ST 371 (VIII): Theory of Joint Distributions

So far we have focused on probability distributions for single random variables. However, we are often interested in probability statements concerning two or more random variables. The following examples are illustrative:

- In ecological studies, counts, modeled as random variables, of several species are often made. One species is often the prey of another; clearly, the number of predators will be related to the number of prey.
- The joint probability distribution of the x, y and z components of wind velocity can be experimentally measured in studies of atmospheric turbulence.
- The joint distribution of the values of various physiological variables in a population of patients is often of interest in medical studies.
- A model for the joint distribution of age and length in a population of fish can be used to estimate the age distribution from the length distribution. The age distribution is relevant to the setting of reasonable harvesting policies.

1 Joint Distribution

The joint behavior of two random variables X and Y is determined by the joint cumulative distribution function (cdf):

$$F_{XY}(x, y) = P(X \leq x, Y \leq y), \tag{1.1}$$

where X and Y are continuous or discrete. For example, the probability that (X, Y) belongs to a given rectangle is

$$\begin{aligned} & P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) \\ = \ & F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1). \end{aligned}$$

In general, if $X_1, \cdots, X_n$ are jointly distributed random variables, the joint cdf is

$$F_{X_1,\cdots,X_n}(x_1, \cdots, x_n) = P(X_1 \le x_1, X_2 \le x_2, \cdots, X_n \le x_n).$$

Two- and higher-dimensional versions of probability distribution functions and probability mass functions exist. We start with a detailed description of joint probability mass functions.

1.1 Jointly Discrete Random Variables

Joint probability mass functions: Let X and Y be discrete random variables defined on the sample space that take on values $\{x_1, x_2, \cdots\}$ and $\{y_1, y_2, \cdots\}$, respectively. The joint probability mass function of (X, Y) is

$$p(x_i, y_j) = P(X = x_i, Y = y_j). \tag{1.2}$$

Example 1 A fair coin is tossed three times independently: let X denote the number of heads on the first toss and Y denote the total number of heads. Find the joint probability mass function of X and Y.

Solution: the joint distribution of (X, Y) can be summarized in the following table:

x/y	0	1	2	3
0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0
1	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$

Marginal probability mass functions: Suppose that we wish to find the pmf of Y from the joint pmf of X and Y in the previous example:

$$\begin{aligned}
p_Y(0) &= P(Y = 0) \\
&= P(Y = 0, X = 0) + P(Y = 0, X = 1) \\
&= 1/8 + 0 = 1/8 \\
p_Y(1) &= P(Y = 1) \\
&= P(Y = 1, X = 0) + P(Y = 1, X = 1) \\
&= 2/8 + 1/8 = 3/8
\end{aligned}$$

In general, to find the frequency function of Y, we simply sum down the appropriate column of the table giving the joint pmf of X and Y. For this reason, p_Y is called the *marginal probability mass function* of Y. Similarly, summing across the rows gives

$$p_X(x) = \sum_i (x, y_i),$$

which is the marginal pmf of X. For the above example, we have

$$p_X(0) = p_X(1) = 0.5.$$

1.2 Jointly Continuous Random Variables

Joint PDF and Joint CDF: Suppose that X and Y are continuous random variables. The joint probability density function (pdf) of X and Y is the function $f(x, y)$ such that for every set C of pairs of real numbers

$$P((X, Y) \in C) = \int\int_{(x,y)\in C} f(x, y)dxdy. \tag{1.3}$$

Another interpretation of the joint pdf is obtained as follows:

$$\begin{aligned} P\{a < X < a+da,\ b < Y < b+db\} &= \int_b^{b+db} \int_a^{a+da} f(x,y)dxdy \\ &\approx f(a,b)dadb, \end{aligned}$$

when da and db are small and $f(x,y)$ is continuous at a, b. Hence $f(a,b)$ is a measure of how likely it is that the random vector (X,Y) will be near (a,b). This is similar to the interpretation of the pdf $f(x)$ for a single random variable X being a measure of how likely it is to be near x.

The joint CDF of (X,Y) can be obtained as follows:

$$\begin{aligned} F(a,b) &= P\{X \in (-\infty, a], Y \in (-\infty, b]\} \\ &= \int\int_{X \in (-\infty,a], Y \in (-\infty,b]} f(x,y)dxdy \\ &= \int_{-\infty}^{b} \int_{-\infty}^{a} f(x,y)dxdy \end{aligned}$$

It follows upon differentiation that

$$f(a,b) = \frac{\partial^2}{\partial a \partial b} F(a,b),$$

wherever the partial derivatives are defined.

<u>Marginal PDF</u>: The cdf and pdf of X can be obtained from the pdf of (X,Y):

$$\begin{aligned} P(X \le x) &= P\{X \le x, Y \in (-\infty, \infty)\} \\ &= \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(x,y)dydx. \end{aligned}$$

Then we have

$$f_X(x) = \frac{d}{dx} P(X \le x) = \frac{d}{dx} \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(x,y)dydx = \int_{-\infty}^{\infty} f(x,y)dy.$$

Similarly, the pdf of Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx.$$

Example 2 Consider the pdf for X and Y

$$f(x,y) = \frac{12}{7}(x^2 + xy),\ 0 \le x \le 1,\ 0 \le y \le 1.$$

Find (a) $P(X > Y)$ (b) the marginal density $f_X(x)$.

Solution: (a). Note that

$$\begin{aligned} P(X > Y) &= \int\int_{X>Y} f(x,y)dxdy \\ &= \int_0^1 \int_y^1 \frac{12}{7}(x^2 + xy)dxdy \\ &= \int_0^1 \frac{12}{7}(\frac{x^3}{3} + \frac{x^2 y}{2})|_y^1 dy \\ &= \int_0^1 (\frac{12}{21} + \frac{12}{14}y - \frac{12}{21}y^3 - \frac{12}{14}y^3)dy \\ &= \frac{9}{14} \end{aligned}$$

(b) For $0 \le x \le 1$,

$$f_X(x) = \int_0^1 \frac{12}{7} f(x^2 + xy)dy = \frac{12}{7}x^2 + \frac{6}{7}x.$$

For $x < 0$ or $x > 1$, we have

$$f_X(x) = 0.$$

Example 3 Suppose the set of possible values for (X, Y) is the rectangle $D = \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$. Let the joint pdf of (X, Y) be

$$f(x, y) = \frac{6}{5}(x + y^2), \text{ for } (x, y) \in D.$$

(a) Verify that $f(x, y)$ is a valid pdf.

(b) Find $P(0 \leq X \leq 1/4, 0 \leq Y \leq 1/4)$.

(c) Find the marginal pdf of X and Y.

(d) Find $P(\frac{1}{4} \leq Y \leq \frac{3}{4})$.

2 Independent Random Variables

The random variables X and Y are said to be *independent* if for any two sets of real numbers A and B,

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B). \tag{2.4}$$

Loosely speaking, X and Y are *independent* if knowing the value of one of the random variables does not change the distribution of the other random variable. Random variables that are not independent are said to be dependent.

For discrete random variables, the condition of independence is equivalent to

$$P(X = x, Y = y) = P(X = x)P(Y = y) \text{ for all } x, y.$$

For continuous random variables, the condition of independence is equivalent to

$$f(x, y) = f_X(x)f_Y(y) \text{ for all } x, y.$$

Example 4 Consider (X, Y) with joint pmf

$$\begin{aligned} &p(10, 1) = p(20, 1) = p(20, 2) = \tfrac{1}{10} \\ &p(10, 2) = p(10, 3) = \tfrac{1}{5}, \text{ and } p(20, 3) = \tfrac{3}{10}. \end{aligned}$$

Are X and Y independent?

Example 5 Suppose that a man and a woman decide to meet at a certain location. If each person independently arrives at a time uniformly distributed between $[0, T]$.

- Find the joint pdf of the arriving times X and Y.
- Find the probability that the first to arrive has to wait longer than a period of τ ($\tau < T$).

2.1 More than two random variables

If $X_1, \cdots, X_n$ are all discrete random variables, the joint pmf of the variables is the function

$$p(x_1, \cdots, x_n) = P(X_1 = x_1, \cdots, X_n = x_n). \tag{2.5}$$

If the variables are continuous, the joint pdf of the variables is the function $f(x_1, \cdots, x_n)$ such that

$$P(a_1 \le X_1 \le b_1, \cdots, a_n \le X_n \le b_n) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1, \cdots, x_n) dx_1 \cdots x_n.$$

One of the most important joint distributions is the multinomial distribution which arises when a sequence of n independent and identical experiments is performed, where each experiment can result in any one of r possible outcomes, with respective probabilities $p_1, \cdots, p_r$. If we let denote the number of the n experiments that result in outcome i, then

$$P(X_1 = n_1, \ldots, X_r = n_r) = \frac{n!}{n_1! \cdots n_r!} p_1^{n_1} \cdots p_r^{n_r}, \tag{2.6}$$

whenever $\sum_{i=1}^{r} n_i = n$.

Example 6 If the allele of each ten independently obtained pea sections is determined with

$$p_1 = P(AA), p_2 = P(Aa), p_3 = P(aa).$$

Let $X_1 =$ number of AA's, $X_2 =$ number of Aa's and $X_3 =$ number of aa's. What is the probability that we have 2 AA's, 5 Aa's and 3 aa's.

Example 7 When a certain method is used to collect a fixed volume of rock samples in a region, there are four rock types. Let X_1, X_2 and X_3 denote the proportion by volume of rock types 1, 2, and 3 in a randomly selected sample (the proportion of rock type 4 is redundant because $X_4 = 1 - X_1 - X_2 - X_3$). Suppose the joint pdf of (X_1, X_2, X_3) is

$$f(x_1, x_2, x_3) = kx_1x_2(1 - x_3)$$

for $(x_1, x_2, x_3) \in [0, 1]^3$ and $x_1 + x_2 + x_3 \leq 1$ and $f(x_1, x_2, x_3) = 0$ otherwise.

- What is k?
- What is the probability that types 1 and 2 together account for at most 50%?

The random variables $X_1, \cdots, X_n$ are independent if for every subset $X_{i_1}, \cdots, X_{i_k}$ of the variables, the joint pmf (pdf) is equal to the product of the marginal pmf's (pdf's).

Example 8 If $X_1, \cdots, X_n$ represent the lifetimes of n independent components, and each lifetime is exponentially distributed with parameter λ. The system fails if any component fails. Find the expected lifetime of system.

2.2 Sum of Independent Random Variables

It is often important to be able to calculate the distribution of $X+Y$ from the distribution of X and Y when X and Y are independent.

Sum of Discrete Independent RV's: Suppose that X and Y are discrete independent rv's. Let $Z = X+Y$. The probability mass function of Z is

$$P(Z=z) = \sum_x P(X=x, Y=z-x) \tag{2.7}$$

$$= \sum_x P(X=x)P(Y=z-x), \tag{2.8}$$

where the second equality uses the independence of X and Y.

Sums of Continuous Independent RV's: Suppose that X and Y are continuous independent rv's. Let $Z = X+Y$. The CDF of Z is

$$\begin{aligned}
F_Z(a) &= P(X+Y \leq a) \\
&= \int\int_{x+y\leq a} f_{XY}(x,y)dxdy \\
&= \int\int_{x+y\leq a} f_X(x)f_Y(y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_X(x)dx f_Y(y)dy \\
&= \int_{-\infty}^{\infty} F_X(a-y)f_Y(y)dy
\end{aligned}$$

By differentiating the cdf, we obtain the pdf of $Z = X+Y$:

$$\begin{aligned}
f_Z(a) &= \frac{d}{da}\int_{-\infty}^{\infty} F_X(a-y)f_Y(y)dy \\
&= \int_{-\infty}^{\infty} \frac{d}{da}F_X(a-y)f_Y(y)dy \\
&= \int_{-\infty}^{\infty} f_X(a-y)f_Y(y)dy.
\end{aligned}$$

Group project (optional):

- *Sum of independent Poisson random variables. Suppose X and Y are independent Poisson random variables with respective means λ_1 and λ_2. Show that $Z = X + Y$ has a Poisson distribution with mean $\lambda_1 + \lambda_2$.*
- *Sum of independent uniform random variables. If X and Y are two independent random variables, both uniformly distributed on (0,1), calculate the pdf of $Z = X + Y$.*

Example 9 Let X and Y denote the lifetimes of two bulbs. Suppose that X and Y are independent and that each has an exponential distribution with $\lambda = 1$ (year).

- What is the probability that each bulb lasts at most 1 year?

- What is the probability that the total lifetime is between 1 and 2 years?

2.3 Conditional Distribution

The use of conditional distribution allows us to define conditional probabilities of events associated with one random variable when we are given the value of a second random variable.

(1) The Discrete Case: Suppose X and Y are discrete random variables. The conditional probability mass function of X given $Y = y_j$ is the conditional probability distribution of X given $Y = y_j$. The conditional probability mass function of $X|Y$ is

$$\begin{aligned} p_{X|Y}(x_i|y_j) &= P(X = x_i|Y = y_j) \\ &= \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} \\ &= \frac{p_{X,Y}(x_i, y_j)}{p_Y(y_j)} \end{aligned}$$

This is just the conditional probability of the event $\{X = x_i\}$ given that $\{Y = y_i\}$.

Example 10 Considered the situation that a fair coin is tossed three times independently. Let X denote the number of heads on the first toss and Y denote the total number of heads.

- What is the conditional probability mass function of X given Y?
- Are X and Y independent?

If X and Y are independent random variables, then the conditional probability mass function is the same as the unconditional one. This follows because if X is independent of Y, then

$$\begin{aligned} p_{X|Y}(x|y) &= P(X = x|Y = y) \\ &= \frac{P(X = x, Y = y)}{P(Y = y)} \\ &= \frac{P(X = x)P(Y = y)}{P(Y = y)} \\ &= P(X = x) \end{aligned}$$

(2) The Continuous Case: If X and Y have a joint probability density function $f(x, y)$, then the conditional pdf of X, given that $Y = y$, is defined for all values of y such that $f_Y(y) > 0$, by

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}. \tag{2.9}$$

To motivate this definition, multiply the left-hand side by dx and the right hand side by $(dxdy)/dy$ to obtain

$$\begin{aligned} f_{X|Y}(x|y)dx &= \frac{f_{X,Y}(x, y)dxdy}{f_Y(y)dy} \\ &\approx \frac{P\{x \le X \le x + dx, y \le Y \le y + dy\}}{P\{y \le Y \le y + dy\}} \\ &= P\{x \le X \le x + dx | y \le Y \le y + dy\}. \end{aligned}$$

In other words, for small values of dx and dy, $f_{X|y}(x|y)$ represents the conditional probability that X is between x and $x + dx$ given that Y is between y and $y + dy$.

That is, if X and Y are jointly continuous, then for any set A,

$$P\{X \in A|Y = y\} = \int_A f_{X|Y}(x|y)dx.$$

In particular, by letting $A = (-\infty, a]$, we can define the conditional cdf of

X given that $Y = y$ by

$$F_{X|Y}(a|y) = P(X \leq a|Y = y) = \int_{-\infty}^{a} f_{X|Y}(x|y)dx.$$

Example 11 Let X and Y be random variables with joint distribution $f(x, y) = \frac{6}{5}(x + y^2)$, for $0 \leq x \leq 1$ and $0 \leq y \leq 1$ and $f(x, y) = 0$ otherwise.

- What is the conditional pdf of Y given that $X = 0.8$?
- What is the conditional expectation of Y given that $X = 0.8$?

Group project (optional): Suppose X and Y are two independent random variables, both uniformly distributed on (0,1). Let $T_1 = \min(X, Y)$ and $T_2 = \max(X, Y)$.

- *What is the joint distribution of T_1 and T_2?*
- *What is the conditional distribution of T_2 given that $T_1 = t$?*
- *Are T_1 and T_2 independent?*

3 Expected Values

Let X and Y be jointly distributed rv's with pmf $p(x,y)$ or pdf $f(x,y)$ according to whether the variables are discrete or continuous. Then the expected value of a function $h(X,Y)$, denoted by $E[h(X,Y)]$, is given by

$$E[h(X,Y)] = \begin{cases} \sum_x \sum_y h(x,y)p(x,y) & \text{Discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y)f(x,y)dxdy & \text{Continuous} \end{cases} \tag{3.10}$$

Example 12 The joint pdf of X and Y is

$$f(x,y) = \begin{cases} 24xy & 0 < x < 1, 0 < y < 1, x+y \le 1 \\ 0 & \text{otherwise} \end{cases} . \tag{3.11}$$

Let $h(X,Y) = 0.5 + 0.5X + Y$. Find $E[h(X,Y)]$.

Rule of expected values: If X and Y are independent random variables, then we have

$$E(XY) = E(X)E(Y).$$

This is in general *not* true for correlated random variables.

Example 13 Five friends purchased tickets to a certain concert. If the tickets are for seats 1-5 in a particular row and the ticket numbers are randomly selected (without replacement). What is the expected number of seats separating any particular two of the five?

4 Covariance and Correlation

Covariance and correlation are related parameters that indicate the extent to which two random variables co-vary. Suppose there are two technology stocks. If they are affected by the same industry trends, their prices will tend to rise or fall together. They co-vary. Covariance and correlation measure such a tendency. We will begin with the problem of calculating the expected values of a function of two random variables.

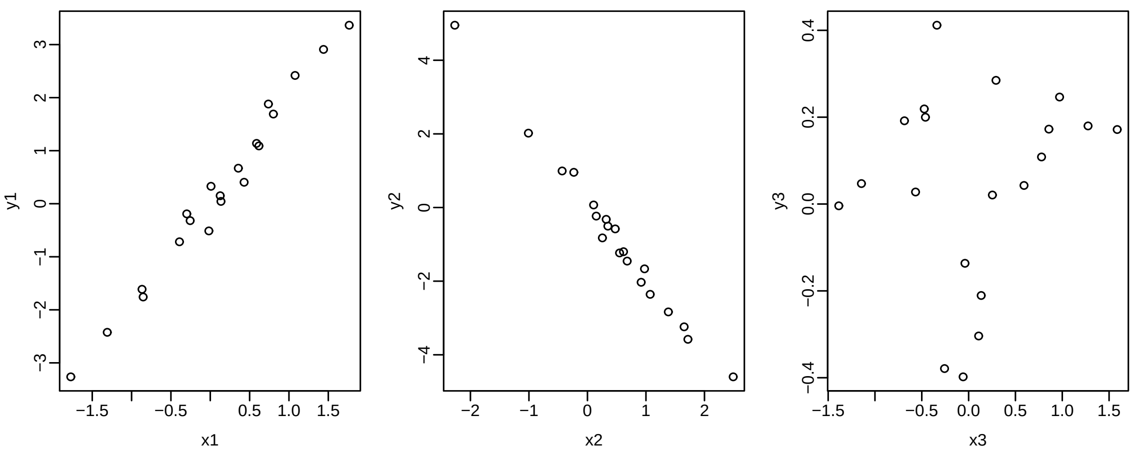

Figure 1: Examples of positive, negative and zero correlation

Examples of positive and negative correlations:

- price and size of a house
- height and weight of a person
- income and highest degree received
- Lung capacity and the number of cigarettes smoked everyday.
- GPA and the hours of TV watched every week
- Expected length of lifetime and body mass index

4.1 Covariance

When two random variables are not independent, we can measure how strongly they are related to each other. The *covariance* between two rv's X and Y is

$$
\begin{aligned}
\mathrm{Cov}(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)] \\
&= \begin{cases} \sum_x \sum_y (x-\mu_X)(y-\mu_Y)p(x,y) & X, Y \text{ discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f(x,y)dxdy & X, Y \text{ continuous} \end{cases}
\end{aligned}
$$

Variance of a random variable can be view as a special case of the above definition: $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$.

Properties of covariance:

1. A shortcut formula: $\mathrm{Cov}(X,Y) = E(XY) - \mu_X\mu_Y$.

2. If X and Y are independent, then $\mathrm{Cov}(X,Y) = 0$.

3. $\mathrm{Cov}(aX + b, cY + d) = ac\mathrm{Cov}(X,Y)$.

Variance of linear combinations: Let X, Y be two random variables, and a and b be two constants, then

$$\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\mathrm{Cov}(X, Y).$$

Special case: when X and Y are independent, then

$$\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y).$$

Example 14 What is $\mathrm{Cov}(X, Y)$ for (X, Y) with joint pmf

$p(x,y)$	$y = 0$	$y = 100$	$y = 200$
$x = 100$	0.20	0.10	0.20
$x = 250$	0.05	0.15	0.30

?

Example 15 The joint pdf of X and Y is $f(x, y) = 24xy$ when $0 \leq x \leq 1$, $0 \leq y \leq 1$ and $x + y \leq 1$, and $f(x, y) = 0$ otherwise. Find $\mathrm{Cov}(X, Y)$.

4.2 Correlation Coefficient

The defect of covariance is that its computed value depends critically on the units of measurement (e.g., kilograms versus pounds, meters versus feet). Ideally, the choice of units should have no effect on a measure of strength of relationship. This can be achieved by scaling the covariance by the standard deviations of X and Y.

The correlation coefficient of X and Y, denoted by $\rho_{X,Y}$, is defined by

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}. \tag{4.1}$$

Example 16 Calculate the correlation of X and Y in Examples 14 and 15.

The correlation coefficient is not affected by a linear change in the units of measurements. Specifically we can show that

- If a and c have the same sign, then $\text{Corr}(aX+b, cY+d) = \text{Corr}(X, Y)$.

- If a and c have opposite signs, then $\text{Corr}(aX+b, cY+d) = -\text{Corr}(X, Y)$.

Additional properties of correlation:

1. For any two rv's X and Y, $-1 \leq \text{Corr}(X, Y) \leq 1$.
2. If X and Y are independent, then $\rho = 0$.

Correlation and dependence: Zero correlation coefficient does not imply that X and Y are independent, but only that there is complete absence of a linear relationship. When $\rho = 0$, X and Y are said to be uncorrelated. Two random variables could be uncorrelated yet highly dependent.

Correlation and Causation: A large correlation does not imply that increasing values of X causes Y to increase, but only that large X values are associated with large Y values. Examples:

- Vocabulary and cavities of children.
- Lung cancer and yellow finger.