Chapter 3
Random Variables and Probability Distributions

吳明龍 副教授
成大資訊系/醫資所
minglong.wu@csie.ncku.edu.tw
辦公室:資訊系館 12 樓

3.1 Concept of a Random Variable

- Random variable: is a function that associates a real number with each element in the sample space, using a capital letter, say *X*, to denote a random variable.
- Example 3.1
  - Two balls are drawn in succession without replacement from an box containing 4 red balls and 3 black balls.
  - The possible outcomes and the values *y* of the random variable *Y*, where *Y* is the number of red balls, are

Sample space	y
RR	2
RB	1
BR	1
BB	0

Concept of a Random Variable

- Definition 3.2: **Discrete sample space**: If a sample space contains a finite number of possibilities or an unending sequence with as many elements as there are whole numbers.
- Definition 3.3: **Continuous sample space**: If a sample space contains an infinite number of possibilities equal to the number of points on a line segment.
- **Discrete random variable**: If the set of possible outcomes of a random variable is countable.
- **Continuous random variable**: If a random variable can take on values on a continuous scale.
- Discrete random variables often represent count data
  - The number of defectives, highway fatalities
- Continuous random variables often represent measured data
  - Heights, weights, temperatures, distance or life periods

3.2 Discrete Probability Distributions

- Frequently, it is convenient to represent all the probabilities of a random variable *X* by a formula.

  $f(x) = P(X = x)$; e.g., $f(3) = P(X = 3)$

- **Probability function** (probability mass function, **probability distribution**) of the discrete random variable *X*: The set of ordered pairs $(x, f(x))$

  1. $f(x) \geq 0$
  2. $\sum_{x \in X} f(x) = 1$
  3. $P(X = x) = f(x)$

Discrete Probability Distributions

- Example 3.8
- A shipment of 20 similar laptops to a retail outlet contains 3 that are defective.
  - If a school make a random purchase of 2 of these computers.
  - Find the probability distribution for the number of defectives.

$$f(0) = P(X = 0) = \frac{\binom{3}{0}\binom{17}{2}}{\binom{20}{2}} = \frac{68}{95}$$

$$f(1) = P(X = 1) = \frac{\binom{3}{1}\binom{17}{1}}{\binom{20}{2}} = \frac{51}{190}$$

$$f(2) = P(X = 2) = \frac{\binom{3}{2}\binom{17}{0}}{\binom{20}{2}} = \frac{3}{190}$$

x	0	1	2
$f(x)$	$\frac{68}{95}$	$\frac{51}{190}$	$\frac{3}{190}$

Discrete Probability Distributions

- Example 3.9
  - If a car agency sells 50% of its inventory of a certain foreign car equipped with airbags.
  - Find a formula for the probability distribution of the number of cars with airbags among the next 4 cars sold by the agency.

$$f(x) == \frac{\binom{4}{x}}{16}, \quad \text{for x} = 0,1,2,3,4$$

Discrete Probability Distributions

- **Cumulative distribution**, $F(x)$ of a discrete random variable X with probability distribution $f(x)$ is $F(x) = P(X \le x) = \sum_{t \le x} f(t)$

- Example 3.10: find the cumulative distribution of the random variable X in example 3.9

$$f(0) = \frac{1}{16},\ f(1) = \frac{4}{16},\ f(2) = \frac{6}{16},\ f(3) = \frac{4}{16},\ f(4) = \frac{1}{16},$$

$$F(0) = f(0) = \frac{1}{16}$$

$$F(1) = f(0) + f(1) = \frac{5}{16}$$

$$F(2) = f(0) + f(1) + f(2) = \frac{11}{16}$$

$$F(3) = f(0) + f(1) + f(2) + f(3) = \frac{15}{16}$$

$$F(4) = f(0) + f(1) + f(2) + f(3) + f(4) = 1$$

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{16} & \text{for } 0 \le x < 1 \\ \frac{5}{16} & \text{for } 1 \le x < 2 \\ \frac{11}{16} & \text{for } 2 \le x < 3 \\ \frac{15}{16} & \text{for } 3 \le x < 4 \\ 1 & \text{for } x \ge 4 \end{cases}$$

Discrete Probability Distributions

- Bar chart and probability histogram

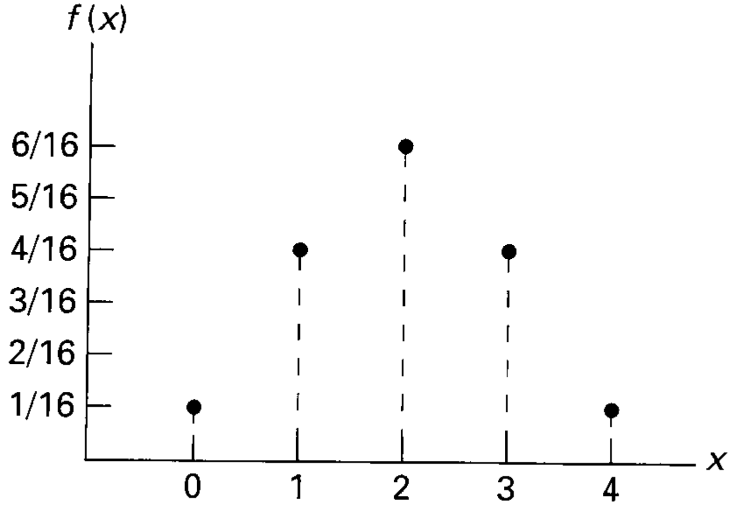

Figure 3.1 Probability mass function plot

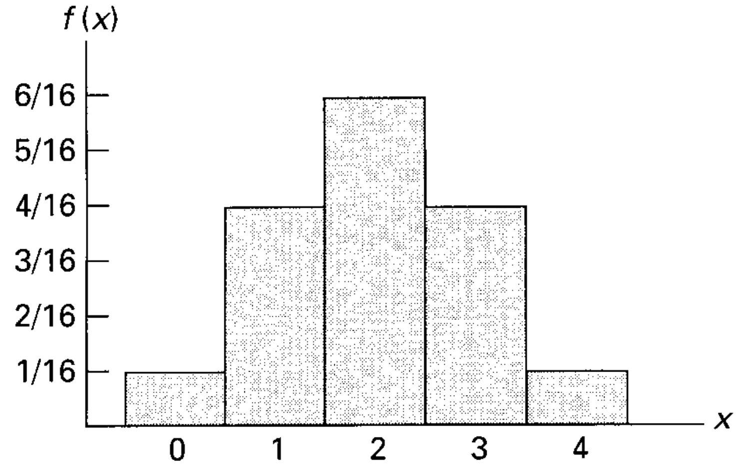

Figure 3.2 Probability histogram.

Discrete Probability Distributions

- Discrete cumulative distribution

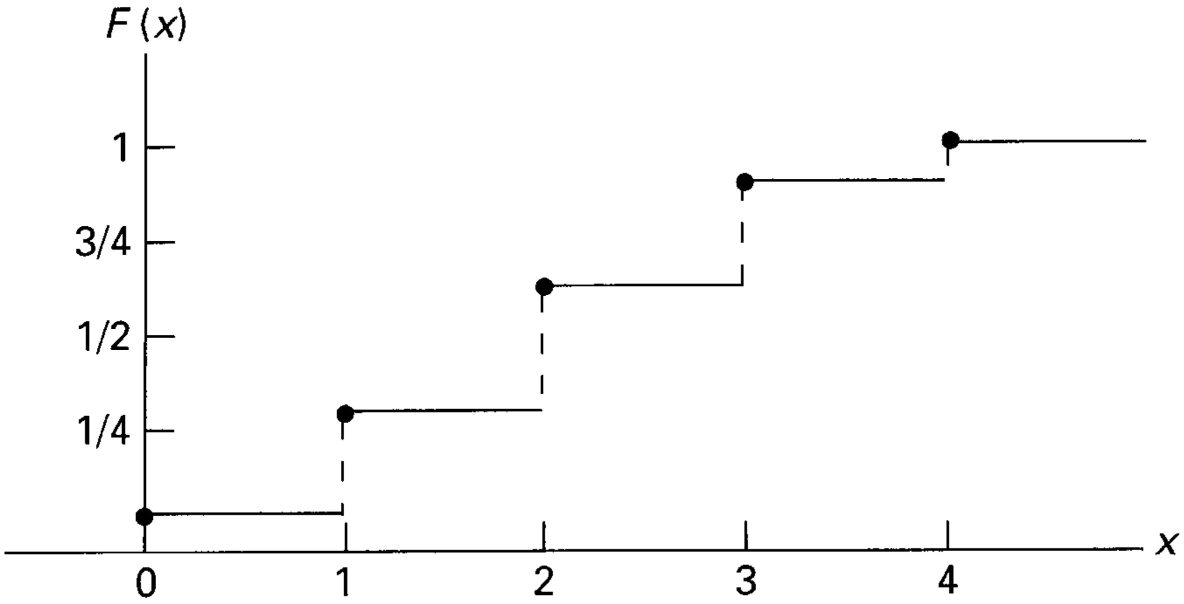

Figure 3.3 Discrete cumulative distribution.

3.3 Continuous Probability Distributions

- Definition 3.6: The function *f*(*x*) is a probability density function (density function, p.d.f) for the continuous random variable *X*, defined over the set of real numbers *R*, if

  1. $f(x) \geq 0$, for all $x \in R$
  2. $\int_{-\infty}^{\infty} f(x)dx = 1$
  3. $P(a < X < b) = \int_a^b f(x)dx$

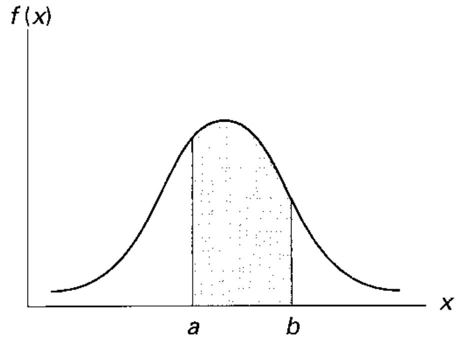

Figure 3.5 $P(a < X < b)$.

- A probability density function is constructed so that the area under its curve bounded by the *x* axis is equal to 1.

Continuous Probability Distributions

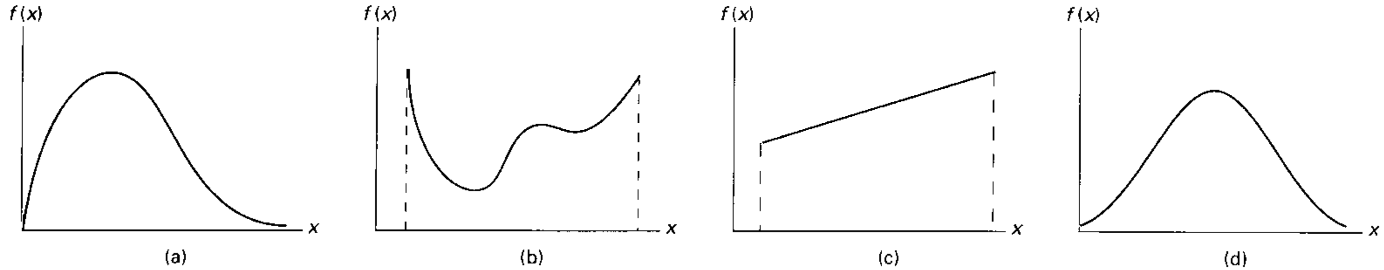

Figure 3.4 Typical density functions.

- Example 3.11: Suppose that the error in reaction temperature in ºC is a continuous random variable X having the probability density function

$$f(x)=\begin{cases}\frac{x^2}{3}, & -1<x<2\\ 0, & \text{elsewhere.}\end{cases}$$

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

(b) Find $P(0<X\leq 1)$.

$$\text{Solution}: (a)\ \int_{-\infty}^{\infty} f(x)dx=\int_{-1}^{2}\frac{x^2}{3}dx=\left.\frac{x^3}{9}\right|_{-1}^{2}=\frac{8}{9}+\frac{1}{9}=1.$$

$$(b)\ P(0<X\leq 1)=\int_{0}^{1}\frac{x^2}{3}dx=\left.\frac{x^3}{9}\right|_{0}^{1}=\frac{1}{9}.$$

Continuous Probability Distributions

- Definition 3.7: The <u>cumulative function</u> *F*(*x*) of a continuous random variable *X* with density function *f*(*x*) is

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt \quad \text{for} -\infty < \mathrm{x} < \infty$$

$$\Rightarrow P(a < X < b) = F(b) - F(a) \text{ and } f(x) = \frac{dF(x)}{dx}$$

Continuous Probability Distributions

- Example 3.12: For the density function of Example 3.11 find *F*(*x*), and use it to evaluate *P*(0 < *X* ≤ 1).

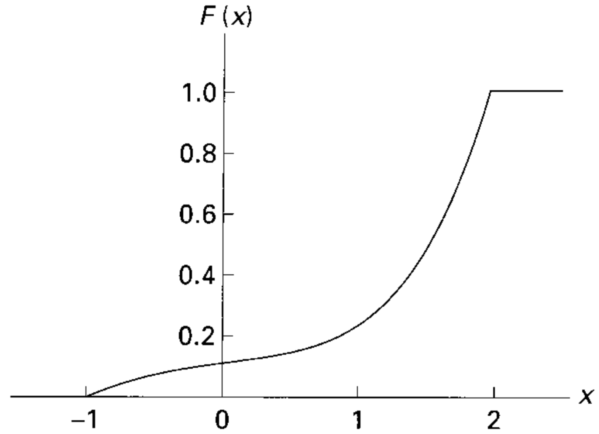

Figure 3.6 Continuous cumulative distribution.

$$\text{Solution: For } -1 < x < 2,$$

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{x} \frac{t^2}{3}dt = \left.\frac{t^3}{9}\right|_{-1}^{x} = \frac{x^3+1}{9}$$

$$\Rightarrow F(x) = \begin{cases} 0, & x \le -1 \\ \frac{x^3+1}{9}, & -1 \le x < 2 \\ 1, & x \ge 2 \end{cases}$$

$$\Rightarrow P(0 < X \le 1) = F(1) - F(0) = \frac{2}{9} - \frac{1}{9} = \frac{1}{9}$$

3.4 Joint Probability Distributions

- Definition 3.8: The function $f(x, y)$ is a <u>joint probability distribution</u> (<u>probability mass function</u>) of the discrete random variables X and Y if 1. $f(x, y) \geq 0$, for all (x, y)

  $$2. \sum_x \sum_y f(x, y) = 1$$

  $$3. P(X = x, Y = y) = f(x, y)$$

  – For any region A in the xy plane,

  $$P[(X, Y) \in \mathrm{A}] = \sum_A \sum f(x, y)$$

Joint Probability Distributions

- Example 3.14: Two refills for a ballpoint pen are selected at random from a box that contains 3 blue refills, 2 red refills, and 3 green refills. If X is the number of blue refills and Y is the number of red refills selected, find (a) the joint probability function *f*(*x*, *y*), and (b) $P[(X,Y)\in \mathrm{A}]$, where A is the region $\{(x,y)\,|\,x+y\le 1\}$.
  - Solution

(a) $f(x,y)=\dfrac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}}$

(b) $P[(X,Y)\in A]=P(X+Y\le 1)$

$= f(0,0)+f(0,1)+f(1,0)$

$=\frac{3}{28}+\frac{3}{14}+\frac{9}{28}=\frac{9}{14}$

f(*x*, *y*)		*x* = 0	*x* = 1	*x* = 2	Row totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$		$\frac{3}{7}$
	2	$\frac{1}{28}$			$\frac{1}{28}$
Column totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Joint Probability Distributions

- Definition 3.9: The function *f*(*x* ,*y*) is a joint density function of the continuous random variables *X* and *Y* if

$$1.\ f(x, y) \geq 0,\ \text{for all}\ (x, y)$$

$$2.\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$$

$$3.\ P[(X, Y) \in A] = \int_A\int f(x,y)dxdy$$

  – for any region A in the *xy* plane.

Joint Probability Distributions

- Example 3.15 (9^{th} ed.)
  - A business operates both drive-in facility and a walk-in facility.
  - For a randomly selected day, let X and Y, respectively, be the proportions of the time that the drive-in and the walk-in facilities are in use.
  - The joint density function is as follows:

Joint Probability Distributions

$$f(x,y)=\begin{cases}\frac{2}{5}(2x+3y), & 0\le \mathrm{x}\le 1,\ 0\le \mathrm{y}\le 1\\ 0, & \text{elsewhere}\end{cases}$$

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

(b) Find $P[(X,Y)\in A]$, where A is the region $\left\{(x,\mathrm{y})\,|\,0<\mathrm{x}<\frac{1}{2},\frac{1}{4}<y<\frac{1}{2}\right\}$.

– Solution

(a) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy$

$=\int_0^1\int_0^1\frac{2}{5}(2x+3y)dxdy$

$=\int_0^1\left(\frac{2x^2}{5}+\frac{6xy}{5}\right)\Big|_{x=0}^{x=1}dy$

$=\int_0^1\left(\frac{2}{5}+\frac{6y}{5}\right)dy=\frac{2y}{5}+\frac{3y^2}{5}\Big|_0^1$

$=\frac{2}{5}+\frac{3}{5}=1.$

(b) $P[(X,Y)\in A]=P(0<X<\frac{1}{2},\frac{1}{4}<Y<\frac{1}{2})$

$=\int_{1/4}^{1/2}\int_0^{1/2}\frac{2}{5}(2x+3y)dxdy$

$=\int_{1/4}^{1/2}\left(\frac{2x^2}{5}+\frac{6xy}{5}\right)\Big|_{x=0}^{x=1/2}dy$

$=\int_{1/4}^{1/2}\left(\frac{1}{10}+\frac{3y}{5}\right)dy=\frac{y}{10}+\frac{3y^2}{10}\Big|_{1/4}^{1/2}$

$=\frac{1}{10}\left[\left(\frac{1}{2}+\frac{3}{4}\right)-\left(\frac{1}{4}+\frac{3}{16}\right)\right]=\frac{13}{160}.$

Joint Probability Distributions

- Definition 3.10: The <u>marginal distributions</u> of *X* alone and of *Y* alone are
$$\begin{cases} g(x) = \sum_y f(x,y) \text{ and } h(\mathrm{y}) = \sum_x f(x,y) \text{ for the discrete case} \\ g(x) = \int_{-\infty}^{\infty} f(x,y)dy \text{ and } h(\mathrm{y}) = \int_{-\infty}^{\infty} f(x,y)dx \text{ for the continuous case} \end{cases}$$

- Example 3.16: Show that the column and row totals of the following table give the marginal distribution of *X* alone and of *Y* alone.
  - $P(X=0) = g(0) = \sum_{y=0}^{2} f(0,y) = f(0,0) + f(0,1) + f(0,2)$
    $= \frac{3}{28} + \frac{3}{14} + \frac{1}{28} = \frac{5}{14}$
    $P(X=1) = g(1) = \sum_{y=0}^{2} f(1,y) = f(1,0) + f(1,1) + f(1,2)$
    $= \frac{9}{28} + \frac{3}{14} + 0 = \frac{15}{28}$
    $P(X=2) = g(2) = \sum_{y=0}^{2} f(2,y) = f(2,0) + f(2,1) + f(2,2)$
    $= \frac{3}{28} + 0 + 0 = \frac{3}{28}$

f(*x*, *y*)		*x* = 0	*x* = 1	*x* = 2	Row totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$		$\frac{3}{7}$
	2	$\frac{1}{28}$			$\frac{1}{28}$
Column totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

x	0	1	2
g(x)	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$

Joint Probability Distributions

- Example 3.17: Find *g*(x) and *h*(y) for the joint density function of Example 3.15.

$$f(x,y)=\begin{cases}\frac{2}{5}(2x+3y), & 0\le x\le 1, 0\le y\le 1\\ 0, & \text{elsewhere}\end{cases}$$

  - (1) $g(x)=\int_{-\infty}^{\infty} f(x,y)dy=\int_0^1 \frac{2}{5}(2x+3y)dy=\frac{4xy}{5}+\frac{6y^2}{10}\Big|_{y=0}^{y=1}=\frac{4x+3}{5}$

    for $0\le x\le 1$, and $g(x)=0$ elsewhere

    (2) $h(y)=\int_{-\infty}^{\infty} f(x,y)dx=\int_0^1 \frac{2}{5}(2x+3y)dx=\frac{2(1+3y)}{5}$

    for $0\le y\le 1$, and $h(y)=0$ elsewhere

Joint Probability Distributions

- Definition 3.11: Let *X* and *Y* be two random variables, discrete or continuous. The conditional distribution of the random variable *Y*, given that *X* = *x*, is $f(y \mid x) = \frac{f(x,y)}{g(x)},\ g(x) > 0.$

  Similarly, the conditional distribution of the random variable *X*, given that *Y* = *y*, is $f(x \mid y) = \frac{f(x,y)}{h(y)},\ h(y) > 0.$

- Evaluate the probability that *X* falls between *a* and *b* given that *Y* is known.

$$\begin{cases} P(a < X < b \mid Y = y) = \sum_{a<x<b} f(x \mid y), \text{ for the discrete case} \\ P(a < X < b \mid Y = y) = \int_a^b f(x \mid y)dx, \text{ for the continuous case} \end{cases}$$

Joint Probability Distributions

- Example 3.18: Referring to Example 3.14, find the conditional distribution of *X*, given that *Y* = 1, and use it to determine *P*(*X* = 0|*Y* = 1).
  - Solution

$$h(y=1)=\sum_{x=0}^{2} f(x,1)=\tfrac{3}{14}+\tfrac{3}{14}+0=\tfrac{3}{7}.$$

$$f(x|1)=\frac{f(x,1)}{h(1)}=\frac{7}{3}f(x,1),\ \ x=0,1,2$$

$$f(0|1)=\tfrac{7}{3}f(0,1)=\tfrac{7}{3}\times\tfrac{3}{14}=\tfrac{1}{2}$$

$$f(1|1)=\tfrac{7}{3}f(1,1)=\tfrac{7}{3}\times\tfrac{3}{14}=\tfrac{1}{2}$$

$$f(2|1)=\tfrac{7}{3}f(2,1)=\tfrac{7}{3}\times 0=0$$

$$\therefore P(X=0|Y=1)=f(0|1)=\tfrac{1}{2}$$

$f(x, y)$		x = 0	1	2	Row totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$		$\frac{3}{7}$
	2	$\frac{1}{28}$			$\frac{1}{28}$
Column totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

x	0	1	2
$f(x\|1)$	$\frac{1}{2}$	$\frac{1}{2}$	0

Joint Probability Distributions

- Example 3.19: The joint density for the random variables (*X* , *Y*), where X is the unit temperature change and Y is the proportion of spectrum shift that a certain atomic particle produces is

$$f(x,y)=\begin{cases}10xy^2, & 0<\text{x}<\text{y}<1\\ 0, & \text{elsewhere}\end{cases}$$

(a) Find the marginal densities g(x), h(y), and the conditional density f(y|x).
(b) Find the probability that the spectrum shifts more than half of the total observations, given the temperature is increased to 0.25 unit.

$$\text{(a) } g(x)=\int_{-\infty}^{\infty} f(x,y)dy=\int_x^1 10xy^2 dy=\frac{10}{3}xy^3\Big|_{y=x}^{y=1}.$$

$$h(y)=\int_{-\infty}^{\infty} f(x,y)dx=\int_0^y 10xy^2 dx$$

$$f(y\,|\,x)=\frac{f(x,y)}{g(x)}=\frac{10xy^2}{\frac{10}{3}x(1-x^3)}=\frac{3y^2}{(1-x^3)}$$

$$\text{(b) } P(Y>\tfrac{1}{2}\,|\,X=0.25)=\int_{1/2}^1 f(y\,|\,x=0.25)dy=\int_{1/2}^1 \frac{3y^2}{(1-0.25^3)}dy=\frac{8}{9}$$

Joint Probability Distributions

- Definition 3.12: Let *X* and *Y* be two random variables, discrete or continuous, with joint probability distribution f(x, y) and marginal distributions g(x) and h(y), respectively. The random variables *X* and *Y* are said to be statistically independent if and only if

$$f(x, y) = g(x)h(y), \text{ for all } (x, y) \text{ within their range.}$$

- Example 3.21: Show that the random variables of Example 3.14 are not statistically independent.

$$f(0,1) = \frac{3}{14}$$

$$g(0) = \sum_{y=0}^{2} f(0, y) = \frac{3}{28} + \frac{3}{14} + \frac{1}{28} = \frac{5}{14}$$

$$h(1) = \sum_{x=0}^{2} f(x,1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}$$

$$\therefore\ f(0,1) \neq g(0)h(1)$$

therefore X and Y are not statistically independent.

Joint Probability Distributions

- Definition 3.13: Let $X_1, X_2, \ldots, X_n$ be n random variables, discrete or continuous, with joint probability distribution $f(x_1, x_2, \ldots, x_n)$ and marginal distributions $f(x_1), f(x_2), \ldots, f(x_n)$, respectively. The random variables $X_1, X_2, \ldots, X_n$ are said to be mutually statistically independent if and only if
$$f(x_1, x_2, \cdots, x_n) = f_1(x_1) f_2(x_2) \cdots f_n(x_n)$$
for all $(x_1, x_2, \cdots, x_n)$ within their range.

- Example 3.22: Suppose that the shelf life, in years, of a certain perishable (易腐爛的) food product packaged in cardboard containers is a random variable whose probability density function is given by

$$f(x) = \begin{cases} e^{-x}, & \mathrm{x} > 0 \\ 0, & \text{elsewhere} \end{cases}$$

Joint Probability Distributions

$$f(x) = \begin{cases} e^{-x}, & \text{x} > 0 \\ 0, & \text{elsewhere} \end{cases}$$

– Let $X_1, X_2, \ldots, X_n$ represent the shelf lives for three of these containers selected <u>independently</u> and find $P(X_1<2, 1< X_2<3, X_3>2)$

– Solution

$$f(x_1, x_2, x_3) = f(x_1) f(x_2) f(x_3) = e^{-x_1} e^{-x_2} e^{-x_3} = e^{-x_1-x_2-x_3}$$

$$P(X_1 < 2, 1 < X_2 < 3, X_3 > 2) = \int_2^\infty \int_1^3 \int_0^2 e^{-x_1-x_2-x_3} dx_1 dx_2 dx_3$$

$$= (1-e^{-2})(e^{-1} - e^{-3})e^{-2} = 0.0372$$