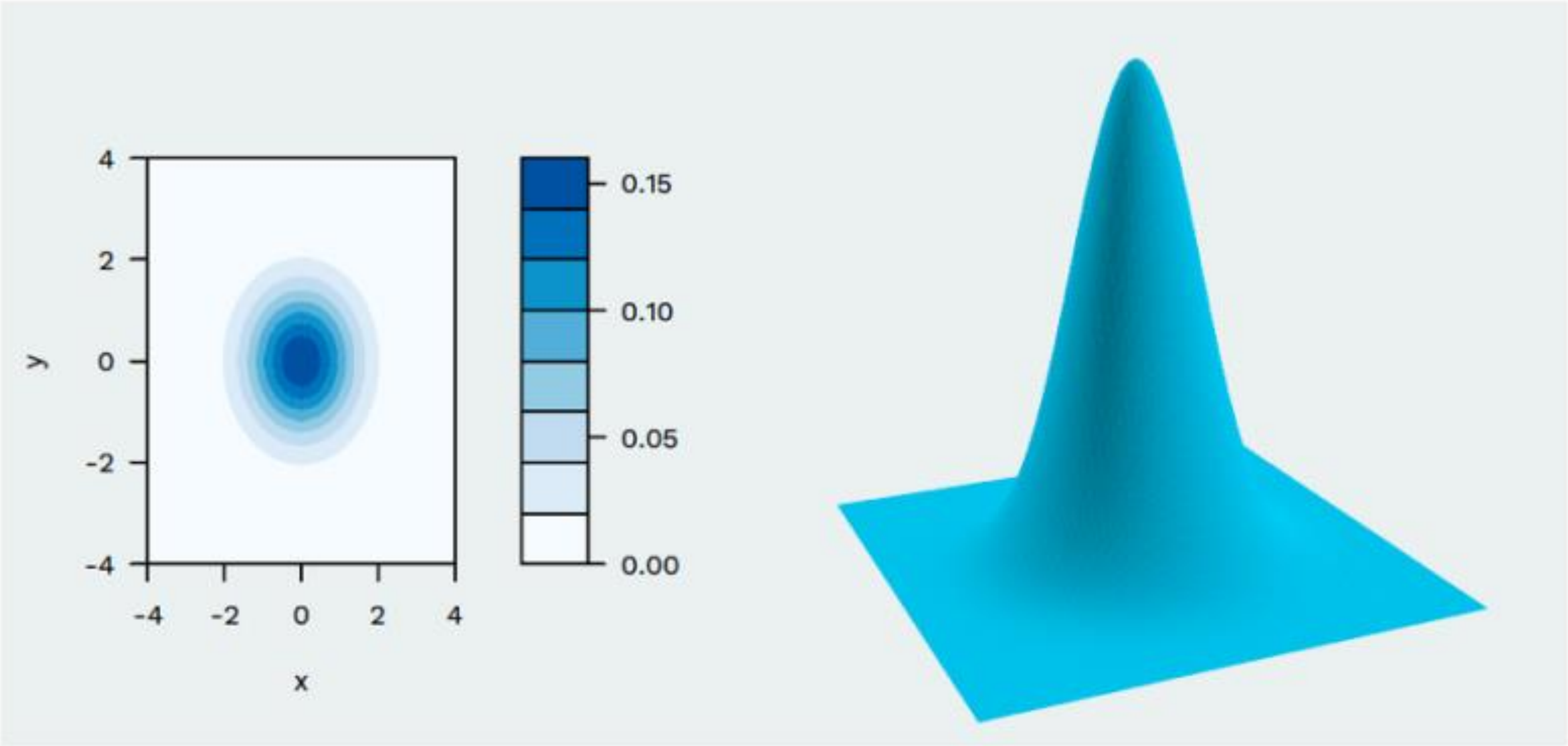
	Rain ($X = 0$)	No Rain ($X = 1$)	Total
Long commute ($Y = 0$)	0.15	0.07	0.22
Short commute ($Y = 1$)	0.15	0.63	0.78
Total	0.30	0.70	1.00

Joint Probability Density Function

Consider two *continuous* random variables X and Y with a joint probability distribution, then the **joint probability density function** of (X, Y) is a function, denoted as $f_{X,Y}(x, y)$ such that:

1. $f_{X,Y}(x, y) \geq 0$
2. $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x, y)\, dxdy = 1$
3. $P(a < X < b, c < Y < d) = \int_c^d \int_a^b f_{X,Y}(x, y)\, dxdy$, thus the probability in the $\{a, b, c, d\}$area.

Joint Probability Density Function

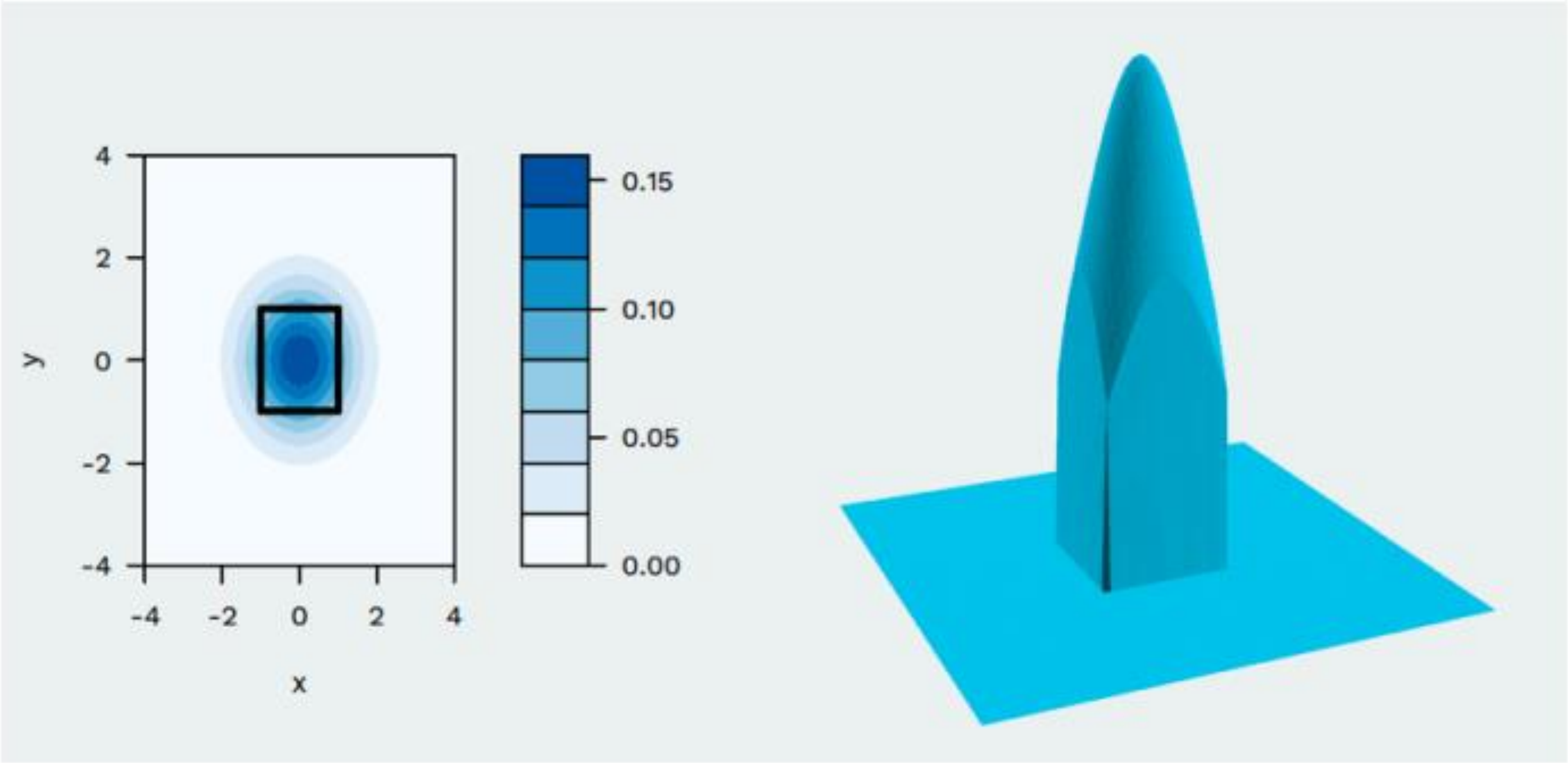


- Y and X axes denote on the "floor", height is the value of $f_{XY}(x, y)$

Joint Probability Density Function

The probability equals to volume above a specific region

$$P(X, Y) \in A) = \int_{(x,y)\in A} f_{X,Y}(x, y)\, dxdy$$



Continuous Marginal Distribution

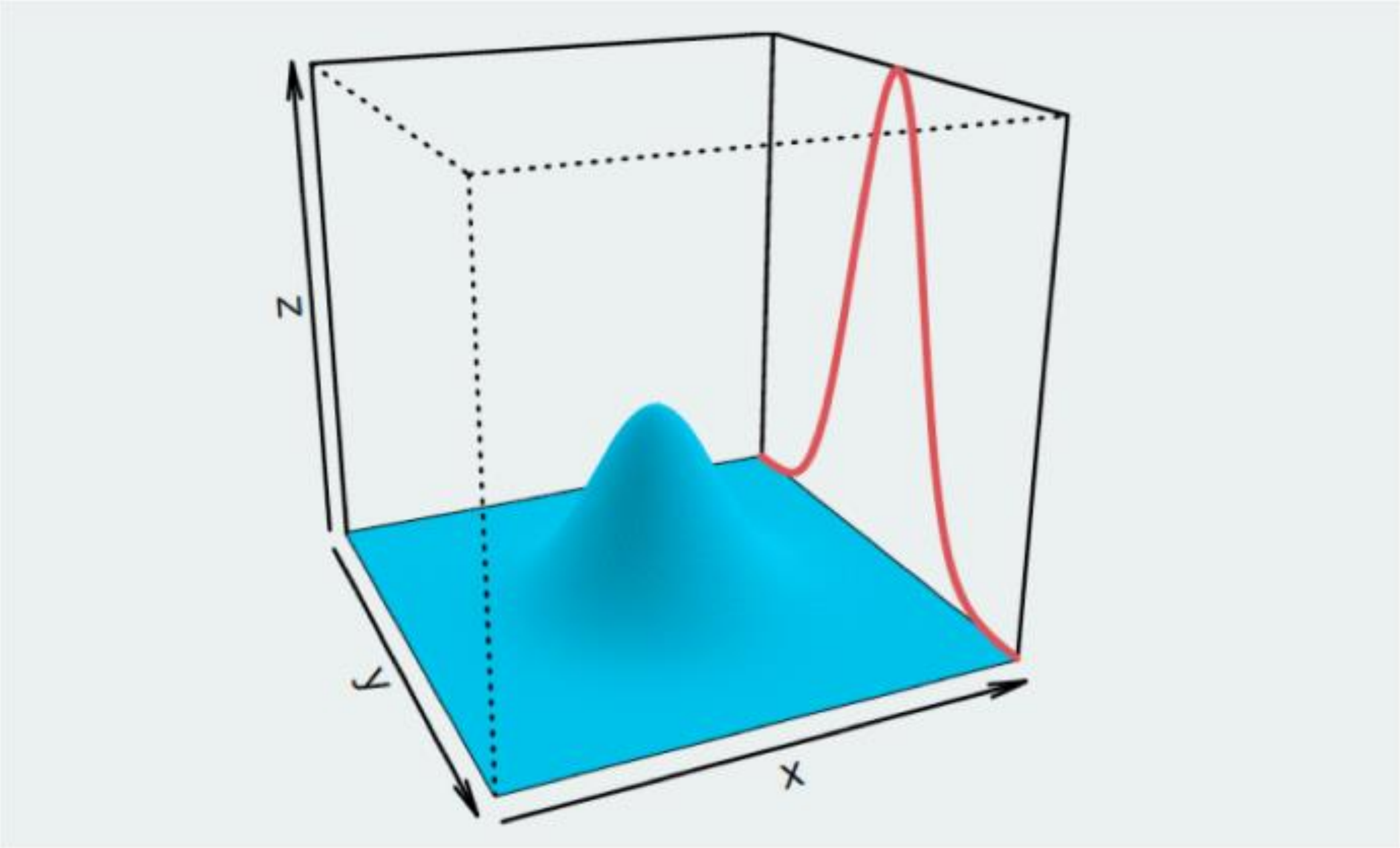
the marginal p.d.f of Y by integrating over the distribution of X:

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y)\, dx$$

the marginal p.d.f of X by integrating over the distribution of Y:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y)\, dy$$

Continuous Marginal Distribution



- Pile up all of the joint density onto a single dimension

Joint Cumulative Distribution Function

The **joint cumulative distribution function** of (X, Y) is

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u, v)\, du dv$$

Transform joint c.d.f and joint p.d.f

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y}$$

Conditional Distributions

The distribution of a random variable Y conditional on another random variable X taking on a specific value is called the **Conditional Distribution** of Y given X. The conditional probability that Y takes on the value y when X takes on the value x is written Pr(Y=y|X=x).

$$\Pr(\mathrm{Y}=\mathrm{y}|\mathrm{X}=\mathrm{x}) = \frac{\Pr(X=x, Y=y)}{\Pr(X=x)}$$

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}$$

TABLE 2.2 Joint Distribution of Weather Conditions and Commuting Times

	Rain ($X = 0$)	No Rain ($X = 1$)	Total
Long commute ($Y = 0$)	0.15	0.07	0.22
Short commute ($Y = 1$)	0.15	0.63	0.78
Total	0.30	0.70	1.00

Conditional expectation

The conditional expectation of Y given X, also called the conditional mean, is the mean of the conditional distribution of Y given X.

Conditional expectation

离散型 $$E(Y|X=x)=\sum y_i \Pr(Y=y_i|X=x)$$

连续型 $$E(Y|X=x)=\int y f_{Y|X}(y|x)dy$$

Intuition:期望就是求平均值，而条件期望是“分组取平均”

迭代期望定律 $$E(Y)=E[E(Y|X)]$$

Conditional Variance

离散型 $$Var(Y|X=x_i)=\sum_{i=1}^{k}[y_i-E(Y|X=x)]^2\Pr(Y=y_i|X=x)$$

连续型 $$Var(Y|X)=\int_y (y-E[Y|X])^2 f_{Y|X}(y|x)$$

Independence

Two random variables X and Y are independently distributed, or independent, if knowing the value of one of the variables provides no information about the other.

$$\Pr(Y = y|X = x) = \Pr(Y = y)$$

$$\Pr(\mathrm{Y} = \mathrm{y}|\mathrm{X} = \mathrm{x}) = \frac{\Pr(X = x, Y = y)}{\Pr(X = x)}$$

$$\Pr(X = x, Y = y) = \Pr(X = x)\Pr(Y = y)$$

Covariance and Correlation

$$cov(X,Y) = \sigma_{xy} = E[(X-\mu_X)(Y-\mu_Y)] = \sum_{i=1}^{k}\sum_{j=1}^{l}(x_j-\mu_X)(y_j-\mu_Y)\Pr(X=x_j, Y=y_i)$$

$$\mathrm{Corr}(\mathrm{X},\mathrm{Y}) = \frac{cov(X,Y)}{\sqrt{var(X)\mathrm{var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

$$-1 \le corr(X,Y) \le 1$$

Means, Variances, and Covariances of Sums of Random Variables

$$E(a + bX + cY) = a + b\mu_X + c\mu_Y$$

$$Var(a + bY) = b^2\sigma_Y^2$$

$$Var(aX + bY) = a^2\sigma_X^2 + 2ab\sigma_{XY} + b^2\sigma_Y^2$$

$$E(Y^2) = \sigma_Y^2 + \mu_Y^2$$

$$cov(a + bX + cY, Y) = b\sigma_{XY} + c\sigma_Y^2$$

$$E(XY) = \sigma_{XY} + \mu_x\mu_Y$$

Random Variables and Probability Distribution

Expected Values, Mean, and Variance

Two Random Variables

The Normal, Chi-Squared, Student t, and F Distributions

The Normal Distribution

- The p.d.f of a normal random variable X is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right], \ -\infty < X < +\infty$$

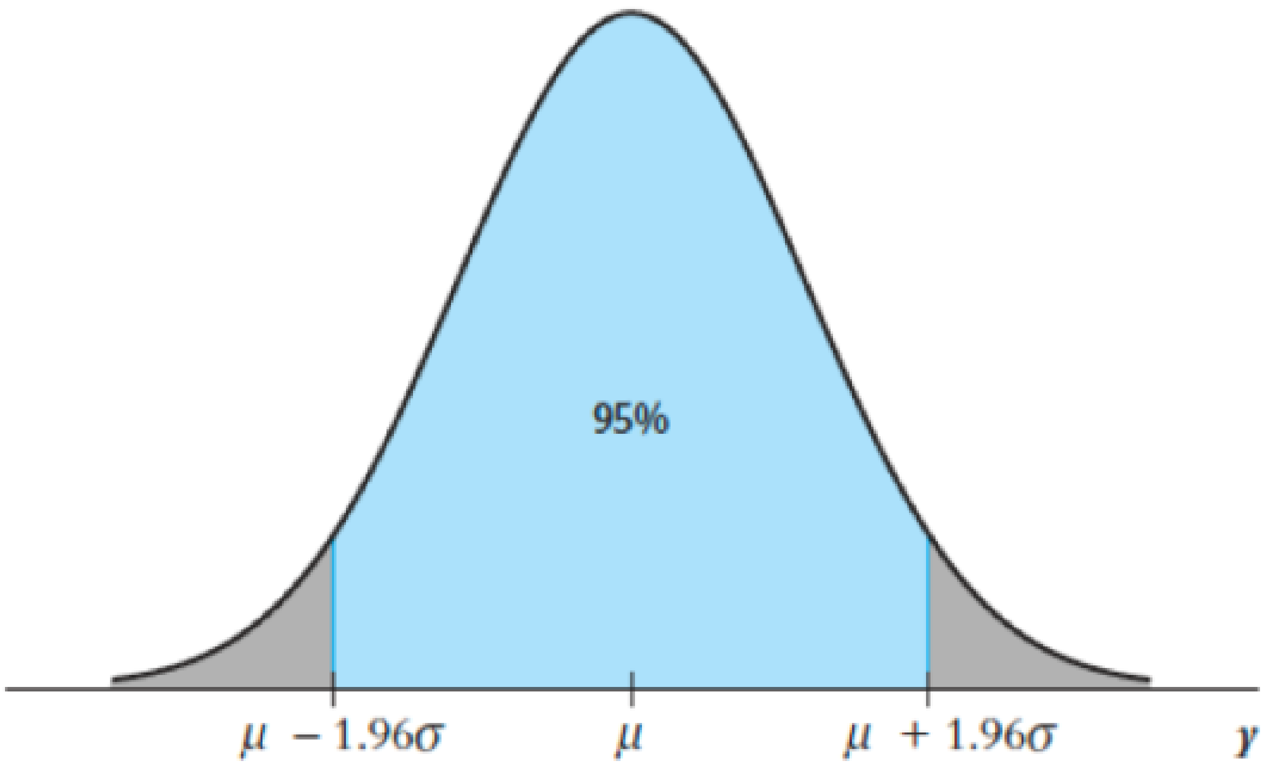
- if X is normally distributed with expected value μ and variance σ^2, denoted as $X \sim N(\mu, \sigma^2)$
    - if we know these two parameters, we know everything about the distribution.
- Examples: Human heights, weights, test scores,
- If X represents wage, income or consumption etc, it will has a log-normal distribution, thus

$$log(X) \sim N(\mu, \sigma^2)$$

The Normal Distribution

FIGURE 2.5 The Normal Probability Density

The normal probability density function with mean μ and variance σ^2 is a bell-shaped curve, centered at μ. The area under the normal p.d.f. between $\mu - 1.96\sigma$ and $\mu + 1.96\sigma$ is 0.95. The normal distribution is denoted $N(\mu, \sigma^2)$.



The Standard Normal Distribution

A special case of the normal distribution where the mean is zero ($\mu = 0$) and the variance is one ($\sigma^2 = \sigma = 1$), then its p.d.f is

$$f_X(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad -\infty < X < +\infty$$

if X is standard normally distributed, then denoted as $X \sim N(0, 1)$

The standard normal cumulative distribution function is denoted

$$\Phi(z) = P(Z \leq z)$$

where z is a standardize r.v. thus $z = \frac{x - \mu_X}{\sigma_X}$

The Standard Normal Distribution

FIGURE B.8 The standard normal cumulative distribution function.

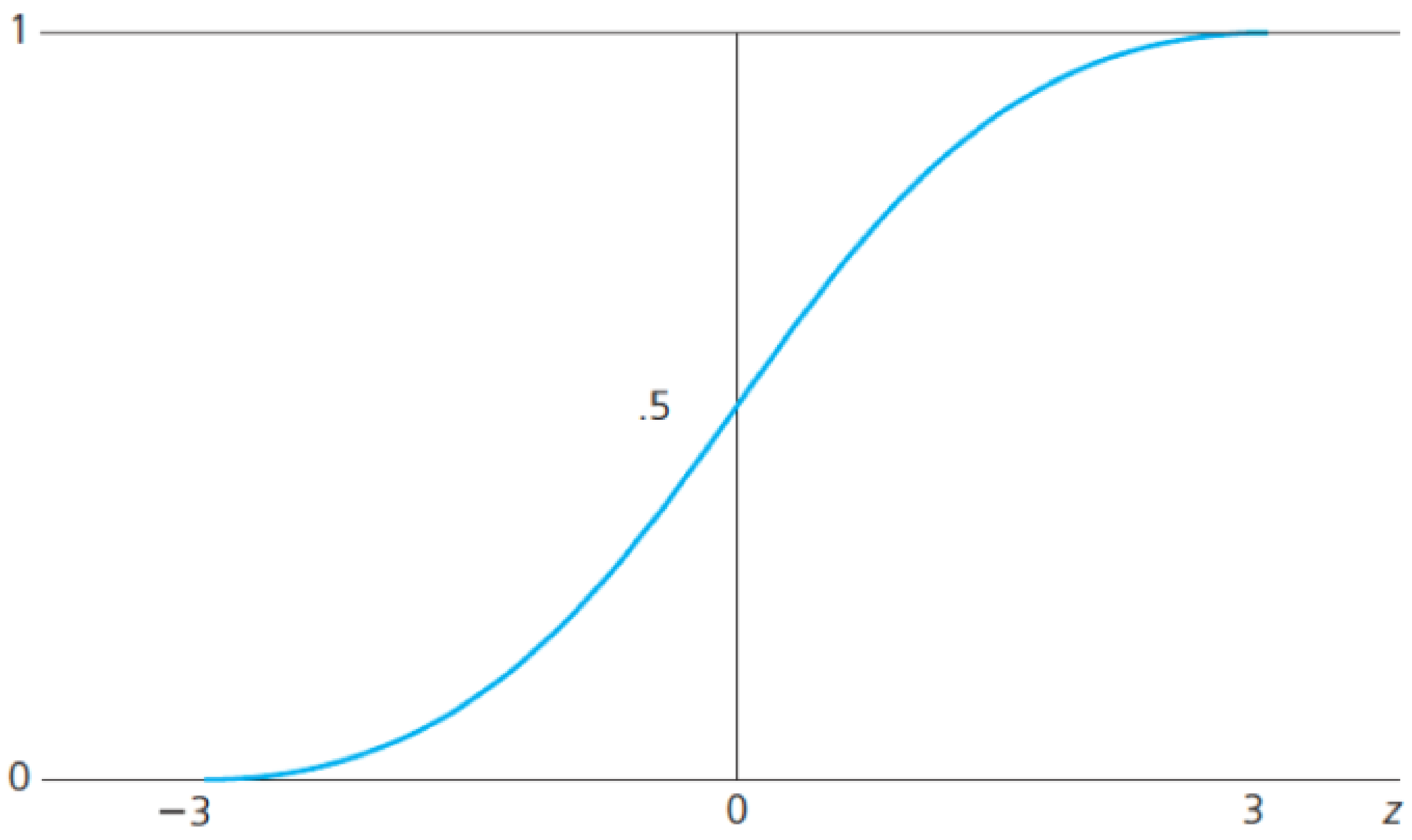
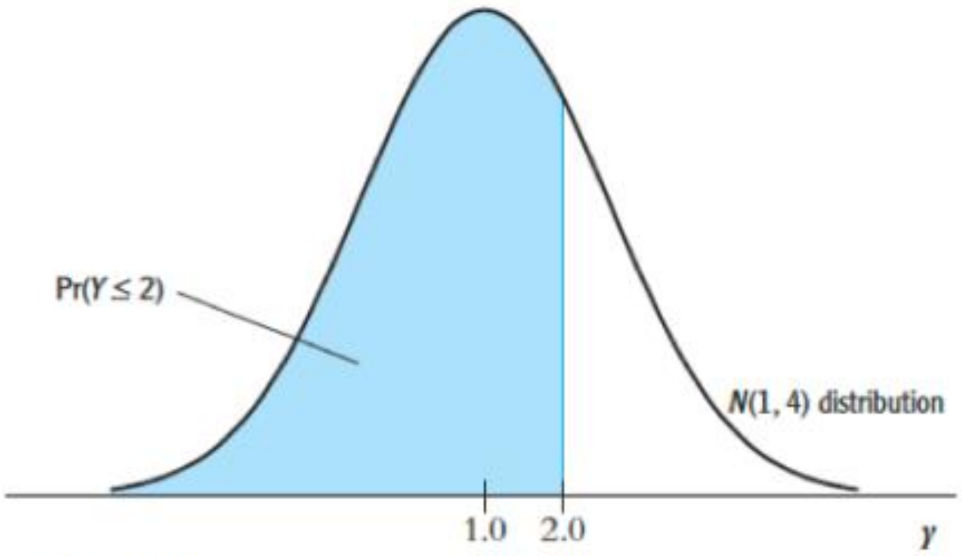
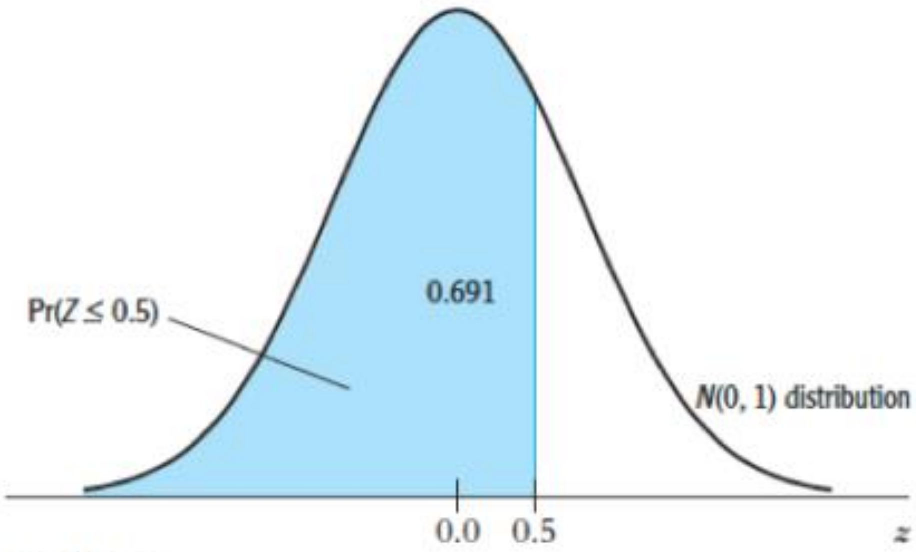


FIGURE 2.6 Calculating the Probability That $Y \leq 2$ When Y Is Distributed $N(1, 4)$

To calculate $\Pr(Y \leq 2)$, standardize Y, then use the standard normal distribution table. Y is standardized by subtracting its mean ($\mu = 1$) and dividing by its standard deviation ($\sigma = 2$). The probability that $Y \leq 2$ is shown in Figure 2.6a, and the corresponding probability after standardizing Y is shown in Figure 2.6b. Because the standardized random variable, $(Y - 1)/2$, is a standard normal (Z) random variable, $\Pr(Y \leq 2) = \Pr\left(\frac{Y-1}{2} \leq \frac{2-1}{2}\right) = \Pr(Z \leq 0.5)$. From Appendix Table 1, $\Pr(Z \leq 0.5) = \Phi(0.5) = 0.691$.



(a) $N(1, 4)$



(b) $N(0, 1)$

The Chi-Square Distribution

Let $Z_i (i = 1, 2, \ldots, m)$ be independent random variables, each distributed as **standard normal**. Then a new random variable can be defined as the sum of the squares of Z_i :

$$X = \sum_{i=1}^{m} Z_i^2$$

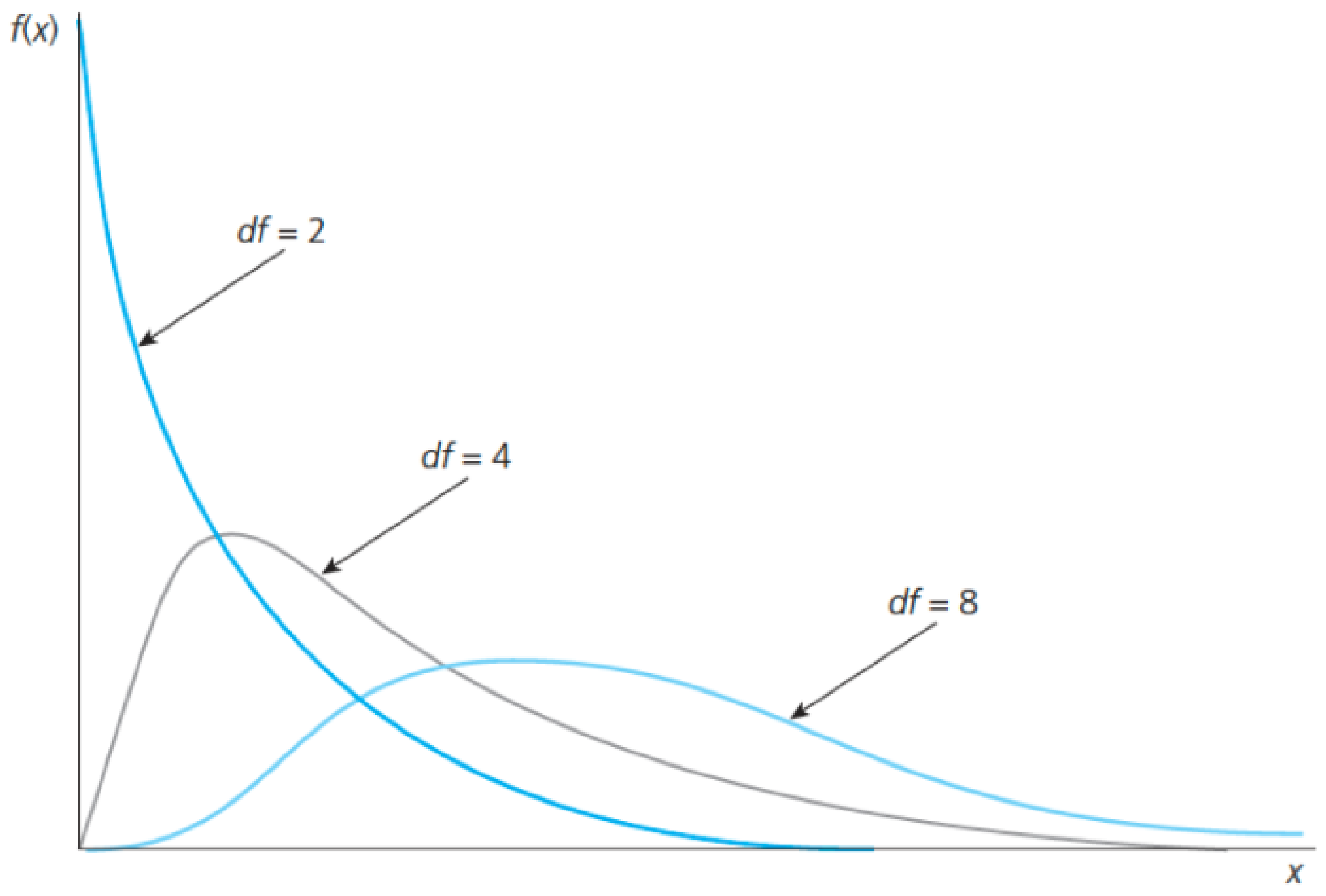
Then X has a **chi-squared distribution** with m **degrees of freedom**

The form of the distribution varies with the number of degrees of freedom, i.e. the number of standard normal random variables Z_i included in X.

The distribution has a long tail, or is skewed, to the right. As the degrees of freedom m gets larger, however, the distribution becomes more symmetric and ''bell-shaped.'' In fact, as m gets larger, the chi-square distribution converges to, and essentially becomes, a **normal distribution**.

The Chi-Square Distribution

FIGURE B.9 The chi-square distribution with various degrees of freedom.



The Student t Distribution

The Student *t* distribution can be obtained from a standard normal and a chi-square random variable.

Let Z have a standard normal distribution, let X have a chi-square distribution with *m* degrees of freedom and assume that Z and X are independent. Then the random variable

$$T = \frac{Z}{\sqrt{X/n}}$$

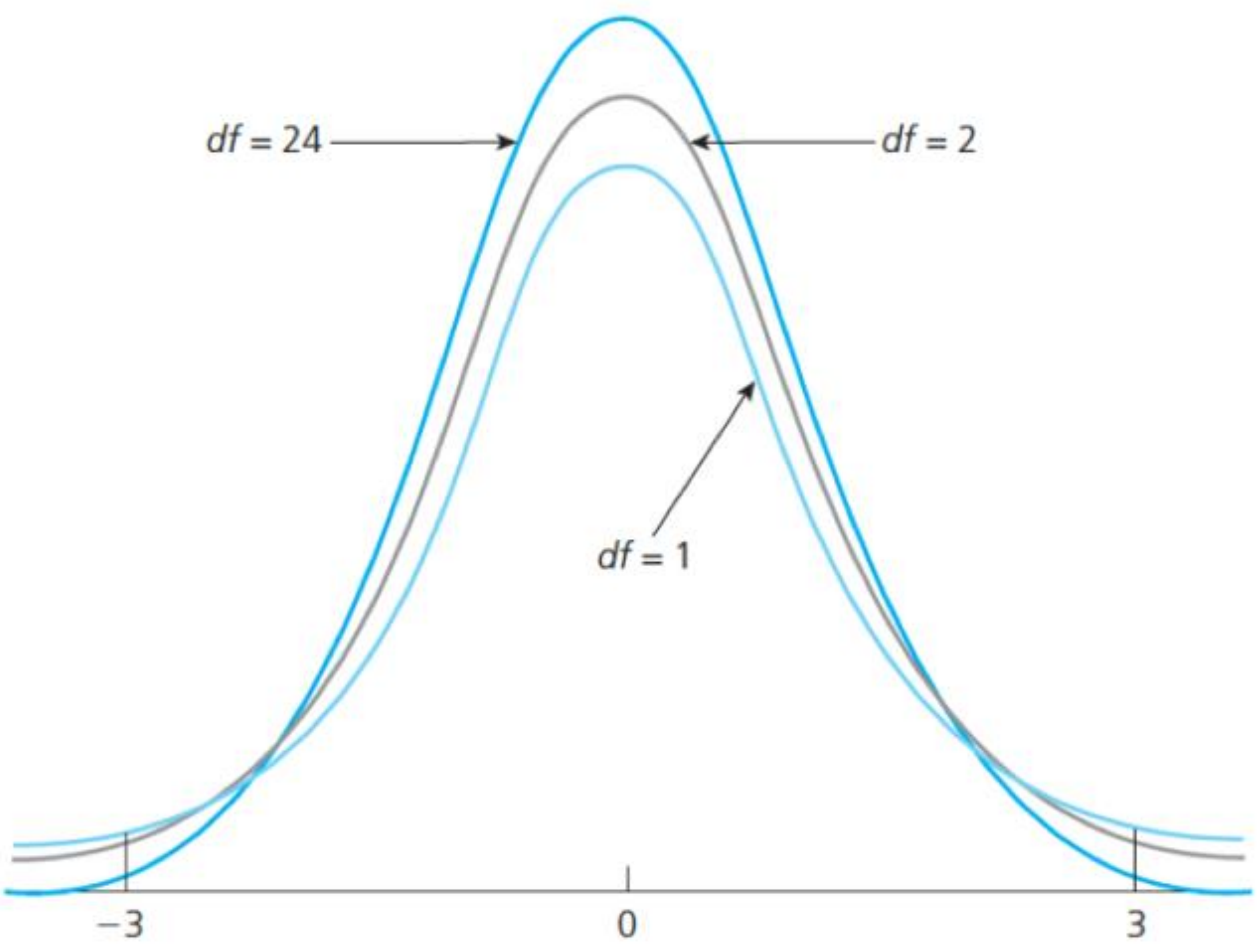
has has a t-distribution with m degrees of freedom, denoted as $T \sim t_n$.

The shape of the t-distribution is similar to that of a normal distribution, except that the t-distribution has more probability mass in the tails.

As the degrees of freedom get large, the t-distribution approaches **the standard normal distribution**.

The Student t Distribution

FIGURE B.10 The *t* distribution with various degrees of freedom.



The F Distribution

Let $X_1 \sim \chi^2_m$ and $X_2 \sim \chi^2_n$, and assume that X_1 and X_2 are independent,

$$Z = \frac{\frac{X_1}{m}}{\frac{X_2}{n}} \sim F_{m,n}$$

thus Z has an F-distribution with (m, n) degrees of freedom.

The F Distribution

FIGURE B.11 The F_{k_1,k_2} distribution for various degrees of freedom, k_1 and k_2.