Example 1.3: As an example, consider the following function:

$$f(x) = \begin{cases} 1, & \text{for } 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Clearly, since $f(x) \geq 0$ for $-\infty < x < \infty$ and $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \int_0^1 f(x)\,\mathrm{d}x = [x]_0^1 = 1$, the above function can be a probability density function.

In fact, it is called a **uniform distribution**.

Example 1.4: As another example, consider the following function:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2},$$

for $-\infty < x < \infty$.

Clearly, we have $f(x) \geq 0$ for all x.

We check whether $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$.

First of all, we define I as $I = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x$.

To show $I = 1$, we may prove $I^2 = 1$ because of $f(x) > 0$ for all x, which is shown as follows:

$$\begin{aligned}
I^2 &= \Big(\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x\Big)^2 = \Big(\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x\Big)\Big(\int_{-\infty}^{\infty} f(y)\,\mathrm{d}y\Big) \\
&= \Big(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2)\,\mathrm{d}x\Big)\Big(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}y^2)\,\mathrm{d}y\Big) \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\Big(-\frac{1}{2}(x^2+y^2)\Big)\,\mathrm{d}x\,\mathrm{d}y \\
&= \frac{1}{2\pi} \int_0^{2\pi}\int_0^{\infty} \exp(-\frac{1}{2}r^2) r\,\mathrm{d}r\,\mathrm{d}\theta \\
&= \frac{1}{2\pi} \int_0^{2\pi}\int_0^{\infty} \exp(-s)\,\mathrm{d}s\,\mathrm{d}\theta = \frac{1}{2\pi} 2\pi[-\exp(-s)]_0^{\infty} = 1.
\end{aligned}$$

< Review > Integration by Substitution (置換積分):

Univariate (1 変数) Case: For a function of x, $f(x)$, we perform integration by substitution, using $x = \psi(y)$.

Then, it is easy to obtain the following formula:

$$\int f(x)\,\mathrm{d}x = \int \psi'(y) f(\psi(y))\,\mathrm{d}y,$$

which formula is called the **integration by substitution**.

Proof:

Let $F(x)$ be the integration of $f(x)$, i.e.,

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t,$$

which implies that $F'(x) = f(x)$.

Differentiating $F(x) = F(\psi(y))$ with respect to y, we have:

$$f(x) \equiv \frac{\mathrm{d}F(\psi(y))}{\mathrm{d}y} = \frac{\mathrm{d}F(x)}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}y} = f(x)\psi'(y) = f(\psi(y))\psi'(y).$$

Bivariate (2 変数) Case: For $f(x, y)$, define $x = \psi_1(u, v)$ and $y = \psi_2(u, v)$.

$$\iint f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \iint J f(\psi_1(u,v), \psi_2(u,v))\,\mathrm{d}u\,\mathrm{d}v,$$

where J is called the **Jacobian (ヤコビアン)**, which represents the following determinant (行列式):

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

< End of Review >

＜ Go back to the Integration ＞

In the fifth equality, integration by substitution (置換積分) is used.

The polar coordinate transformation (極座標変換) is used as $x = r\cos\theta$ and $y = r\sin\theta$.

Note that $0 \le r < +\infty$ and $0 \le \theta < 2\pi$.

The Jacobian is given by:

$$J = \begin{vmatrix} \dfrac{\partial x}{\partial r} & \dfrac{\partial x}{\partial \theta} \\ \dfrac{\partial y}{\partial r} & \dfrac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r.$$

In the inner integration of the sixth equality, again, integration by substitution is utilized, where transformation is $s = \dfrac{1}{2}r^2$.

Thus, we obtain the result $I^2 = 1$ and accordingly we have $I = 1$ because of $f(x) \ge 0$.

Therefore, $f(x) = e^{-\frac{1}{2}x^2}/\sqrt{2\pi}$ is also taken as a probability density function.

Actually, this density function is called the **standard normal probability density function (標準正規分布)**.

Distribution Function: The **distribution function (分布関数)** or the **cumulative distribution function (累積分布関数)**, denoted by $F(x)$, is defined as:

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

which represents the probability less than x.

The properties of the distribution function $F(x)$ are given by:

$F(x_1) \le F(x_2)$, for $x_1 < x_2$, —> nondecreasing function

$P(a < X \le b) = F(b) - F(a)$, for $a < b$,

$F(-\infty) = 0$, $F(+\infty) = 1$.

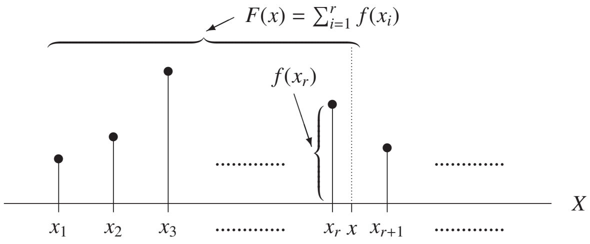
The difference between the discrete and continuous random variables is given by:

1. Discrete random variable (Figure 1):
   - $F(x) = \displaystyle\sum_{i=1}^{r} f(x_i) = \sum_{i=1}^{r} p_i$,
     where r denotes the integer which satisfies $x_r \le x < x_{r+1}$.
   - $F(x_i) - F(x_i - \epsilon) = f(x_i) = p_i$,
     where ϵ is a small positive number less than $x_i - x_{i-1}$.

2. Continuous random variable (Figure 2):
   - $F(x) = \displaystyle\int_{-\infty}^{x} f(t)\,\mathrm{d}t$,
   - $F'(x) = f(x)$.

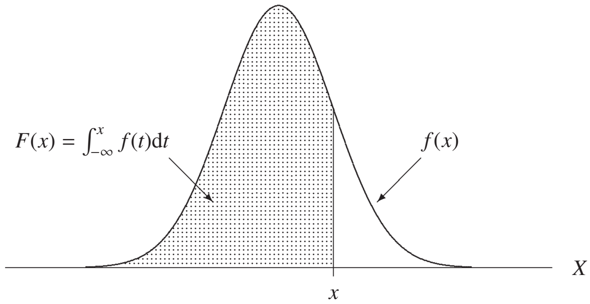
$f(x)$ and $F(x)$ are displayed in Figure 1 for a discrete random variable and Figure 2 for a continuous random variable.

Figure 1: Probability Function $f(x)$ and Distribution Function $F(x)$— Discrete Case



Note that r is the integer which satisfies $x_r \leq x < x_{r+1}$.

Figure 2: Density Function $f(x)$ and Distribution Function $F(x)$ — Continuous Case



2.2 Multivariate Random Variable (多変量確率変数) and Distribution

We consider two random variables X and Y in this section. It is easy to extend to more than two random variables.

Discrete Random Variables: Suppose that discrete random variables X and Y take $x_1, x_2, \cdots$ and $y_1, y_2, \cdots$, respectively. The probability which event $\{\omega; X(\omega) = x_i$ and $Y(\omega) =$

$y_j\}$ occurs is given by:

$$P(X = x_i, Y = y_j) = f_{xy}(x_i, y_j),$$

where $f_{xy}(x_i, y_j)$ represents the **joint probability function (結合確率関数)** of X and Y. In order for $f_{xy}(x_i, y_j)$ to be a joint probability function, $f_{xy}(x_i, y_j)$ has to satisfies the following properties:

$$f_{xy}(x_i, y_j) \geq 0, \quad i, j = 1, 2, \cdots$$
$$\sum_i \sum_j f_{xy}(x_i, y_j) = 1.$$

Define $f_x(x_i)$ and $f_y(y_j)$ as:

$$f_x(x_i) = \sum_j f_{xy}(x_i, y_j), \quad i = 1, 2, \cdots,$$
$$f_y(y_j) = \sum_i f_{xy}(x_i, y_j), \quad j = 1, 2, \cdots.$$

Then, $f_x(x_i)$ and $f_y(y_j)$ are called the **marginal probability functions (周辺確率関数)** of X and Y.

$f_x(x_i)$ and $f_y(y_j)$ also have the properties of the probability functions, i.e.,

$f_x(x_i) \geq 0$ and $\sum_i f_x(x_i) = 1$, and $f_y(y_j) \geq 0$ and $\sum_j f_y(y_j) = 1$.

Continuous Random Variables: Consider two continuous random variables X and Y. For a domain D, the probability which event $\{\omega; (X(\omega), Y(\omega)) \in D\}$ occurs is given by:

$$P((X, Y) \in D) = \iint_D f_{xy}(x, y)\, \mathrm{d}x\, \mathrm{d}y,$$

where $f_{xy}(x, y)$ is called the **joint probability density function (結合確率密度関数)** of X and Y or the **joint density function** of X and Y.

$f_{xy}(x, y)$ has to satisfy the following properties:

$$f_{xy}(x, y) \geq 0,$$
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{xy}(x, y)\,\mathrm{d}x\,\mathrm{d}y = 1.$$

Define $f_x(x)$ and $f_y(y)$ as:

$$f_x(x) = \int_{-\infty}^{\infty} f_{xy}(x, y)\,\mathrm{d}y, \qquad \text{for all } x \text{ and } y,$$
$$f_y(y) = \int_{-\infty}^{\infty} f_{xy}(x, y)\,\mathrm{d}x,$$

where $f_x(x)$ and $f_y(y)$ are called the **marginal probability**

density functions (周辺確率密度関数) of X and Y or the **marginal density functions (周辺密度関数)** of X and Y.

For example, consider the event $\{\omega; a < X(\omega) < b,\ c < Y(\omega) < d\}$, which is a specific case of the domain D. Then, the probability that we have the event $\{\omega; a < X(\omega) < b,\ c < Y(\omega) < d\}$ is written as:

$$P(a < X < b,\ c < Y < d) = \int_a^b\int_c^d f_{xy}(x, y)\,\mathrm{d}x\,\mathrm{d}y.$$

The mixture of discrete and continuous RVs is also possible. For example, let X be a discrete RV and Y be a continuous RV. X takes $x_1,\ x_2,\ \cdots$. The probability which both X takes x_i and Y takes real numbers within the interval I is given by:

$$P(X = x_i, Y \in I) = \int_I f_{xy}(x_i, y)\,\mathrm{d}y.$$

Then, we have the following properties:

$$f_{xy}(x_i, y) \geq 0, \quad \text{for all } y \text{ and } i = 1, 2, \cdots,$$
$$\sum_i \int_{-\infty}^{\infty} f_{xy}(x_i, y)\,\mathrm{d}y = 1.$$

The marginal probability function of X is given by:

$$f_x(x_i) = \int_{-\infty}^{\infty} f_{xy}(x_i, y)\,\mathrm{d}y,$$

for $i = 1, 2, \cdots$. The marginal probability density function of Y is:

$$f_y(y) = \sum_i f_{xy}(x_i, y).$$

2.3 Conditional Distribution

Discrete Random Variable: The **conditional probability function (条件付確率関数)** of X given $Y = y_j$ is represented as:

$$P(X = x_i|Y = y_j) = f_{x|y}(x_i|y_j) = \frac{f_{xy}(x_i, y_j)}{f_y(y_j)} = \frac{f_{xy}(x_i, y_j)}{\sum_i f_{xy}(x_i, y_j)}.$$

The second equality indicates the definition of the conditional probability.

The features of the conditional probability function $f_{x|y}(x_i|y_j)$ are:

$$f_{x|y}(x_i|y_j) \geq 0, \quad i = 1, 2, \cdots,$$
$$\sum_i f_{x|y}(x_i|y_j) = 1, \quad \text{for any } j.$$

Continuous Random Variable: The **conditional probability density function (条件付確率密度関数)** of X given $Y = y$ (or the **conditional density function (条件付密度関数)** of X given $Y = y$) is:

$$f_{x|y}(x|y) = \frac{f_{xy}(x, y)}{f_y(y)} = \frac{f_{xy}(x, y)}{\int_{-\infty}^{\infty} f_{xy}(x, y)\,\mathrm{d}x}.$$

83

The properties of the conditional probability density function $f_{x|y}(x|y)$ are given by:

$$f_{x|y}(x|y) \geq 0,$$
$$\int_{-\infty}^{\infty} f_{x|y}(x|y)\,\mathrm{d}x = 1, \quad \text{for any } Y = y.$$

84

Independence of Random Variables: For discrete random variables X and Y, we say that X is **independent (独立)** (or **stochastically independent (確率的に独立)**) of Y if and only if $f_{xy}(x_i, y_j) = f_x(x_i) f_y(y_j)$.

Similarly, for continuous random variables X and Y, we say that X is independent of Y if and only if $f_{xy}(x, y) = f_x(x) f_y(y)$.

When X and Y are stochastically independent, $g(X)$ and $h(Y)$ are also stochastically independent, where $g(X)$ and $h(Y)$ are functions of X and Y.

85