Multivariate probability distributions
Outline of Topics

1. Background
2. Discrete bivariate distribution
3. Continuous bivariate distribution
Multivariate analysis

- When one measurement is made on each observation in a dataset, **univariate** analysis is used, *e.g.*, survival time of patients.

- If more than one measurement is made on each observation, a **multivariate** analysis is used, *e.g.*, survival time, age, cancer subtype, size of cancer, *etc.*

- We focus on **bivariate** analysis, where exactly two measurements are made on each observation.

- The two measurements will be called $X$ and $Y$. Since $X$ and $Y$ are obtained for each observation, the data for one observation is the pair $(X, Y)$. 
Bivariate data can be represented as:

<table>
<thead>
<tr>
<th>Observation</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$X_1$</td>
<td>$Y_1$</td>
</tr>
<tr>
<td>2</td>
<td>$X_2$</td>
<td>$Y_2$</td>
</tr>
<tr>
<td>3</td>
<td>$X_3$</td>
<td>$Y_3$</td>
</tr>
<tr>
<td>4</td>
<td>$X_4$</td>
<td>$Y_4$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>n</td>
<td>$X_n$</td>
<td>$Y_n$</td>
</tr>
</tbody>
</table>

Each observation is a pair of values, e.g., $(X_4, Y_4)$ is the 4-th observation.

$X$ and $Y$ can be both discrete, both continuous, or one discrete and one continuous. We focus on the first two cases.

Some examples:
- $X$ (survived $> 1$ year) and $Y$ (cancer subtype) of each patient in a sample
- $X$ (length of job training) and $Y$ (time to find a job) for each unemployed individual in a job training program
- $X$ (income) and $Y$ (happiness) for each individual in a survey
Bivariate distributions

- We can study $X$ and $Y$ separately, i.e., we can analyse $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ separately using probability distribution function, probability density function or cumulative distribution function. These are examples of univariate analyses. When $X$ and $Y$ are studied separately, their distribution and probability are called **marginal**.

- When $X$ and $Y$ are considered together, many interesting questions can be answered, e.g.,
  - Is subtype I cancer ($X$) associated with a higher chance of survival beyond 1 year ($Y$)?
  - Does longer job training ($X$) result in shorter time to find a job ($Y$)?
  - Do people with higher income ($X$) lead a happier life ($Y$)?

- The joint behavior of $X$ and $Y$ is summarized in a **bivariate probability distribution**. A bivariate distribution is an example of a **joint distribution**.
Review of a discrete distribution: Drawing a marble from an urn

Probability distribution tells us the long run frequency for \( \bullet \) is higher than \( \circ \).
Discrete bivariate distribution - Drawing 2 marbles with replacement

- **Draw 1 (X)**
  - 1
  - 2
  - 3
  - 4
  - 5

- **Draw 2 (Y)**
  - 1
  - 2
  - 3

**Probabilities:**

\[
P(X = \bullet \text{ and } Y = \bullet) = P(\bullet, \bullet) = \left(\frac{3}{5}\right) \left(\frac{3}{5}\right) = \frac{9}{25}
\]

\[
P(X = \bullet \text{ and } Y = \bullet) = P(\bullet, \bullet) = \left(\frac{2}{5}\right) \left(\frac{3}{5}\right) = \frac{6}{25}
\]

\[
P(\bullet, \bullet) + P(\bullet, \bullet) + P(\bullet, \bullet) + P(\bullet, \bullet) = \frac{9}{25} + \frac{6}{25} + \frac{6}{25} + \frac{4}{25} = 1
\]
Discrete bivariate distribution - Drawing 2 marbles without replacement

\[
P(X = \text{blue} \text{ and } Y = \text{green}) = P(\text{blue}, \text{blue}) = \left(\frac{2}{4}\right) \left(\frac{3}{5}\right) = \frac{6}{20}
\]
\[
P(X = \text{green} \text{ and } Y = \text{blue}) = P(\text{green}, \text{blue}) = \left(\frac{3}{4}\right) \left(\frac{2}{5}\right) = \frac{6}{20}
\]
\[
P(\text{blue}, \text{blue}) + P(\text{green}, \text{blue}) + P(\text{blue}, \text{green}) + P(\text{green}, \text{green}) = \frac{6}{20} + \frac{6}{20} + \frac{6}{20} + \frac{2}{20} = 1
\]
A **discrete bivariate distribution** is used to model the joint behavior of two variables, \( X \) and \( Y \), both of which are discrete.

\( X \) and \( Y \) are **discrete random variables** if there is a countable number of possible values for \( X \): \( a_1, a_2, ..., a_k \) and for \( Y \): \( b_1, b_2, ..., b_l \).

\((X, Y)\) is the unknown outcome if we randomly draw an observation from the population.

\( P(X = a_i, Y = b_j) \) is the **joint probability distribution function** of observing \( X = a_i, Y = b_j \).

A valid joint probability distribution function must satisfy the following rules:

- \( P(X = a_i, Y = b_j) \) must be between 0 and 1

- We are certain that one of the values will appear, therefore:
  \[ P[(X = a_1, Y = b_1) \text{ or } (X = a_2, Y = b_1) \text{ or } ... \text{ or } (X = a_k, Y = b_l)] = P(X = a_1, Y = b_1) + P(X = a_2, Y = b_1) + ... + P(X = a_k, Y = b_l) = 1 \]
Discrete joint distribution: Example 1

\[ P(X = a, Y = b) = \frac{a+b}{48}, \text{ if } a, b = 0, 1, 2, 3 \]

\[ \begin{array}{cccc|c}
X & 0 & 1 & 2 & 3 & P(X = a) \\
\hline
0 & \frac{6}{48} & \frac{10}{48} & \frac{14}{48} & \frac{18}{48} & 1 \\
1 & \frac{1}{48} & \frac{2}{48} & \frac{3}{48} & \frac{4}{48} & \frac{10}{48} \\
2 & \frac{2}{48} & \frac{3}{48} & \frac{4}{48} & \frac{5}{48} & \frac{14}{48} \\
3 & \frac{3}{48} & \frac{4}{48} & \frac{5}{48} & \frac{6}{48} & \frac{18}{48} \\
\hline
Y & 0 & 1 & 2 & 3 & P(Y = b) \\
\hline
0 & \frac{48}{48} & \frac{48}{48} & \frac{48}{48} & \frac{48}{48} & 1 \\
1 & \frac{1}{48} & \frac{2}{48} & \frac{3}{48} & \frac{4}{48} & \frac{10}{48} \\
2 & \frac{2}{48} & \frac{3}{48} & \frac{4}{48} & \frac{5}{48} & \frac{14}{48} \\
3 & \frac{3}{48} & \frac{4}{48} & \frac{5}{48} & \frac{6}{48} & \frac{18}{48} \\
\end{array} \]

- \( P(X = a, Y = b) \) are the **joint probabilities**
- \( P(X = a) \), \( P(Y = b) \) are called **marginal probabilities**. \( P(X = a) \) gives us information about \( X \) ignoring \( Y \) and \( P(Y = b) \) gives us information about \( Y \) ignoring \( X \)
- We can always find marginal probabilities from joint probabilities (as in Example 1) but not the other way around unless \( X \) and \( Y \) are independent (see next slide)
Discrete joint distribution - Independence

- $X$ and $Y$ are **independent** if, for all $X = a, Y = b$:

\[ P(X = a | Y = b) = P(X = a) \]

\[ \iff P(Y = b | X = a) = P(Y = b) \]

\[ \iff P(X = a, Y = b) = P(X = a)P(Y = b) \]

- $P(Y = b | X = a)$ and $P(X = a | Y = b)$ are **conditional probabilities**

- If $X$ and $Y$ are independent, then we can easily

  (a) calculate $P(X = a, Y = b)$ by $P(X = a)P(Y = b)$

  (b) write $P(X = a | Y = b)$ as $P(X = a)$

  (c) write $P(Y = b | X = a)$ as $P(Y = b)$
Try,

\[
P(Y = 1|X = 3) = \frac{P(Y = 1, X = 3)}{P(X = 3)} = \frac{4/48}{18/48} = \frac{4}{18} \neq P(Y = 1) = \frac{10}{48}.
\]

Alternatively, try

\[
P(X = 3, Y = 1) = \frac{4}{48} \neq P(X = 3)P(Y = 1) = \frac{18}{48} \times \frac{10}{48}
\]

Either way is sufficient to show \(X\) and \(Y\) are not independent. Furthermore, we can try any combination of \(X = a, Y = b\) to disprove independence.
Discrete joint distribution: Example 2

\[ P(X = a, Y = b) = \frac{ab}{18}, \text{ if } a = 1, 2, 3; b = 1, 2 \]

\[
\begin{array}{c|cc}
\mathbf{X} & 1 & 2 \\
\hline
1 & \frac{1}{18} & \frac{2}{18} \\
2 & \frac{2}{18} & \frac{4}{18} \\
3 & \frac{3}{18} & \frac{6}{18} \\
\hline
P(Y = b) & \frac{6}{18} & \frac{12}{18} & 1
\end{array}
\]

\[ P(X = a) = \frac{3}{18}, \frac{6}{18}, \frac{9}{18} \]

\[ P(Y = b) = \frac{3}{18} + \frac{6}{18} \]

\[ P(X = 1, Y = 1) = \frac{1}{18} = P(X = 1)P(Y = 1) = \frac{3}{18} \times \frac{6}{18} = \frac{18}{18^2} = \frac{1}{18} \]

\[ \vdots \]

\[ P(X = 3, Y = 2) = \frac{6}{18} = P(X = 3)P(Y = 2) = \frac{9}{18} \times \frac{12}{18} = \frac{108}{18^2} = \frac{6}{18} \]

To show independence, we must show \( P(X = a, Y = b) = P(X = a)P(Y = b) \) for all combinations of \( X = a, Y = b \)
Probability under a univariate probability density function (PDF)

- \( P(X \leq 1) \) can be found by integration:

\[
P(X \leq 1) = \int_{0}^{1} f(x) \, dx
\]

\[
= \int_{0}^{1} 1.5 e^{-1.5x} \, dx
\]

\[
= \left[ -e^{-1.5x} \right]_{0}^{1}
\]

\[
= 1 - e^{-1.5}
\]

\[
\approx 0.776
\]

- It turns out, for any \( x > 0 \),

\[
P(X \leq x) = 1 - e^{-1.5x}
\]

- we often write \( P(X \leq x) \) as \( F(x) \) and call \( F(x) \) the cumulative distribution function (CDF)

- \( F(1) \) is a probability but \( f(1) \) (a point on \( f(x) \)) is not a probability
The univariate cumulative distribution function (CDF)

- $F(x)$ can be used to find $P(X \leq x)$ for any $x$, e.g., $F(1) = P(X \leq 1)$

- A plot of $F(x)$ is a convenient way for finding probabilities.

- Probability is found by drawing a line (-----) from the horizontal axis until it meets the CDF and then drawing a horizontal line until it meets the vertical axis.

- All CDF plots have an asymptote at 1 (-----): $F(x) \leq 1$ because it is a probability.
A **continuous bivariate distribution** is used to model the joint behavior of two variables, $X$ and $Y$, both of which are continuous.

$X$ and $Y$ are **continuous random variables** if the possible values of $X$ fall in a range $(a, b) \subseteq (-\infty, \infty)$ and those for $Y$ are in a range $(c, d) \subseteq (-\infty, \infty)$.

The bivariate distribution of $(X, Y)$ is defined by a **joint** PDF, $f(x, y)$, with the characteristics:

- $f(x, y) \geq 0$ if $(x, y) \in (a, b) \times (c, d)$, $f(x, y) = 0$ otherwise
- $f(x, y) \neq P(X = x, Y = y) = 0$ for all values of $x, y$
- $P(a \leq X \leq b, c \leq Y \leq d) = 1$ since $X$ must be in $(a, b)$ and $Y$ must be in $(c, d)$
- $F(x, y) = P(X \leq x, Y \leq y)$ is the joint CDF
The joint CDF

Probabilities are often found by manipulating the joint CDF $F(x, y)$, obtained using *double integration* on the joint PDF $f(x, y)$:

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{y \leq y} \int_{x \leq x} f(x, y) \, dx \, dy = \int_{x \leq x} \int_{y \leq y} f(x, y) \, dy \, dx.$$

**Example 1** $f(x, y) = e^{-y-x}, \ x, y > 0$.

The red shaded figure is

$$F(a, b) = \int_{Y \leq b} \int_{X \leq a} e^{-y-x} \, dx \, dy$$

$$= \int_{0}^{b} e^{-y} \left[ \int_{0}^{a} e^{-x} \, dx \right] \, dy$$

$$= \int_{0}^{b} e^{-y} \left[ -e^{-x} \right]_{0}^{a} \, dy$$

$$= \int_{0}^{b} e^{-y} \left[ 1 - e^{-a} \right] \, dy$$

$$= [1 - e^{-a}] \int_{0}^{b} e^{-y} \, dy$$

$$= [1 - e^{-a}] \left[ -e^{-y} \right]_{0}^{b}$$

$$\Rightarrow F(x, y) = [1 - e^{-x}][1 - e^{-y}], \ x, y > 0$$

Suppose we wish to find $P(X \leq 2, Y \leq 1)$:

$$P(X \leq 2, Y \leq 1) \equiv F(2, 1) = [1 - e^{-2}][1 - e^{-1}] \approx 0.547$$
Example 2 \( f(x, y) = 2, \ 0 < x < y, 0 < y < 1 \)

\[
F(a, b) = \int_{y \leq b} \int_{x \leq a} 2 \, dx \, dy
\]

\[
= \int_0^a \left[ \int_x^b 2 \, dy \right] \, dx
\]

\[
= \int_0^a [2y]_x^b \, dx
\]

\[
= \int_0^a 2(b - x) \, dx
\]

\[
= [2bx - x^2]_0^a
\]

\[
= 2ba - a^2
\]

\( \Rightarrow F(x, y) = 2yx - x^2, \ 0 < x < y, 0 < y < 1 \)

Suppose we wish to find \( P(X \leq \frac{1}{2}, Y \leq \frac{3}{4}) : \)

\[
P \left( X \leq \frac{1}{2}, Y \leq \frac{3}{4} \right) \equiv F \left( \frac{1}{2}, \frac{3}{4} \right)
\]

\[
= 2 \left( \frac{3}{4} \right) \left( \frac{1}{2} \right) - \left( \frac{1}{2} \right)^2 = \frac{1}{2}
\]
Marginal PDF and CDF

- As in the discrete case (cf. slide 10), the marginal PDFs \( f(x) \) and \( f(y) \) for \( X \) and \( Y \) can be obtained from the joint PDF \( f(x, y) \)

- To find marginal probabilities of \( X \) and \( Y \), we need the marginal CDFs \( F(x) \) and \( F(y) \)

- Two ways to obtain marginal CDFs
  1. \( F(x) \) can be obtained from \( f(x) \); similarly for \( F(y) \) (cf. slide 15)
  2. Using joint CDF \( F(x, y) \):

\[
F(x) = F(x, \infty), \quad F(y) = F(\infty, y)
\]

Reason:

\[
F(x) \equiv P(X \leq x) \\
= P(X \leq x, \text{ don't care about } Y) \\
= P(X \leq x, \forall Y) \\
= P(X \leq x, Y \leq \infty) \equiv F(x, \infty)
\]

Same reasoning for \( F(y) = F(\infty, y) \)
### Marginal CDF: Example 2

Given: $f(x, y) = 2, \quad F(x, y) = 2yx - x^2, \quad 0 < x < y, \quad 0 < y < 1$

#### 1. $f(x)$

\[
f(x) = \int_{-\infty}^{\infty} f(x, y) \, dy \\
= \int_{0}^{1} 2 \, dy \\
= 2(1 - x), \quad 0 < x < 1
\]

\[
F(x) = \int_{0}^{x} 2(1 - x) \, dx \\
= 2x - x^2, \quad 0 < x < 1
\]

#### 2. $F(x)$

\[
F(x) = F(x, \infty) \\
= F(x, 1), \quad \text{since } y < 1 \\
= 2(1)x - x^2 \\
= 2x - x^2, \quad 0 < x < 1
\]
Conditional probabilities

\[ P(X \leq x|Y \leq y) = \frac{P(X \leq x, Y \leq y)}{P(Y \leq y)} = \frac{F(x, y)}{F(y)} \]

\[ P(Y \leq y|X \leq x) = \frac{P(X \leq x, Y \leq y)}{P(X \leq x)} = \frac{F(x, y)}{F(x)} \]

Example 2: \( F(x, y) = 2yx - x^2, \quad F(y) = y^2, \quad 0 < x < y, \quad 0 < y < 1 \)

\[ P(Y \leq 0.8) \]

\[ P(X \leq x|Y \leq y) = \frac{2yx - x^2}{y^2} \]

\[ P(X \leq 0.4|Y \leq 0.8) = \frac{2(0.8)(0.4) - (0.4)^2}{0.8^2} = 0.75 \]
Conditional probabilities (2)

\[ P(X \leq x | Y = y) = \int_{x \leq x} f(x|y) \, dx = \int_{x \leq x} \frac{f(x, y)}{f(y)} \, dx \]

\( f(x|y) \) is called a **conditional density**. \( P(X \leq x | Y = y) \) refers to the probability of the event “\( X \leq x \)”, given the information “\( Y = y \)” and hence it is defined despite \( P(Y = y) = 0 \). We can similarly define \( P(Y \leq y | X = x) \).

**Example 2** \( f(x, y) = 2, \ f(y) = 2y, \ 0 < x < y, \ 0 < y < 1 \)

\[ f(x, Y = 0.8) \]

\[ f(x, Y = 0.8) \]

\[ P(X \leq 0.4 | Y = 0.8) \]

\[ P(X \leq 0.4 | Y = 0.8) = \frac{0.4}{0.8} = 0.5 \]

\[ P(X \leq a | Y = y) = \int_{x \leq a} \frac{2}{2y} \, dx \]

\[ = \int_{0}^{a} \frac{1}{y} \, dx \]

\[ = \left[ \frac{x}{y} \right]_{0}^{a} \]

\[ = \frac{x}{y} \]

\[ = \frac{0.4}{0.8} = 0.5 \]
**Independence**

*A and B are independent if and only if*

\[
P(A \text{ and } B) = P(A)P(B)
\]

Analogously, if *X* and *Y* are independent, we can simplify \(P(X \leq x, Y \leq y)\) as \(P(X \leq x)P(Y \leq y)\). Two ways of proving independence

**Method 1**

\[
P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)
\]

\[
\iff
\]

\[
F(x, y) = F(x)F(y)
\]

\[
\iff
\]

\[
f(x, y) = f(x)f(y)
\]

**Method 2** Method 1 requires us to know the marginal PDFs and CDFs. A simpler way to establish independence is to show that \(f(x, y)\) can be factorized as

\[
f(x, y) = g(x)h(y)
\]

for *any* positive functions \(g, h\)
Example 1

\[ f(x, y) = e^{-y-x}, \quad F(x, y) = [1 - e^{-x}][1 - e^{-y}], \quad x, y > 0 \]

Method 1

\[ f(x, y) = e^{-x-y} = \frac{e^{-x}}{f(x)} \cdot \frac{e^{-y}}{f(y)}; \quad F(x, y) = \left[1 - e^{-x}\right]\left[1 - e^{-y}\right] \]

where \( f(x), x > 0, f(y), y > 0 \) and \( F(x), x > 0, F(y), y > 0 \) are the marginal PDFs and CDFs.

Method 2

\( f(x, y) \) can also be factorized as \( f(x, y) = g(x)h(y) \), where

\[
\begin{align*}
    g(x) &= 2e^{-x}, \quad 0 < x < 1 \\
    h(y) &= \frac{1}{2}e^{-y}, \quad 0 < y < 1
\end{align*}
\]

Using either method, we show \( X \) and \( Y \) are independent.
Example 2

\( f(x, y) = 2, \quad F(x, y) = 2yx - x^2, \quad 0 < x < y, \quad 0 < y < 1 \)

**Method 1** Let \( I \) be the indicator function such that \( I(A) = 1 \) if \( A \) is true, and 0 otherwise, then

\[
  f(x, y) = 2I(0 < x < y < 1) \neq f(x)f(y) = [2(1 - x)][2y].
\]

Similarly,

\[
  F(x, y) = [2yx - x^2]I(0 < x < y < 1) \neq F(x)F(y) = [2x - x^2][y^2].
\]

**Method 2** \( I(0 < x < y < 1) \) cannot be factorized into two functions \( g(x) \) and \( h(y) \)

Using either method, \( X \) and \( Y \) are not independent