Lecture 4
Multivariate Distributions

Chao Song

College of Ecology

Lanzhou University

September 29, 2025

Multivariate distribution

In many practical cases, it is possible, and often desirable, to take more than one measurement of a random observation. Moreover, we sometimes want to use these measurements to predict a third one. For example, we measure the GPA and extracurriculum activities of a student, and we give each of them a comprehensive evaluation score.

Definition: Let X and Y be two discrete random variables. Let S denote the two-dimensional space of X and Y. The probability that $X = x$ and $Y = y$ is denoted by $f(x, y) = P(X = x, Y = y)$. The function $f(x, y)$ is called the joint probability mass function.

Joint probability mass function

Example: Roll a pair of fair dice. For each of the 36 sampling points with probability $1/36$, let X denote the smaller and Y the larger outcome on the dice. For example, if the outcome is $(3, 2)$, then the observed values are $X = 2,\ Y = 3$. What is the joint PMF of X and Y?

The event $X = 2,\ Y = 3$ can happen in one of two ways $(2, 3)$ or $(3, 2)$. So its probability is $2/36$. However, for event such as $X = 2,\ Y = 2$, it can only happen in one way. Thus, in general, the joint probability mass function is

$$f(x, y) = \begin{cases} \frac{1}{36} & x = y \\ \frac{1}{18} & x \neq y \end{cases}$$

Multinomial distribution

Suppose we have three mutually exclusive and exhaustive ways for an experiment to end: perfect, seconds, and defective. We repeat the experiment n independent times and the probability p_X, p_Y, $1 - p_X - p_Y$ of the three type of results. Let X and Y be the number of perfect and seconds. What is the joint probability mass function of X and Y?

The probability of having x perfects, y seconds, and $n - x - y$ defective is

$$p_X^x p_Y^y (1 - p_X - p_Y)^{n-x-y}$$

And it can be achieved in

$$\mathbf{C}_n^x \mathbf{C}_{n-x}^y = \frac{n!}{x!(n-x)!} \frac{(n-x)!}{y!(n-x-y)!} = \frac{n!}{x!y!(n-x-y)!}$$

Thus, the joint PMF is

$$f(x,y) = \frac{n!}{x!y!(n-x-y)!} p_X^x p_Y^y (1 - p_X - p_Y)^{n-x-y}$$

Marginal probability mass function

Let X and Y have the joint probability mass function $f(x, y)$ with space S. The probability mass function of X alone is called the marginal probability mass function of X and is defined by

$$f_X(x) = \sum_y f(x, y) \;\; x \in S_X$$

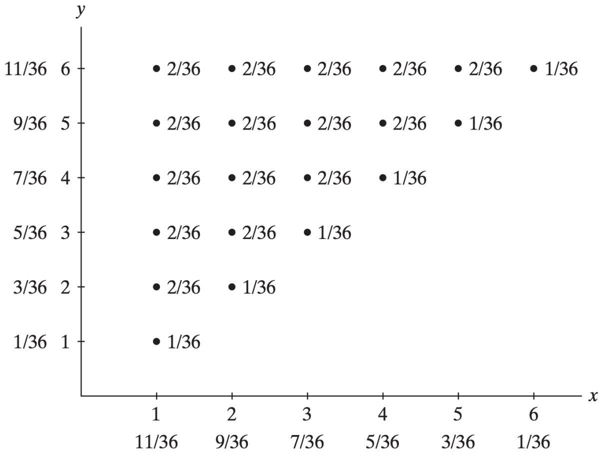
The random variables X and Y are independent if and only if, for every $x \in S_X$ and $y \in S_Y$,

$$f(x, y) = f_X(x) f_Y(y)$$

Otherwise, X ad Y are said to be dependent.

Marginal probability mass function

Example: In the dice rolling example mentioned above, what is the marginal probability mass function of X and Y? Are X and Y independent?



Marginal probability mass function

If X and Y has a multinomial distribution, are they independent?

It is easy to see by logic that X and Y both have a binomial distribution.

$$f_X(x) = \mathbf{C}_n^x p_X^x (1 - p_X)^{n-x}$$

$$f_Y(y) = \mathbf{C}_n^y p_Y^y (1 - p_Y)^{n-y}$$

Therefore,

$$f_X(x) f_Y(y) = \mathbf{C}_n^x \mathbf{C}_n^y p_X^x (1 - p_X)^{n-x} p_Y^y (1 - p_Y)^{n-y} \neq f(xy)$$

Thus, X and Y are not indepenent.

Mathematical expectation

Let X_1 and X_2 be random variables of the discrete type with the joint PMF $f(x_1, x_2)$ on the space S. If $u(X_1, X_2)$ is a function of these two random variables, then

$$E\left[u(X_1, X_2)\right] = \sum_{(x_1,x_2)\in S}\sum u(x_1, x_2) f(x_1, x_2)$$

if it exists, is called the mathematical expectation of $u(X_1, X_2)$.

If $u(X_1, X_2) = X_i$, then $E[u(X_1, X_2)] = E(X_i) = \mu_i$; if $u(X_1, X_2) = (X_i - \mu_i)^2$, then $E[u(X_1, X_2)] = E[(X_i - \mu_i)^2] = Var(X_i)$

Mathematical expectation

Example: There are eight chips in a bow: three marked $(0,0)$, two marked $(1,0)$, two marked $(0,1)$, and one marked $(1,1)$. A player selects a chip at random and is given the sum of the two coordinates in dollars as a prize. What is the expected prize money a play can get?

Let X_1 and X_2 denote the two coordinates. Their joint PMF is

$$f(x,y) = \frac{3 - x_1 - x_2}{8}, \quad x_1 = 0,1 \;\; and \;\; x_2 = 0,1$$

Thus,

$$\begin{aligned} E(X_1 + X_2) &= \sum_{x_2=0}^{1} \sum_{x_1=0}^{1} (x_1 + x_2) \frac{3 - x_1 - x_2}{8} \\ &= (0)(\frac{3}{8}) + (1)(\frac{2}{8}) + (1)(\frac{2}{8}) + (2)(\frac{1}{8}) = \frac{3}{4} \end{aligned}$$

Correlation coefficient

Let $u(X, Y) = (X - \mu_X)(Y - \mu_Y)$, then

$$E[u(X, Y)] = E[(X - \mu_X)(Y - \mu_Y)] = Cov(X, Y) = \sigma_{XY}$$

is called the covariance of X and Y.

$$\rho = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

is called the correlation coefficient of X and Y.

A commonly used formula to calculate covariance:

$$\begin{aligned} Cov(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E(XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y) \\ &= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X \mu_Y \\ &= E(XY) - \mu_X \mu_Y \end{aligned}$$

Correlation coefficient

Example: Let X and Y have the joint PMF

$$f(x,y) = \frac{x+2y}{18}, \quad x = 1,2 \;\; and \;\; y = 1,2$$

What is the correlation coefficient of X and Y?

The marginal PMF are respectively

$$f_X(x) = \sum_{y=1}^{2} \frac{x+2y}{18} = \frac{x+3}{9}$$

$$f_Y(y) = \sum_{x=1}^{2} \frac{x+2y}{18} = \frac{3+4y}{18}$$

The mean and variance of X are

$$\mu_X = \sum_{x=1}^{2} x\frac{x+3}{9} = (1)\frac{4}{9} + (2)\frac{5}{9} = \frac{14}{9}$$

Correlation coefficient

$$\sigma_X^2 = E(X^2) - \mu_X^2 = \sum_{x=1}^{2} x^2 \frac{x+3}{9} - \left(\frac{14}{9}\right)^2 = \frac{20}{81}$$

Similarly, we get the mean and variance of Y

$$\mu_Y = \frac{29}{18} \quad \sigma_Y^2 = \frac{77}{324}$$

The covariance of X and Y

$$\begin{aligned} Cov(X,Y) &= \sum_{x=1}^{2}\sum_{y=1}^{2} xy\frac{x+2y}{18} - \frac{14}{9}\frac{29}{18} \\ &= (1)(1)\frac{3}{18} + (2)(1)\frac{4}{18} + (1)(2)\frac{5}{18} + (2)(2)\frac{6}{18} - \frac{14}{9}\frac{29}{18} \\ &= -\frac{1}{162} \end{aligned}$$

$$\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = -0.025$$

Correlation coefficient

Proposition: If X and Y are independent, $Cov(X, Y) = 0$.

$$\begin{aligned} E(XY) &= \sum_{S_X} \sum_{S_Y} xyf(x, y) \\ &= \sum_{S_X} \sum_{S_Y} xyf_X(x)f_Y(y) \\ &= \sum_{S_X} xf_X(x) \sum_{S_Y} yf_Y(y) \\ &= \mu_X \mu_Y \end{aligned}$$

Thus, we have

$$Cov(X, Y) = E(XY) - \mu_X \mu_Y = 0$$

Correlation coefficient

If $Cov(X, Y) = 0$, are X and Y necessarily independent?

Example: Let X and Y have the joint PMF

$$f(x, y) = \frac{1}{3}, \quad (x, y) = (0, 1), (1, 0), (2, 1).$$

It is easy to get the marginal PMF of X and Y:

$$f_X(x) = \frac{1}{3},\ x = 0, 1, 2; \quad f_Y(y) = \begin{cases} \frac{1}{3}, & y = 0 \\ \frac{2}{3}, & y = 1 \end{cases}$$

Thus, $\mu_X = 1$ amd $\mu_Y = 2/3$. Then

$$\begin{aligned} Cov(X, Y) &= E(XY) - \mu_X \mu_Y \\ &= (0)(1)\frac{1}{3} + (1)(0)\frac{1}{3} + (2)(1)\frac{1}{3} - (1)\frac{2}{3} \\ &= 0 \end{aligned}$$

It is obvious that $f(x, y) \neq f_X(x) f_Y(y)$. Thus, X and Y are dependent.

Conditional distributions

Let X and Y have a joint discrete distribution with PMF $f(x, y)$ on space S. Say the marginal PMF are $f_X(x)$ and $f_Y(y)$ respectively. Let event $A = \{X = x\}$ and event $B = \{Y = y\}$. Thus $A \cap B = \{X = x, Y = y\}$. Because $P(A \cap B) = P(X = x, Y = y) = f(x, y)$ and $P(B) = P(Y = y) = f_Y(y)$, the conditional probability of A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{f(x, y)}{f_Y(y)}$$

Definition: The conditional probability mass function of X, given that $Y = y$, is defined by

$$g(x|y) = \frac{f(x, y)}{f_Y(y)}$$

provided that $f_Y(y) > 0$

Conditional distributions

Example: Let X and Y have the joint PMF

$$f(x,y) = \frac{x+y}{21}, \quad x = 1,2,3, \quad y = 1,2.$$

Find the conditional distribution $g(x|y)$.

We first calculate marginal PMF of y:

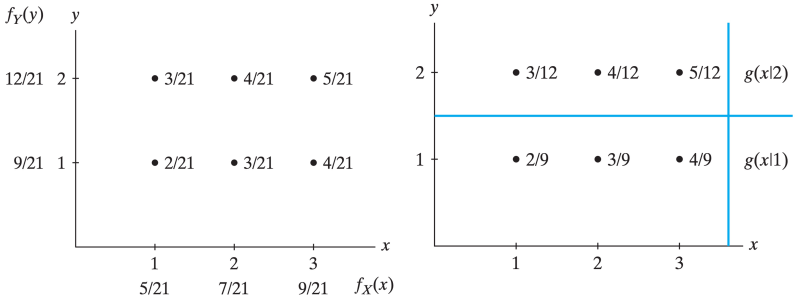
$$f_Y(y) = \sum_{x=1}^{3} \frac{x+y}{21} = \frac{y+2}{7}, \quad y = 1,2$$

Thus, the conditional PMF of X given Y is

$$g(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{(x+y)/21}{(y+2)/7} = \frac{x+y}{3y+6}$$

Conditional distribution

Similar to conditional probability, we can visualize the joint, marginal, and conditional PMF.



(Graphic illustration of joint, marginal and conditional PMF.)

Conditional expectation

Because conditional PMF is a PMF, we thus can define conditional expectation the same way we define mathematical expectation:

$$E\left[u(Y)|X=x\right]=\sum_{y}u(y)g(y|x)$$

Conditional mean and conditional variance are defined by

$$\mu_{Y|X}=E(Y|X)=\sum_{y}yg(y|x)$$

$$\sigma^2_{Y|X}=E\left[(Y-\mu_{Y|X})^2|X\right]=\sum_{y}(y-\mu_{Y|X})^2g(y|x)$$

Conditional expectation

Example: Let X and Y have a multinomial PMF with parameters n, p_X, and p_Y. That is,

$$f(x,y) = \frac{n!}{x!y!(n-x-y)!} p_X^x p_Y^y (1 - p_x - p_Y)^{n-x-y}$$

What is the conditional mean of X given Y?

We know that the marginal distribution of Y is binomial,. i.e.,

$$f_Y(y) = \frac{n!}{y!(n-y)!} p_Y^y (1 - p_Y)^{n-y}$$

Thus, the conditional PMF of X given Y is

$$g(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{(n-y)!}{x!(n-y-x)!} (\frac{p_X}{1-p_Y})^x (1 - \frac{p_X}{1-p_Y})^{n-y-x}$$

This is a binomial distribution with parameters $n-y$ and $\frac{p_X}{1-p_Y}$. Thus, the conditional mean is $(n-y)\frac{p_X}{1-p_Y}$.

Multivariate distribution of continuous random variables

The idea of joint distributions of discrete random variables can be extended to that of continuous random variables. The **joint probability density function** of two continuous random variables is an integrable function $f(x, y)$ such that

- $f(x, y) \geqslant 0$, where $f(x, y) = 0$ when (x, y) is not in the space of X and Y;
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dxdy = 1$;
- $P(X, Y) \in A = \int \int_{A} f(x, y) dxdy$

Multivariate distribution of continuous random variables

The **marginal probability density** function of X and Y are given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy, \quad x \in S_X;$$
$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx, \quad y \in S_Y;$$

X and Y are **independent** if and only if $f(x, y) = f_X(x)f_Y(y)$

Multivariate distribution of continuous random variables

The correlation coefficient of two continuous random variables X and Y is defined in the same way as the discrete random variables as

$$\rho = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

The **Conditional probability density function** of X, given that $Y = y$, is

$$f(x|y) = \frac{f(x, y)}{f_Y(y)},$$

provided that $f_Y(y) > 0$.

Multivariate distribution of continuous random variables

Example: Let X and Y have the joint PDF

$$f(x, y) = 1, \quad x \leqslant y \leqslant x + 1, \quad 0 \leqslant x \leqslant 1.$$

Find the marginal PDF and the correlation coefficient of X and Y.

The marginal PDFs of X and Y are

$$f_X(x) = \int_x^{x+1} 1 dy = 1, \quad 0 \leqslant x \leqslant 1$$

$$f_Y(y) = \begin{cases} \displaystyle\int_0^y 1 dx = y, \quad 0 \leqslant y \leqslant 1, \\ \displaystyle\int_{y-1}^1 1 dx = 2 - y, \quad 1 \leqslant y \leqslant 2. \end{cases}$$

Multivariate distribution of continuous random variables

The mean and variance of X and Y are

$$\mu_X = \int_0^1 x \cdot 1 dx = \frac{1}{2}$$

$$\mu_Y = \int_0^1 y \cdot y dy + \int_1^2 y \cdot (2-y) dy = \frac{1}{3} + \frac{2}{3} = 1$$

$$E(X^2) = \int_0^1 x^2 \cdot 1 dx = \frac{1}{3}$$

$$E(Y^2) = \int_0^1 y^2 \cdot y dy + \int_1^2 y^2 \cdot (2-y) dy = \frac{7}{6}$$

$$E(XY) = \int_0^1 \int_x^{x+1} xy \cdot 1 dy dx = \int_0^1 \frac{1}{2} x(2x+1) dx = \frac{7}{12}$$

Multivariate distribution of continuous random variables

$$\sigma_X^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$$

$$\sigma_Y^2 = \frac{7}{6} - 1^2 = \frac{1}{6}$$

$$\sigma_{XY} = \frac{7}{12} - \left(\frac{1}{2}\right)(1) = \frac{1}{12}$$

Therefore, the correlation coefficient is

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{1/12}{\sqrt{(1/12)(1/6)}} = \frac{\sqrt{2}}{2}$$

Multivariate normal distribution

A very commonly used multivariate distribution is the multivariate normal distribution. Random variables X and Y have a bivariate normal distribution if its joint PDF is

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[-\frac{q(x,y)}{2}\right],$$

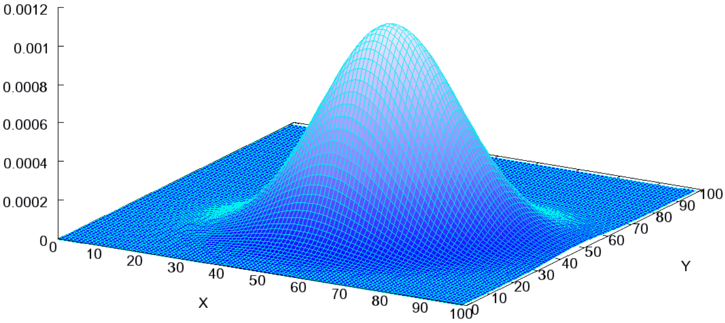
where

$$q(x,y) = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]$$

Here, μ_X and μ_Y are the mean of X and Y, σ_X and σ_Y are the standard deviation of X and Y, and ρ is the correlation coefficient.

Multivariate normal distribution

A bivariate normal distribution has a typical PDF figure as follows.



Multivariate normal distribution

If random variables X and Y have a bivariate normal distribution, then the marginal distribution of X and Y are both normal.

$$
\begin{aligned}
q(x,y) &= \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_x}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right] \\
&= \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_X}{\sigma_X} - \rho\frac{y-\mu_Y}{\sigma_Y}\right)^2 + (1-\rho^2)\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right] \\
&= \frac{1}{\sigma_X^2(1-\rho^2)}\left(x-\mu_X-\rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2
\end{aligned}
$$

Multivariate normal distribution

Thus, the marginal distribution of Y is

$$
\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f(x,y)dx = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[-\frac{q(x,y)}{2}\right]dx \\
&= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}\right] \\
&\quad \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\sigma_X^2(1-\rho^2)}\left(x-\mu_X-\rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)\right)^2\right]dx \\
&= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}\right](\sigma_X\sqrt{2\pi}\sqrt{1-\rho^2}) \\
&= \frac{1}{\sigma_Y\sqrt{2\pi}} \exp\left[-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}\right]
\end{aligned}
$$

Thus, the marginal distribution of Y is $N(\mu_Y, \sigma_Y^2)$. Using the procedure, it is obvious that $X \sim N(\mu_X, \sigma_X^2)$.

Multivariate normal distribution

If If random variables X and Y have a bivariate normal distribution, then the conditional distribution of X given Y is normal.

The joint PDF is

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[-\frac{q(x,y)}{2}\right],$$

where

$$q(x,y) = \frac{1}{\sigma_X^2(1-\rho^2)}\left(x-\mu_X-\rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2$$

The marginal PDF of Y is

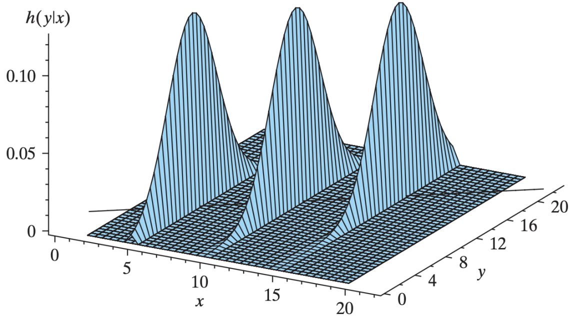
$$f_Y(y) = \frac{1}{\sigma_Y\sqrt{2\pi}} \exp\left[-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}\right]$$

Multivariate normal distribution

The conditional distribution of X given Y is thus

$$g(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{1}{\sigma_X\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left[-\frac{[x-\mu_X-\rho(\sigma_X/\sigma_Y)(y-\mu_Y)]^2}{2\sigma_X^2(1-\rho^2)}\right]$$

Thus, $g(x|y)$ is $N\left(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y), (1-\rho^2)\sigma_X^2\right)$.



(Illustration of conditional distribution of a bivariate normal distribution)

Multivariate normal distribution

We can extend the case of bivariate normal distribution to more than two variables. For more k variables, we write the PDF of a multivariate normal distribution in matrix notation:

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^k |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right)$$

where $\boldsymbol{x}$ and $\boldsymbol{\mu}$ are column vectors of the variables and its means, $\boldsymbol{\Sigma}$ is the $k \times k$ variance covariance matrix, i.e.,

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}; \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix}; \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2_{X_1} & \sigma_{X_1 X_2} & \cdots & \sigma_{X_1 X_k} \\ \sigma_{X_1 X_1} & \sigma^2_{X_2} & \cdots & \sigma_{X_2 X_k} \\ \vdots & \vdots & & \vdots \\ \sigma_{X_k X_1} & \sigma^2_{X_k X_2} & \cdots & \sigma^2_{X_k} \end{bmatrix}$$