Some Formulas of Mean and Variance: We consider two random variables X and Y.

1. **Theorem:** $\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$.

   Proof:

   For discrete random variables X and Y, it is given by:

$$
\begin{aligned}
\mathrm{E}(X+Y) &= \sum_i \sum_j (x_i + y_j) f_{xy}(x_i, y_j) \\
&= \sum_i \sum_j x_i f_{xy}(x_i, y_j) + \sum_i \sum_j y_j f_{xy}(x_i, y_j) \\
&= \mathrm{E}(X) + \mathrm{E}(Y).
\end{aligned}
$$

For continuous random variables X and Y, we can show:

$$
\begin{aligned}
\mathrm{E}(X+Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y) f_{xy}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{xy}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&\quad + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{xy}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \mathrm{E}(X) + \mathrm{E}(Y).
\end{aligned}
$$

2. **Theorem:** $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$, when X is independent of Y.

   Proof:

   For discrete random variables X and Y,

$$
\begin{aligned}
\mathrm{E}(XY) &= \sum_i \sum_j x_i y_j f_{xy}(x_i, y_j) = \sum_i \sum_j x_i y_j f_x(x_i) f_y(y_j) \\
&= \Big(\sum_i x_i f_x(x_i)\Big)\Big(\sum_j y_j f_y(y_j)\Big) = \mathrm{E}(X)\mathrm{E}(Y).
\end{aligned}
$$

   If X is independent of Y, the second equality holds, i.e., $f_{xy}(x_i, y_j) = f_x(x_i) f_y(y_j)$.

For continuous random variables X and Y,

$$
\begin{aligned}
\mathrm{E}(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{xy}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_x(x) f_y(y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \Big(\int_{-\infty}^{\infty} x f_x(x)\,\mathrm{d}x\Big)\Big(\int_{-\infty}^{\infty} y f_y(y)\,\mathrm{d}y\Big) = \mathrm{E}(X)\mathrm{E}(Y).
\end{aligned}
$$

When X is independent of Y, we have $f_{xy}(x,y) = f_x(x) f_y(y)$ in the second equality.

3. **Theorem:** $\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$.

   Proof:

   For both discrete and continuous random variables, we can rewrite as follows:

$$
\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathrm{E}((X-\mu_x)(Y-\mu_y)) \\
&= \mathrm{E}(XY - \mu_x Y - \mu_y X + \mu_x\mu_y) \\
&= \mathrm{E}(XY) - \mathrm{E}(\mu_x Y) - \mathrm{E}(\mu_y X) + \mu_x\mu_y \\
&= \mathrm{E}(XY) - \mu_x \mathrm{E}(Y) - \mu_y \mathrm{E}(X) + \mu_x\mu_y
\end{aligned}
$$

$$
\begin{aligned}
&= \mathrm{E}(XY) - \mu_x\mu_y - \mu_y\mu_x + \mu_x\mu_y \\
&= \mathrm{E}(XY) - \mu_x\mu_y \\
&= \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y).
\end{aligned}
$$

In the fourth equality, the theorem in Section 3.1 is used, i.e., $\mathrm{E}(\mu_x Y) = \mu_x \mathrm{E}(Y)$ and $\mathrm{E}(\mu_y X) = \mu_y \mathrm{E}(X)$.

4. **Theorem:** $\mathrm{Cov}(X, Y) = 0$, when X is independent of Y.

Proof:

From the above two theorems, we have $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$ when X is independent of Y and $\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$.

Therefore, $\mathrm{Cov}(X, Y) = 0$ is obtained when X is independent of Y.

5. **Definition:** The **correlation coefficient (相関係数)** between X and Y, denoted by ρ_{xy}, is defined as:

$$\rho_{xy} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{V}(X)}\sqrt{\mathrm{V}(Y)}} = \frac{\mathrm{Cov}(X, Y)}{\sigma_x \sigma_y}.$$

$\rho_{xy} > 0 \Longrightarrow$ **positive correlation** between X and Y

$\rho_{xy} \longrightarrow 1 \Longrightarrow$ **strong positive correlation**

$\rho_{xy} < 0 \Longrightarrow$ **negative correlation** between X and Y

$\rho_{xy} \longrightarrow -1 \Longrightarrow$ **strong negative correlation**

6. **Theorem:** $\rho_{xy} = 0$, when X is independent of Y.

Proof:

When X is independent of Y, we have $\mathrm{Cov}(X, Y) = 0$.

We obtain the result $\rho_{xy} = \dfrac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{V}(X)}\sqrt{\mathrm{V}(Y)}} = 0$.

However, note that $\rho_{xy} = 0$ does not mean the independence between X and Y.

7. **Theorem:** $\mathrm{V}(X \pm Y) = \mathrm{V}(X) \pm 2\mathrm{Cov}(X, Y) + \mathrm{V}(Y)$.

Proof:

For both discrete and continuous random variables, $\mathrm{V}(X \pm Y)$ is rewritten as follows:

$$\begin{aligned}
\mathrm{V}(X \pm Y) &= \mathrm{E}\Big(\big((X \pm Y) - \mathrm{E}(X \pm Y)\big)^2\Big) \\
&= \mathrm{E}\Big(\big((X - \mu_x) \pm (Y - \mu_y)\big)^2\Big) \\
&= \mathrm{E}\big((X - \mu_x)^2 \pm 2(X - \mu_x)(Y - \mu_y) + (Y - \mu_y)^2\big)
\end{aligned}$$

$$\begin{aligned}
&= \mathrm{E}((X - \mu_x)^2) \pm 2\mathrm{E}((X - \mu_x)(Y - \mu_y)) \\
&\qquad + \mathrm{E}((Y - \mu_y)^2) \\
&= \mathrm{V}(X) \pm 2\mathrm{Cov}(X, Y) + \mathrm{V}(Y).
\end{aligned}$$

8. **Theorem:** $-1 \le \rho_{xy} \le 1$.

Proof:

Consider the following function of t: $f(t) = \mathrm{V}(Xt - Y)$, which is always greater than or equal to zero because of the definition of variance. Therefore, for all t, we have $f(t) \ge 0$. $f(t)$ is rewritten as follows:

$$
\begin{aligned}
f(t) &= \mathrm{V}(Xt - Y) = \mathrm{V}(Xt) - 2\mathrm{Cov}(Xt, Y) + \mathrm{V}(Y) \\
&= t^2\mathrm{V}(X) - 2t\mathrm{Cov}(X, Y) + \mathrm{V}(Y) \\
&= \mathrm{V}(X)\Big(t - \frac{\mathrm{Cov}(X, Y)}{\mathrm{V}(X)}\Big)^2 + \mathrm{V}(Y) - \frac{(\mathrm{Cov}(X, Y))^2}{\mathrm{V}(X)}.
\end{aligned}
$$

In order to have $f(t) \geq 0$ for all t, we need the following condition:

$$
\mathrm{V}(Y) - \frac{(\mathrm{Cov}(X, Y))^2}{\mathrm{V}(X)} \geq 0,
$$

because the first term in the last equality is nonnega-

tive, which implies:

$$
\frac{(\mathrm{Cov}(X, Y))^2}{\mathrm{V}(X)\mathrm{V}(Y)} \leq 1.
$$

Therefore, we have:

$$
-1 \leq \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{V}(X)}\sqrt{\mathrm{V}(Y)}} \leq 1.
$$

From the definition of correlation coefficient, i.e., $\rho_{xy} = \dfrac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{V}(X)}\sqrt{\mathrm{V}(Y)}}$, we obtain the result: $-1 \leq \rho_{xy} \leq 1$.

9. **Theorem:** $\mathrm{V}(X \pm Y) = \mathrm{V}(X) + \mathrm{V}(Y)$, when X is independent of Y.

Proof:

From the theorem above, $\mathrm{V}(X \pm Y) = \mathrm{V}(X) \pm 2\mathrm{Cov}(X, Y) + \mathrm{V}(Y)$ generally holds. When random variables X and Y are independent, we have $\mathrm{Cov}(X, Y) = 0$. Therefore, $\mathrm{V}(X + Y) = \mathrm{V}(X) + \mathrm{V}(Y)$ holds, when X is independent of Y.

10. **Theorem:** For n random variables $X_1, X_2, \cdots, X_n$,

$$
\begin{aligned}
\mathrm{E}(\sum_i a_i X_i) &= \sum_i a_i \mu_i, \\
\mathrm{V}(\sum_i a_i X_i) &= \sum_i \sum_j a_i a_j \mathrm{Cov}(X_i, X_j),
\end{aligned}
$$

where $\mathrm{E}(X_i) = \mu_i$ and a_i is a constant value. Especially, when $X_1, X_2, \cdots, X_n$ are mutually independent, we have the following:

$$
\mathrm{V}(\sum_i a_i X_i) = \sum_i a_i^2 \mathrm{V}(X_i).
$$

Proof:

For mean of $\sum_i a_i X_i$, the following representation is obtained.

$$
\mathrm{E}(\sum_i a_i X_i) = \sum_i \mathrm{E}(a_i X_i) = \sum_i a_i \mathrm{E}(X_i) = \sum_i a_i \mu_i.
$$

The first and second equalities come from the previous theorems on mean.

For variance of $\sum_i a_i X_i$, we can rewrite as follows:

$$
\begin{aligned}
\mathrm{V}(\sum_i a_i X_i) &= \mathrm{E}\Big(\sum_i a_i (X_i - \mu_i)\Big)^2 \\
&= \mathrm{E}\Big(\sum_i a_i (X_i - \mu_i)\Big)\Big(\sum_j a_j (X_j - \mu_j)\Big) \\
&= \mathrm{E}\Big(\sum_i \sum_j a_i a_j (X_i - \mu_i)(X_j - \mu_j)\Big) \\
&= \sum_i \sum_j a_i a_j \mathrm{E}\Big((X_i - \mu_i)(X_j - \mu_j)\Big) \\
&= \sum_i \sum_j a_i a_j \mathrm{Cov}(X_i, X_j).
\end{aligned}
$$

When $X_1, X_2, \cdots, X_n$ are mutually independent, we obtain $\mathrm{Cov}(X_i, X_j) = 0$ for all $i \neq j$ from the previous theorem. Therefore, we obtain:

$$\mathrm{V}(\sum_i a_i X_i) = \sum_i a_i^2 \mathrm{V}(X_i).$$

Note that $\mathrm{Cov}(X_i, X_i) = \mathrm{E}((X_i - \mu)^2) = \mathrm{V}(X_i)$.

11. **Theorem:** n random variables $X_1, X_2, \cdots, X_n$ are mutually independently and identically distributed with mean μ and variance σ^2. That is, for all $i = 1, 2, \cdots, n$, $\mathrm{E}(X_i) = \mu$ and $\mathrm{V}(X_i) = \sigma^2$ are assumed. Consider arithmetic average $\overline{X} = (1/n)\sum_{i=1}^n X_i$. Then, mean and variance of $\overline{X}$ are given by:

$$\mathrm{E}(\overline{X}) = \mu, \qquad \mathrm{V}(\overline{X}) = \frac{\sigma^2}{n}.$$

Proof:

The mathematical expectation of $\overline{X}$ is given by:

$$\begin{aligned}\mathrm{E}(\overline{X}) &= \mathrm{E}(\frac{1}{n}\sum_{i=1}^n X_i) = \frac{1}{n}\mathrm{E}(\sum_{i=1}^n X_i) = \frac{1}{n}\sum_{i=1}^n \mathrm{E}(X_i) \\ &= \frac{1}{n}\sum_{i=1}^n \mu = \frac{1}{n} n\mu = \mu.\end{aligned}$$

$\mathrm{E}(aX) = a\mathrm{E}(X)$ in the second equality and $\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$ in the third equality are utilized, where X and Y are random variables and a is a constant value.

The variance of $\overline{X}$ is computed as follows:

$$\begin{aligned}\mathrm{V}(\overline{X}) &= \mathrm{V}(\frac{1}{n}\sum_{i=1}^n X_i) = \frac{1}{n^2}\mathrm{V}(\sum_{i=1}^n X_i) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{V}(X_i) \\ &= \frac{1}{n^2}\sum_{i=1}^n \sigma^2 = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.\end{aligned}$$

We use $\mathrm{V}(aX) = a^2\mathrm{V}(X)$ in the second equality and $\mathrm{V}(X+Y) = \mathrm{V}(X) + \mathrm{V}(Y)$ for X independent of Y in the third equality, where X and Y denote random variables and a is a constant value.

4 Transformation of Variables (変数変換)

Transformation of variables is used in the case of continuous random variables. Based on a distribution of a random variable, a distribution of the transformed random variable is derived. In other words, when a distribution of X is known, we can find a distribution of Y using the transformation of variables, where Y is a function of X.

4.1 Univariate Case

Distribution of $Y = \psi^{-1}(X)$: Let $f_x(x)$ be the probability density function of continuous random variable X and $X = \psi(Y)$ be a one-to-one (一対一) transformation. Then, the probability density function of Y, i.e., $f_y(y)$, is given by:

$$f_y(y) = |\psi'(y)| f_x\big(\psi(y)\big).$$

We can derive the above transformation of variables from X to Y as follows. Let $f_x(x)$ and $F_x(x)$ be the probability density function and the distribution function of X, respectively. Note that $F_x(x) = P(X \le x)$ and $f_x(x) = F'_x(x)$.

When $X = \psi(Y)$, we want to obtain the probability density function of Y. Let $f_y(y)$ and $F_y(y)$ be the probability density function and the distribution function of Y, respectively.

In the case of $\psi'(X) > 0$, the distribution function of Y, $F_y(y)$, is rewritten as follows:

$$\begin{aligned} F_y(y) &= P(Y \le y) = P\big(\psi(Y) \le \psi(y)\big) \\ &= P\big(X \le \psi(y)\big) = F_x\big(\psi(y)\big). \end{aligned}$$

The first equality is the definition of the cumulative distribution function. The second equality holds because of $\psi'(Y) > 0$. Therefore, differentiating $F_y(y)$ with respect to y, we can obtain the following expression:

$$f_y(y) = F'_y(y) = \psi'(y)F'_x\big(\psi(y)\big) = \psi'(y)f_x\big(\psi(y)\big). \qquad (4)$$

Next, in the case of $\psi'(X) < 0$, the distribution function of Y, $F_y(y)$, is rewritten as follows:

$$\begin{aligned} F_y(y) &= P(Y \le y) = P\big(\psi(Y) \ge \psi(y)\big) = P\big(X \ge \psi(y)\big) \\ &= 1 - P\big(X < \psi(y)\big) = 1 - F_x\big(\psi(y)\big). \end{aligned}$$

Thus, in the case of $\psi'(X) < 0$, pay attention to the second equality, where the inequality sign is reversed. Differentiating $F_y(y)$ with respect to y, we obtain the following result:

$$f_y(y) = F'_y(y) = -\psi'(y)F'_x\big(\psi(y)\big) = -\psi'(y)f_x\big(\psi(y)\big). \qquad (5)$$

Note that $-\psi'(y) > 0$.

Thus, summarizing the above two cases, i.e., $\psi'(X) > 0$ and $\psi'(X) < 0$, equations (4) and (5) indicate the following result:

$$f_y(y) = |\psi'(y)|f_x\big(\psi(y)\big),$$

which is called the **transformation of variables**.

Example 1.9: When $X \sim N(0,1)$, we derive the probability density function of $Y = \mu + \sigma X$.

Since we have:

$$X = \psi(Y) = \frac{Y - \mu}{\sigma},$$

$\psi'(y) = 1/\sigma$ is obtained. Therefore, $f_y(y)$ is given by:

$$f_y(y) = |\psi'(y)|f_x\big(\psi(y)\big) = \frac{1}{\sigma\sqrt{2\pi}}\exp\Big(-\frac{1}{2\sigma^2}(y-\mu)^2\Big),$$

which indicates the normal distribution with mean μ and variance σ^2, denoted by $N(\mu, \sigma^2)$.

On Distribution of $Y = X^2$: As an example, when we know the distribution function of X as $F_x(x)$, we want to obtain the distribution function of Y, $F_y(y)$, where $Y = X^2$. Using $F_x(x)$, $F_y(y)$ is rewritten as follows:

$$\begin{aligned} F_y(y) &= P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) \\ &= F_x(\sqrt{y}) - F_x(-\sqrt{y}). \end{aligned}$$

The probability density function of Y is obtained as follows:

$$f_y(y) = F'_y(y) = \frac{1}{2\sqrt{y}}\Big(f_x(\sqrt{y}) + f_x(-\sqrt{y})\Big).$$

4.2 Multivariate Cases

Bivariate Case: Let $f_{xy}(x, y)$ be a joint probability density function of X and Y. Let $X = \psi_1(U, V)$ and $Y = \psi_2(U, V)$ be a one-to-one transformation from (X, Y) to (U, V). Then, we obtain a joint probability density function of U and V, denoted by $f_{uv}(u, v)$, as follows:

$$f_{uv}(u, v) = |J| f_{xy}\Big(\psi_1(u, v), \psi_2(u, v)\Big),$$

where J is called the **Jacobian** of the transformation, which is defined as:

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

Multivariate Case: Let $f_x(x_1, x_2, \cdots, x_n)$ be a joint probability density function of X_1, X_2, $\cdots X_n$. Suppose that a one-to-one transformation from $(X_1, X_2, \cdots, X_n)$ to $(Y_1, Y_2, \cdots, Y_n)$ is given by:

$$\begin{aligned} X_1 &= \psi_1(Y_1, Y_2, \cdots, Y_n), \\ X_2 &= \psi_2(Y_1, Y_2, \cdots, Y_n), \\ &\vdots \\ X_n &= \psi_n(Y_1, Y_2, \cdots, Y_n). \end{aligned}$$

Then, we obtain a joint probability density function of Y_1, Y_2, $\cdots$, Y_n, denoted by $f_y(y_1, y_2, \cdots, y_n)$, as follows:

$$\begin{aligned} &f_y(y_1, y_2, \cdots, y_n) \\ &\quad = |J| f_x\Big(\psi_1(y_1, \cdots, y_n), \psi_2(y_1, \cdots, y_n), \cdots, \psi_n(y_1, \cdots, y_n)\Big), \end{aligned}$$

where J is called the Jacobian of the transformation, which is defined as:

$$J = \begin{vmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_1}{\partial y_2} & \cdots & \dfrac{\partial x_1}{\partial y_n} \\ \\ \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_2}{\partial y_2} & \cdots & \dfrac{\partial x_2}{\partial y_n} \\ \\ \vdots & \vdots & \ddots & \vdots \\ \\ \dfrac{\partial x_n}{\partial y_1} & \dfrac{\partial x_n}{\partial y_2} & \cdots & \dfrac{\partial x_n}{\partial y_n} \end{vmatrix}.$$