Advanced Statistics: Theory and Methods

Dr. Nandini Kannan

email : nandini.kannan@plaksha.edu.in

Chapter 2

Bivariate and Multivariate Distributions

In many experiments or studies, we often record the simultaneous outcomes of several random variables. For example, to determine the success of students in college, information on national test scores and board exam scores are analyzed. In assessing air quality standards, measurements on sulphur oxides, nitrogen oxides, and particulate matter are obtained.

We are interested in the simultaneous behaviour of these random variables.

2.1 Bivariate Random Variables

Let X and Y be two discrete random variables. The **joint probability mass function** of X and Y is denoted by $f(x, y)$ and is given by

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

Definition: The function $f(x, y)$ is a joint pmf of the discrete random variables X and Y if

1. $f(x, y) \geq 0, \forall (x, y)$
2. $\sum_x \sum_y f(x, y) = 1.$

For any region in the $x - y$ plane,

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

Example: An experiment consists of tossing a tetrahedron (4 sided die) twice. Let X denote the number on the first toss, and let Y denote the maximum of the two tosses. Find the joint pmf of X and Y.

When X and Y are continuous random variables, we may define the joint probability density function $f(x, y)$.

Definition: The function $f(x, y)$ is a joint pdf of the continuous random variables X and Y if

1. $f(x, y) \geq 0$ for all (x, y).
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1.$

For any region in the $x - y$ plane,

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

Example: Assume that for a certain type of washer, both the thickness and the hole diameter vary from item to item. Let X denote the thickness, and let Y denote the diameter. Assume that the joint pdf of X and Y is given by

$$f(x, y) = \begin{cases} \frac{1}{6}(x + y) & 1 \leq x \leq 2, 4 \leq y \leq 5 \\ o & o.w. \end{cases}$$

Find the probability that a randomly chosen washer has a thickness between 1 and 1.5 mm and a diameter between 4.5 and 5 mm.

Definition: The **marginal distributions** of X and Y alone are

1. $g(x) = \sum_y f(x,y)$ and $h(y) = \sum_x f(x,y)$ for the discrete case, and

2. $g(x) = \int_{-\infty}^{\infty} f(x,y)dy$ and $h(y) = \int_{-\infty}^{\infty} f(x,y)dx$.

Definition: The **conditional distribution** of the random variable Y given that $X = x$ is

$$f(y|x) = \frac{f(x,y)}{g(x)}, \qquad g(x) > 0.$$

The **conditional distribution** of the random variable X given that $Y = y$ is

$$f(x|y) = \frac{f(x,y)}{h(y)}, \qquad h(y) > 0.$$

Example: X and Y are continuous random variables with joint pdf given by

$$f(x) = \begin{cases} k(x+y^2) & 0 \le x \le 1, 0 \le y \le 1 \\ o & o.w. \end{cases}$$

a. Find k.

b. Find the marginal densities of X and Y.

c. Find the conditional density of X given $Y = y$.

Definition: Let X and Y be two random variables with joint probability distribution $f(x,y)$ and marginals $g(x)$ and $h(y)$, respectively. X and Y are said to be **statistically independent** if and only if

$$f(x,y) = g(x)h(y)$$

for all (x,y) within their range.

2.2 Expectation

The concept of expectation extends to random vectors in a straightforward manner. Let (X,Y) be a random vector, and $U = g(X,Y)$ be some real valued function. Then U is a random variable and we can compute its expectation by finding its distribution. We can also find the expectation by using the joint distribution of (X,Y).

Definition: Let (X,Y) be a bivariate random vector. The expected value of the random variable $U = g(X,Y)$ is given by

$$E[g(X,Y)] = \begin{cases} \sum_x \sum_y g(x,y)p(x,y), & (X,Y) \text{ discrete;} \\ \int \int g(x,y)f(x,y)dxdy, & (X,Y) \text{ continuous} \end{cases} \tag{2.1}$$

provided the sum or integral exist. If $E|g(X,Y)| = \infty$, we say that $E(g(X,Y))$ **does not exist**.

If two random variables are dependent, can we measure the strength of their relationship?

In order to quantify the relationship, we introduce the ideas of covariance and correlation.

<u>**Definition:**</u> Let X and Y be two random variables with finite variances. The **covariance** of X and Y is defined as

$$Cov(X,Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X\mu_Y. \tag{2.2}$$

$Cov(X,Y)$ will be positive when $X - \mu_X$ and $Y - \mu_Y$ tend to have the same sign with high probability. $Cov(X,Y)$ will be negative when $X - \mu_X$ and $Y - \mu_Y$ tend to have opposite signs with high probability. Therefore, the sign of the covariance provides information about the relationship between X and Y. However, the magnitude of the covariance does not in itself provide information on the strength of the relationship since it depends on the variability.

<u>**Remark:**</u> Equation (4.21) defines an inner product on the linear space spanned by X and Y. We have

$$< X, X >= ||X||^2 \qquad Cov(X,X) = Var(X).$$

<u>**Definition:**</u> Let X and Y be two random variables with finite variances. The **correlation** of X and Y is the number defined by

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X\sigma_Y}. \tag{2.3}$$

The value $\rho(X,Y)$ is called the **correlation coefficient**.

Theorem 2.2.1. *If X and Y are independent, then $Cov(X,Y) = 0$ and $\rho(X,Y) = 0$.*

Proof: If X and Y are independent,

$$\begin{aligned} Cov(X,Y) &= E(XY) - \mu_X\mu_Y \\ &= E(X)E(Y) - \mu_X\mu_Y \\ &= 0. \end{aligned}$$

If $Cov(X,Y) = 0$, then clearly $\rho(X,Y) = 0$.

■

The converse of the result is not true. If $\rho(X,Y) = 0$, X and Y may still exhibit some form of dependence. Covariance and correlation measure a particular kind of linear relationship between X and Y.

Theorem 2.2.2. *For any random variables X and Y,*

(a) $-1 \leq \rho(X,Y) \leq 1$.

(b) $|\rho(X,Y)| = 1$ *if and only if there exist numbers* $a \neq 0$ *and* b *such that* $P(Y = aX + b) = 1$. *If* $\rho(X,Y) = 1$, *then* $a > 0$, *and if* $\rho(X,Y) = -1$, *then* $a < 0$.

2.3 Multivariate Distributions

In the previous sections, we have considered bivariate distributions. We will now extend the definitions given earlier to the case of three or more random variables.

Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be an $n-$dimensional random vector. Then $\boldsymbol{X}$ is a function defined from the sample space $\mathcal{S}$ into $\mathcal{R}^n$.

If the sample space of $\boldsymbol{X}$ is countable, then we can define the joint pmf by

$$p(\boldsymbol{x}) = P(X_1 = x_1, \ldots, X_n = x_n)$$

for each $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathcal{R}^n$. For any $\boldsymbol{A} \in \mathcal{B}^n$,

$$P(\boldsymbol{X} \in A) = \sum_{\boldsymbol{x} \in A} p(\boldsymbol{x}.$$

If $\boldsymbol{X}$ is a continuous random vector, then the joint pdf is a function $f(x_1, \ldots, x_n)$ that satisfies

$$P(\boldsymbol{X} \in A) = \int \ldots \int_A f(x_1, \ldots, x_n) dx_1 \ldots dx_n.$$

The joint pmf and pdf satisfy the standard properties: they are non-negative functions and either sum to one or integrate to one.

The joint cumulative distribution function is defined as

$$F(\boldsymbol{x}) = P(X_1 \leq x_1, \ldots, X_n \leq x_n).$$

If $\boldsymbol{X}$ is continuous, then the joint pdf is obtained as

$$f(\boldsymbol{x}) = \frac{\partial^n}{\partial x_1 \dots \partial x_n} F(\boldsymbol{x}).$$

Let $g(\boldsymbol{x})$ be a real valued function defined on the sample space of $\boldsymbol{X}$. Then $g(\boldsymbol{X})$ is a random variable and

$$E[g(\boldsymbol{X})] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x}; \qquad E[g(\boldsymbol{X})] = \sum_{\boldsymbol{x} \in \mathcal{R}^n} p(\boldsymbol{x}),$$

for the continuous and discrete cases, respectively. The properties of expectation carry over to the case of multiple random variables.

We can define the marginal pmf or pdf of any subset of the random vector. We have

$$f(x_1, \dots x_k) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_{k+1} \dots dx_n$$

as the marginal distribution of $(X_1, \dots, X_k)$. The discrete case can be similarly defined.

The conditional pdf of a subset of random variables given the values of the remaining variables is defined as

$$f(x_{k+1}, \dots, x_n | x_1, \dots x_k) = \frac{f(x_1, \dots, x_n)}{f(x_1, \dots x_k)},$$

provided the denominator is not zero.

Definition: Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be random vectors with joint pdf $f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$. Let $f_{\mathbf{X}_i}(\boldsymbol{x}_i)$ be the marginal pdf of $\boldsymbol{X}_i$. Then $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are said to be **mutually independent random vectors** if, for every $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$,

$$f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \prod_{i=1}^{n} f_{\mathbf{X}_i}(\boldsymbol{x}_i).$$

If the X_i's are all one-dimensional, then $X_1, \ldots, X_n$ are called **mutually independent random variables**.

The definition for discrete random variables is similar.

2.3.1 The Multinomial Distribution

Consider the following experiment:

1. The experiment consists of n identical trials.
2. Each trial can result in one of k possible outcomes.
3. The probability of the $i-$th outcome is p_i and remains constant from trial to trial.
4. The trials are independent.

We have

$$\sum_{i=1}^{k} p_i = 1.$$

Let X_i be the random variable that records the total number of times outcome i is observed in the n trials. We have

$$\sum_{i=1}^{k} = n.$$

Then

$$P(X_1 = x_1, \ldots, X_k = x_k) = \begin{cases} \begin{pmatrix} n \\ x_1,\ x_2, \ldots, x_k \end{pmatrix} p_1^{x_1} \ldots p_k^{x_k}, & n = \sum x_i; \\ 0, & \text{ow.} \end{cases} \tag{2.4}$$

where

$$\begin{pmatrix} n \\ x_1,\ x_2, \ldots, x_k \end{pmatrix} = \frac{n!}{x_1! \ldots x_k!}$$

is called a multinomial coefficient.

$(X_1, \ldots, X_k)$ is said to have a **multinomial distribution** with n trials and probabilities $(p_1, \ldots, p_k)$.

Theorem 2.3.1. *The marginal pmf of X_i is $Bin(n, p_i)$, i=1, …, k.*

2.3.2 The Multivariate Normal

Let $Y_1, \ldots, Y_p$ be p random variables. Let $\boldsymbol{Y}$ be a $p \times 1$ column vector, with the $i-$ th element equal to Y_i. Then $\boldsymbol{Y}$ is a **random vector**.

Let $\boldsymbol{\mu}$ be the $p \times 1$ vector

$$\boldsymbol{\mu} = \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_p) \end{bmatrix}.$$

This is called the **mean vector**. The dispersion matrix or variance-covariance matrix is given by

$$\boldsymbol{\Sigma} = E[(\boldsymbol{Y} - \boldsymbol{\mu})(\boldsymbol{Y} - \boldsymbol{\mu})^T] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \ldots & \sigma_{pp} \end{bmatrix},$$

where

$$\sigma_{ii} = \sigma_i^2 = Var(Y_i)$$

and

$$\sigma_{ij} = Cov(Y_i, Y_j).$$

The dispersion matrix is a symmetric $p \times p$ matrix, with $\boldsymbol{\Sigma}$ being non-negative definite (nnd). We have

$$\boldsymbol{x}^T \boldsymbol{\Sigma} \boldsymbol{x} = \sum_i \sum_j x_i x_j \sigma_{ij} = Var(x_1 Y_1 + \ldots + x_p Y_p)$$

which is ≥ 0 and $= 0$ if $x_1 Y_1 + \ldots + x_p Y_p = 0$..

The correlation matrix is given by

$$\boldsymbol{P} == \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{21} & 1 & \dots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \dots & 1 \end{bmatrix},$$

where

$$\rho_{ij} = Corr(Y_i, Y_j).$$

<u>Definition:</u> A random vector $\boldsymbol{Y}$ is said to have a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and dispersion matrix $\boldsymbol{\Sigma}$ (nnd) if it has a pdf given by

$$f(\boldsymbol{y}) = \left(\frac{1}{2\pi}\right)^{p/2} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} exp\left\{-\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu})\right\}.$$

We write $\boldsymbol{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

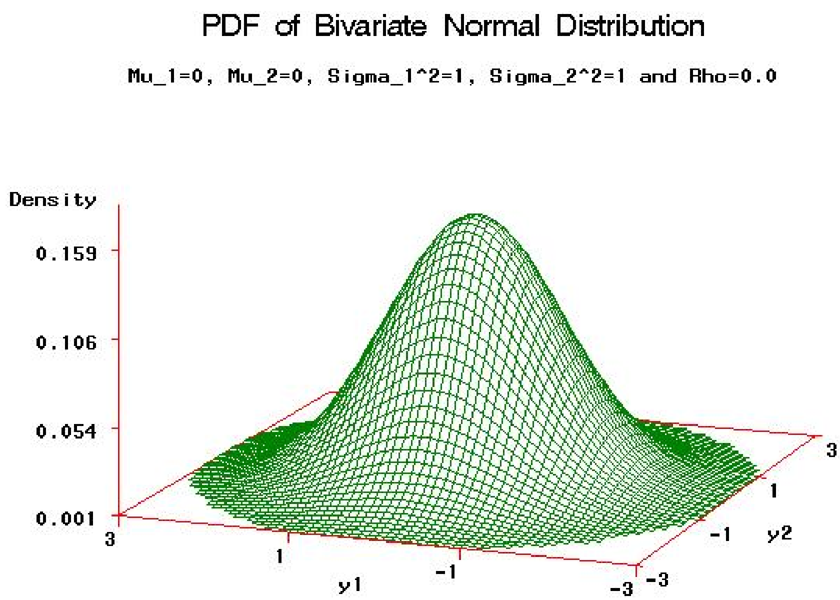

Figure 2.1: Bivariate Normal PDF: $\rho = 0$

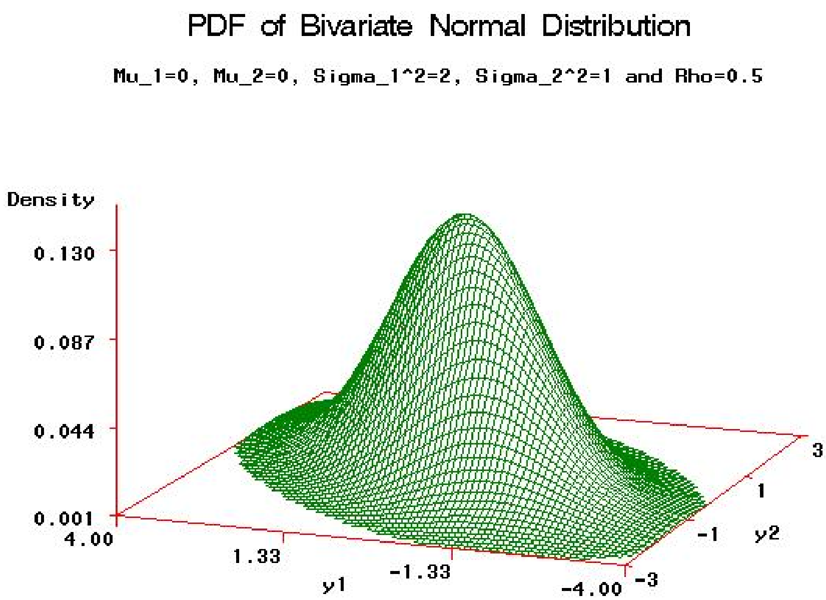

Figure 2.2: Bivariate Normal PDF: $\rho = 0.5$

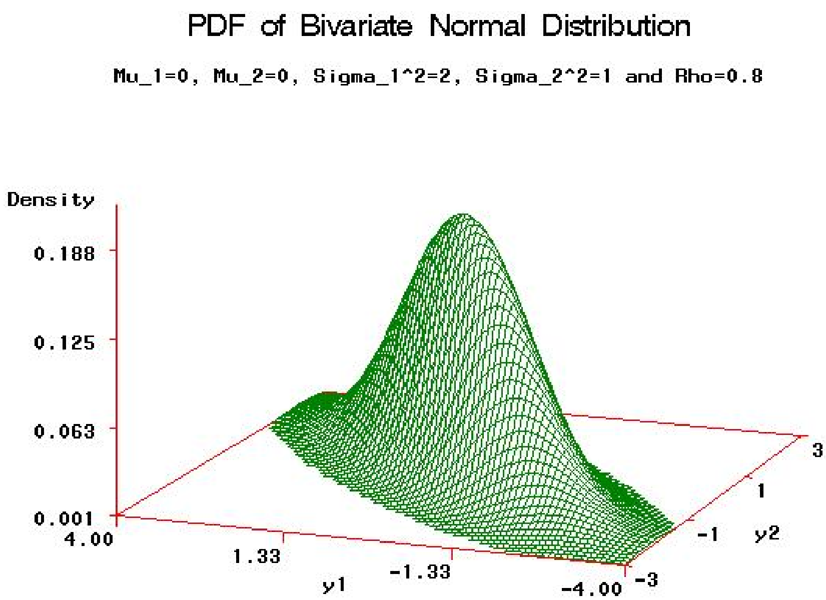

Figure 2.3: Bivariate Normal PDF: $\rho = 0.8$