Imperial College London

Mathematics: Year 2

# Joint Distributed Random Variables

Xin Wang

September 21, 2020

**Abstract**

Previous chapters have covered concepts to group data together in a coherent format such as cumulative distribution function and probability density function. This chapter combines probability and concepts like CDF and PDF.

Given random variables $X$ and $Y$ that are defined on a probability space, the joint probability distribution for $X$ and $Y$ is a probability distribution that gives the probability in that each of the random variables $X$ and $Y$ falls in any particular range or discrete set of specified values.

# Contents

# 1 Introduction

The previous chapter has only seen situations with single random variables but useful probability statements often involve more than one random variable, usually pairs of random variables. For example, in signal transmission, there will be a random variable $X$ to be the number of high- quality signals received and the random variable $Y$ to be the number of low-quality signals received - it is of interest the probabilities that can be expressed in terms of both $X$ and $Y$.

# 2 Two discrete random variables

## 2.1 Joint probability distribution

When dealing with **pairs** of random variables, the joint probability distribution is called **bivariate distribution**.

**Case Study**: In a new receiver transmitting digital information, each received bit is classified as: Acceptable - *Probability* 0.9, Unacceptable - *Probability* 0.08 and Suspect - *Probability* 0.02. The ratings of each bit are **independent**.

Four bits are transmitted. In the case study, the notations are defined as:

- $X$ - Number of acceptable bits.

  The distribution of $X$ is **binomial** where $n = 4$ and $p = 0.9$.

- $Y$ - Number of suspect bits.

  The distribution of $X$ is **binomial** where $n = 4$ and $p = 0.08$.

The probabilities in a joint probability distribution is determined by the following example: Find the probability that **two acceptable bits** and **one suspect bit** are received among the four bits transmitted - $P(X = 2, Y = 1)$ - assuming the bits are independent of each other.

1. Find the probability $P(aasu)$:

$$P(aasu) = 0.9 \times (0.9) \times (0.08) \times (0.02) = 0.0013$$

2. Find the number of possible sequences with two $a$ and one $s$:

$$\frac{4!}{2! \times 1! \times 1!} = 12$$

3. Find the joint probability:

$$P(aasu) = f_{XY}(2,1) = P(X = 2, Y = 1) = 12(0.0013) = 0.0156$$

The following image describes the set of points $(x, y)$ in the range of $(X, Y)$ along with the probability of each point within the **sample space** $S$.

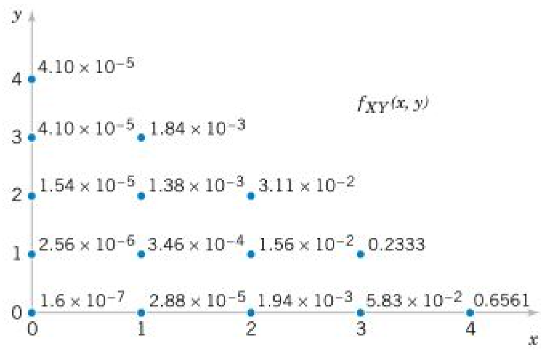



Figure 1: Joint Probability Distribution of $X$ and $Y$

**Joint probability distribution**: The probability distributions that defines how probability is assigned to all pairs of random variables $X$ and $Y$.

$$F_{X,Y}(a,b) = P(X \leq a, Y \leq b) \text{ where } -\infty < a \text{ and } b < \infty$$

The joint probability distribution must satisfy the **characteristics of discrete probability**:

1. PMF must be non-negative.
2. Sum of probabilities must be 1.

## 2.2 Marginal probability distributions (Marginal PMF)

When more than one random variable is defined in a random experiment, it is important to distinguish between the **joint probability distribution of $X$ and $Y$**, and the **probability distribution of each individual variable**.

The **marginal probability distribution** of $X$ can be determined from the **joint probability distribution** of $X$ by summing the joint PMF across the range of the other variable.

**Example 1**: Using the joint probability distribution of $X$ and $Y$ in Fig 1, find the marginal probability distribution of $X = 3$.

$$\begin{aligned} P(X = 3) &= P(X = 3,\ Y = 0) + P(X = 3,\ Y = 1) \\ &= 0.0583 + 0.2333 \\ &= 0.292 \end{aligned}$$

Which is verified by:

$$P(X = 3) = \binom{4}{3} \times 0.9^3 \times 0.1^1 = 0.292$$

**Marginal probability distribution** ($f_X(x)$): The individual probability distribution of a random variable $X$.

- For random variable $X$: Summing probabilities in **each column**.

$$f_X(x) = P(X = x) = \sum_{R_x} f_{XY}(x,\ y)$$

where $R_x$ denotes the set of all points in the range of $(X, Y)$ for $X = x$.

- For random variable $Y$: Summing probabilities in **each row**.

$$f_Y(y) = P(Y = y) = \sum_{R_y} f_{XY}(x,\ y)$$

where $R_y$ denotes the set of all points in the range of $(X, Y)$ for $Y = y$.

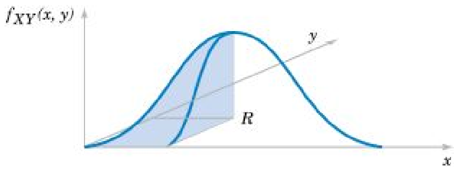
# 3 Two continuous random variables

## 3.1 Joint probability distribution

The joint probability distribution of two continuous random variables is similar to the concept of two discrete random variables. For example, consider injection modelling where the following **continuous random variables** denote: $X$: The length of one dimension of an injection-molded part and $Y$: the length of another dimension. The **sample space** of the random experiment consists of points in two dimensions.

Similar to the **probability density function** of **a single continuous random variable**, the **joint probability density function** of **two continuous random variable** can be defined over 2-D space.

The double integral of over a region $R$ provides the probability that $(X, Y)$ assumes a value in $R$. This integral can be interpreted as the volume under the surface over the region $R$.




Probability that $(X, Y)$ is in the region $R$ is determined by the volume of $f_{XY}(x, y)$ over the region $R$.

Figure 2: Joint probability density function for random variables $X$ and $Y$

**Jointly continuous**: Two random variables X and Y are jointly continuous if there exists a nonnegative function $f_{XY} : \Re^2 \to \Re$, such that, for any set $A \in \Re^2$:

$$P[(X, Y) \in A] = \int \int_A f_{XY}(x, y) dx\, dy$$

**Joint probability density function** $f_{XY}(x, y)$: A function that satifies the following:

- $f_{XY}(x, y) \geq 0$ for all $x, y$
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dx\, dy = 1$
- For any region $R$ of $2D$ space:

$$P([X, Y] \in \Re) = \int_R \int f_{XY}(x, y) dx\, dy$$

**Example 1**: Let the random variable $X$ denote the time until a computer server connects to your machine (in milliseconds), and let $Y$ denote the time until the server authorizes you as a valid user (in milliseconds). Each of these random variables measures the wait from a common starting time and $X < Y$.

Assume that the joint probability density function for X and Y:

$$f_{XY}(x, y) = 6 \times 10^{-6} e^{(-0.001x - 0.002y)} \text{ for } x < y$$

If $X$ and $Y$ are continuous with a joint cumulative distribution function $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$, the **joint probability density function** is defined as:

$$f_{X,Y}(x, y) = \frac{\partial^2 F_(X, Y)(x, y)}{\partial x \partial y}$$

- The region with nonzero probability is shaded. The property that this joint probability density function integrates to 1 can be verified by the integral of $f_{XY}(x, y)$ over this region.

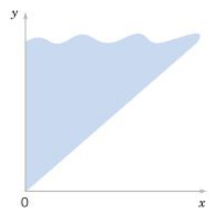
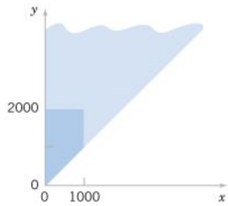



Figure 3: The joint probability density function of $X$ and $Y$ is non-zero over the shaded region

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)dy\,dx = \int_{0}^{\infty}\left(\int_{x}^{\infty} 6\times 10^{-6} e^{(-0.001x-0.002y)dy}\right)dx$$
$$= 6\times 10^{-6}\int_{0}^{\infty}\left(\int_{x}^{\infty} e^{-0.002x}dy\right)e^{-0.001x}dx$$
$$= 6\times 10^{-6}\int_{0}^{\infty}\left(\frac{e^{-0.002y}}{0.002}\right)e^{-0.001x}dx$$
$$= 0.003\left(\int_{0}^{\infty} e^{-0.003x}dx\right)$$
$$= 0.003\left(\frac{1}{0.003}\right) = 1$$

- The probability that $X < 1000$ and $Y < 2000$ is determined as the integral over the darkly shaded region:




$$P(X \leq 1000, Y \leq 2000) = \int_{0}^{1000}\int_{x}^{2000} f_{XY}(x,y)dy\,dx$$
$$= 6\times 10^{-6}\int_{0}^{1000}\left(\int_{x}^{2000} e^{-0.002y}dy\right)e^{-0.001x}dx$$
$$= 6\times 10^{-6}\int_{0}^{1000}\left(\frac{e^{-0.002x}-e^{-4}}{0.002}\right)e^{-0.001x}dx$$
$$= 0.003\int_{0}^{1000} e^{-0.003x} - e^{-4}e^{-0.001x}dx$$
$$= 0.003\left[\left(\frac{1-e^{-3}}{0.003}\right) - e^{-4}\left(\frac{1-e^{-1}}{0.001}\right)\right]$$
$$= 0.003(316.738 - 11.578) = 0.915$$

## 3.2 Marginal probability distributions (Marginal PMF)

**The marginal probability density**: The probability distribution of random variables.

The marginal probability density functions of $X$ and $Y$, denoted by $f_X(x)$ and $f_Y(y)$ respectively, are obtained from the joint PDF:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \text{ for } -\infty < x < \infty$$
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx \text{ for } -\infty < y < \infty$$

As the marginal probability density describes the probability distribution of random variables, the marginal density can be used to compute probabilities.

**Example 1**: Consider continuous random variables X and Y with joint PDF:

$$f_{X,Y}(x,y) = \begin{cases} \frac{6}{5}(x+y^2) & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{Otherwise} \end{cases}$$

1. Find $f_X(x)$:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \\ &= \frac{6}{5}\int_0^1 (x+y^2)dy \\ &= \frac{6}{5}\left[xy + \frac{y^3}{3}\right]_0^1 \\ &= \frac{6}{5}x + \frac{2}{5} \end{aligned}$$

2. Find $f_Y(y)$:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx \\ &= \frac{6}{5}\int_0^1 (x+y^2)dx \\ &= \frac{6}{5}\left[\frac{x^2}{2} + y^2x\right]_0^1 \\ &= \frac{6}{5}y^2 + \frac{3}{5} \end{aligned}$$

3. Compute probability $P\left(\frac{1}{4} \le Y \le \frac{3}{4}\right)$ from the marginal probabilities:

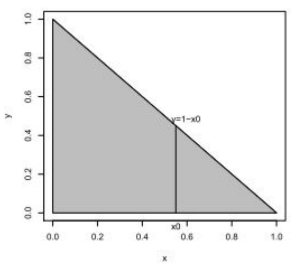
$$\begin{aligned} P\left(\frac{1}{4} \le Y \le \frac{3}{4}\right) &= \int_{\frac{1}{4}}^{\frac{3}{4}} f_Y(y)dy \\ &= \frac{37}{80} \\ &\approx 0.4265 \end{aligned}$$

If the region of integration **is not rectangular**, additional precautions need to be taken.

**Example 2**: Consider the continuous random variables X and Y with joint PDF given by:

$$f_{X,Y}(x,y) = \begin{cases} 24xy & 0 \leq x \leq 1,\ 0 \leq y \leq 1,\ x+y \leq 1 \\ 0 & \text{Otherwise} \end{cases}$$

Note the extra constraint $y \leq 1-x$, resulting in the region of integration being:




1. Establish that it is a **valid PDF** since the region is triangular:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = \int_0^1 \left(\int_0^{1-x} 24xydy\right) dx$$
$$\int_0^1 24x\left[\frac{y^2}{2}\right]_{y=0}^{y=1-x} dx$$
$$= \int_0^1 12x(1-x)^2 dx$$
$$= 1$$

2. Obtain marginal distribution of $X$:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \\ &= \int_0^{1-x} 24xydy \\ &= 12x(1-x)^2 \end{aligned}$$

## 3.3 Independence

The definition of independence for continuous random variables is similar to the definition for discrete random variables i.e. If $f_{XY}(x,y) = f_X(x)f_Y(y)$ for all $x$ and $y$, $X$ and $Y$ are **independent**.

For continuous random variables X and Y are said to be **independent**, if it satisfies the following:
1. $f_{XY}(x, y) = f_X(x)f_Y(y)$ for all x and y.
2. $f_{Y|x}(y) = f_Y(y)$ for all $x$ and $y$ with $f_X(x) > 0$
3. $f_{X|y}(x) = f_X(x)$ for all $x$ and $y$ with $f_Y(y) > 0$
4. $P(X \leq x \cap Y \leq y) = P(X \leq x)P(Y \leq y)$

Random variables are **independent** if the joint PMF or PDF can be expressed as a **product** of the marginal PMF or PDF.

**Example 1**: Given the following discrete random variables, determine if $X$ and $Y$ are dependent.

| | | X | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | $f_Y(y)$ |
| | 5 | 0.2 | 0.1 | 0.0 | 0.3 |
| Y | 6 | 0.2 | 0.1 | 0.1 | 0.4 |
| | 7 | 0.1 | 0.1 | 0.1 | 0.3 |
| | $f_X(x)$ | 0.5 | 0.3 | 0.2 | 1.0 |

1. Calculate the product of any marginal PDF:

$$f_X(1) \times f_Y(7) = 0.5 \times 0.3 = 0.15$$

2. Calculate the joint PDF:

$$f_{X,Y}(1, 7) = 0.1$$

3. Conclude: $X$ and $Y$ are dependent.

$$f_X(1) \times f_Y(7) \neq f_{X,Y}(1, 7)$$

## 3.4 More than 2 variables

The **joint PDF and CDF** are defined similarly to the previous concept:

- **Joint CDF**:

$$F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$$

- **Joint PDF**:

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \frac{\partial^n F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)}{\partial x_1, \partial x_2, \ldots, \partial x_n}$$

**Mutual independence**: Every subset of variables are independent i.e. the joint PMF can be expressed as a product of appropriate marginals (similar to previous chapter).

$$F_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\ldots F_{X_n}(x_n)$$
$$f_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\ldots f_{X_n}(x_n)$$

## 3.5 Conditional distributions

Analogous to discrete random variables, we can define the **conditional probability distribution** of $Y$ given $X = x$. This examines how one random variable behaves when a condition is placed on the other random variable.

For random variables (**discrete** or **continuous**) $X$ and $Y$ with:

- PMF (discrete) or PDF (continuous) denoted $f_{X,Y}(x, y)$
- Marginals denoted $f_X(x)$ and $f_Y(y)$

the **conditional PMF/PDF of** $Y$ **given** $X = x$:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} \text{ where } f_X(x) > 0$$

Since the function $f_{Y|x}(x)$ finds the probabilities of the all possible values for $Y$ given that $X = x$, $f_{Y|x}(x)$ is a PDF for continuous random variables and a PMF for discrete random variables. It has the following properties:

- $f_{Y|x}(y) \geq 0$
- $\int_{R_x} f_{Y|x}(y)\, dy = 1$
- $P(Y \in B|X = x) = \int_B f_{Y|x}(y)\, dy$

**Example 1**: Consider the continuous random variables $X$ and $Y$ with joint PDF:

$$f_{X,Y}(x,y) = \begin{cases} \frac{6}{5}(x+y^2) & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{Otherwise} \end{cases}$$

find the PDF of $X$ given $Y = 0.3$:

$$\begin{aligned} f_{X|Y}(x|0.3) &= \frac{f(x,0.3)}{f_Y(0.3)} \\ &= \frac{\frac{6}{5}(x+0.3^2)}{\frac{6}{5}(0.3)^2+\frac{3}{5}} \\ &= \frac{100}{59}(x+0.3) \text{ for } 0 \leq x \leq 1 \end{aligned}$$

# 4 Expectation and Variance

## 4.1 Expectation

The **mean**, **expected value**, or **expectation** $E(X)$ of a random variable $X$ is a weighted average of the possible values that $X$ can take, each value being weighted according to the probability of that event occurring i.e. the long-term average of the random variable.

Imagine observing many thousands of independent random values from the random variable of interest. Take the average of these random values. The expectation is the value of this average as the sample size tends to infinity.

The result for the expected value of a function of a random variable extends to joint distributions by weighting the PDF or PMF appropriately.

For random variables $X$ and $Y$, the expected value of $g(X, Y)$ is:

$$E[X] = \mu_x = \begin{cases} \sum_x \sum_y x f_{X,Y}(x, y) & \text{X and Y are discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X,Y}(x, y)\, dxdy & \text{X and Y are continuous} \end{cases}$$

and

$$E[Y] = \mu_y = \begin{cases} \sum_x \sum_y y f_{X,Y}(x, y) & \text{X and Y are discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{X,Y}(x, y)\, dxdy & \text{X and Y are continuous} \end{cases}$$

Properties of Expectation:

1. $E(X + Y) = E(X) + E(Y)$ for any random variable $X$ and $Y$.

2. $Var(X \pm Y) = Var(X) + Var(Y)$ if random variable $X$ and $Y$ are **uncorrelated**.

**Example 1**: Consider the data table, compute $E[g(X, Y)]$ where $g(X, Y) = X + Y$.

| | | X | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | $f_Y(y)$ |
| | 5 | 0.2 | 0.1 | 0.0 | 0.3 |
| Y | 6 | 0.2 | 0.1 | 0.1 | 0.4 |
| | 7 | 0.1 | 0.1 | 0.1 | 0.3 |
| | $f_X(x)$ | 0.5 | 0.3 | 0.2 | 1.0 |

$$\begin{aligned} E[g(X, Y)] &= \sum_x \sum_y g(x, y) f_{X,Y}(x, y) \\ &= 6(0.2) + 7(0.2) + 8(0.1) + \cdots + 9(0.1) + 10(0.1) \\ &= 7.7 \end{aligned}$$

### 4.1.1 Conditional expectation

**Conditional expectation** of $X$ with respect to $Y$ $E[X|Y]$: The conditional expectation $g(y)$ of $X$ given that $Y = y$ is given by:

$$g(y) = E[X|Y = y] = \begin{cases} \sum_i x_i P(X = x_i|Y = y) & \text{discrete} \\ \int_{-\infty}^{+\infty} x f_{X|Y}(x|y)\, dx & \text{continuous} \end{cases}$$

Note that $E[X|Y = y]$ depends on the value of $y$ i.e. by changing $y$, $E[X|Y = y]$ can also change. It is correct to say $E[X|Y = y]$ is a function of $y$:

$$g(y) = E[X|Y = y]$$

It is also possible to think of $g(y) = E[X|Y = y]$ as a function of the value of random variable $Y$:

$$g(Y) = E[X|Y]$$

Thus notation to indicate that $E[X|Y]$ is a random variable whose value equals $g(y) = E[X|Y = y]$ when $Y = y$. If $Y$ is a random variable with range $R_Y = y_1, y_2, \ldots$, then $E[X|Y]$ is **also** a random variable

**Example 1**: Given the condition PDF of $X$ given that $Y = 0.3$

$$f_{X|Y}(x|y) = \frac{100}{59}(x + 0.3)$$

find the **conditional expectation** of $X$ given that $Y = 0.3$.

$$\begin{aligned} E[X|Y = 0.3] &= \int_{-\infty}^{\infty} x f_{X|Y}(x|0.3)\, dx \\ &= \left[\frac{100}{77}x^3 + \frac{9}{118}x^2\right]_0^1 \\ &= \frac{227}{354} \end{aligned}$$

Common properties of conditional expectation:

- $E_Y[E[X|Y]] = E(X)$
- $E[ag(X) + h(X)|Y] = aE[g(X)|Y] + E[h(X)|Y]$
- $E[g(X)h(Y)|Y = y] = h(y)E[g(X)|Y = y]$

## 4.2 Conditional Variance

Similar to the conditional expectation, the conditional variance of $X$ is defined as $Var(X|Y = y)$, which is the variance of $X$ in the conditional space where $Y = y$.

For random variables $X$ and $Y$, the conditional variance $Var(X|Y)$ of $X$ with w.r.t $Y$ is the random variable whose value at point $Y = y$ is defined by:

$$Var[X|Y=y] = \begin{cases} \sum_i [x_i - E[X|Y=y]]^2 P(X = x_i | Y = y) & \text{for discrete} \\ \int_{-\infty}^{+\infty} [x - E[X|Y=y]]^2 f_{X|Y}(x|y) dx & \text{for continuous} \end{cases}$$

The **Law of Total Variance** states:

$$Var(X) = E[Var(X|Y)] + Var(E[X|Y])$$

# 5 Covariance and Correlation

In probability theory and statistics, the concepts of **covariance** and **correlation** are very similar and often used. Both the terms measure the relationship and the dependency between two variables between dependent random variables $X$ and $Y$.

When comparing data samples from different populations:

- **Covariance** measures the direction of the linear relationship between variables.
- **Correlation** measures both the strength and direction of the linear relationship between two variables.

Correlation is a function of the covariance. What sets them apart is the fact that correlation values are standardized whereas, covariance values are not.

## 5.1 Covariance

Covariance is a measure used to determine how much two variables change in tandem, the direction of the **linear relationship** between the two variables. It generalizes the concept of variance to multiple random variables. Instead of measuring the fluctuation of a single random variable, the covariance measures the fluctuation of two variables with each other.

Recall that the variance is the mean squared deviation from the mean for a single random variable $X$:

$$Var(X) = E[(X - E[X])^2].$$

**Covariance** between random variables $X$ and $Y$:

$$\text{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$$
$$\Rightarrow \text{Cov}(X,Y) = E[(X - \mu_x)(Y - \mu_y)]$$

thus

$$\text{Cov}(X,Y) = \begin{cases} \sum_X \sum_Y (x-\mu_x)(y-\mu_y) f_{X,Y}(x,y) & \text{discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_x)(y-\mu_y) f_{X,Y}(x,y)dx\,dy & \text{continuous} \end{cases}$$

where $\mu_x$ and $\mu_y$ are the expected value of $X$ and $Y$ respectively.

Common properties of covariance:

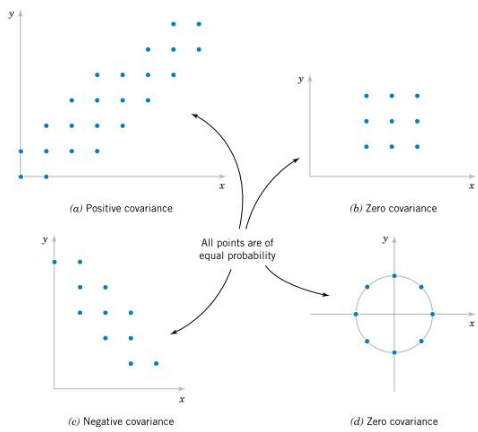
- $\text{Cov}(X,Y) = \text{Cov}(Y,X)$
- $\text{Cov}(X,X) = Var(X)$
- $\text{Cov}(X,a) = 0$ for any constant $a$
- $\text{Cov}(aX+b, cY+d) = ac\text{Cov}(X,Y)$

It is generally simpler to find the covariance by expanding the definition of covariance:

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

Interpreting the results:

- $\text{Cov}(X,Y) > 0$: Large values of $X$ tend to be associated with large values of $Y$ i.e. **higher the covariance, the stronger the relationship**.
- $\text{Cov}(X,Y) < 0$: Large values of $X$ tend to be association with small values of $Y$.
- $\text{Cov}(X,Y) = 0$: No simple linear relationship between the variables.



**Example 1:** Consider discrete random variables $X$ and $Y$ with joint PMF:

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2} & x=3, y=4 \\ \frac{1}{3} & x=3, y=6 \\ \frac{1}{6} & x=5, y=6 \\ 0 & \text{Otherwise} \end{cases}$$

find the covariance:

1. Find expected values $E(X)$ and $E(Y)$:

$$\begin{aligned} E(X) &= \frac{1}{2}(3) + \frac{1}{3}(3) + \frac{1}{6}(5) + 0 \\ &= \frac{10}{3} \end{aligned}$$

and

$$\begin{aligned} E(Y) &= \frac{1}{2}(4) + \frac{1}{3}(6) + \frac{1}{6}(6) + 0 \\ &= 5 \end{aligned}$$

2. Find the $\mathrm{Cov}(X,Y)$:

$$\begin{aligned} \mathrm{Cov}(X,Y) &= E[(X-\mu_x)(Y-\mu_y)] \\ &= \frac{\left(3-\frac{10}{3}\right)(4-5)}{2} + \frac{\left(3-\frac{10}{3}\right)(6-5)}{3} + \frac{\left(5-\frac{10}{3}\right)(6-5)}{6} + 0 \\ &= \frac{1}{6} - \frac{1}{9} + \frac{5}{18} + 0 \\ &= \frac{1}{3} \end{aligned}$$

**Example 2**: Recall the example with $X$ and $Y$ continuous, with joint PDF:

$$f_{X,Y}(x,y) = 24xy$$

for $0 \le x \le 1$, $0 \le y \le 1$, $x + y \le 1$, and 0 otherwise.

Find the covariance.

1. Find expect value $\mu_x$:

$$\begin{aligned} \mu_x &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)\, dy\, dx \\ &= \int_0^1 \int_0^{1-x} x24xy\, dy\, dx \\ &= \int_0^1 24x^2 \left[\frac{y^2}{2}\right]_0^{1-x} dx \\ &= \int_0^1 (12x^2 - 24x^3 + 12x^4) dx \\ &= \frac{2}{5} \end{aligned}$$

2. Find expect value $\mu_y$:

$$\begin{aligned}
\mu_y &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x,y)\,dy\,dx \\
&= \int_0^1\int_0^{1-x} y24xy\,dy\,dx \\
&= \int_0^1 24x\left[\frac{y^3}{3}\right]_0^{1-x} dx \\
&= \int_0^1 (8x - 24x^2 + 24x^3 - 8x^4)dx \\
&= \frac{2}{5}
\end{aligned}$$

3. Find the expect value $\mu_{xy}$:

$$\begin{aligned}
\mu_{xy} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\,dy\,dx \\
&= \int_0^1\int_0^{1-x} xy24xy\,dy\,dx \\
&= 8\int_0^1 x^2(1-x)^3 dx \\
&= \frac{2}{15}
\end{aligned}$$

4. Find covariance:

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= E(XY) - \mu_x\mu_y \\
&= \frac{2}{15} - \left(\frac{2}{5}\right)^2 \\
&= -\frac{2}{75}
\end{aligned}$$

If $X$ and $Y$ are **independent** then:

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

thus

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = 0$$

**Example 3**: Consider the functions $X = \cos(\theta)$ and $Y = \sin(\theta)$ which $\theta$ is uniformly distributed in $[0, 2\pi]$, verify if $X$ and $Y$ are independent or not:

1. Find $E(X)$:

$$E[X] = \frac{1}{2\pi}\int_0^{2\pi} \cos(\theta)\,d\theta = 0$$

2. Find $E(Y)$:

$$E[Y] = \frac{1}{2\pi}\int_0^{2\pi} \sin(\theta)\, d\theta = 0$$

3. Find $E(XY)$:

$$E[XY] = \int_0^{2\pi} \sin(\theta) cos(\theta) f(\theta)\, d\theta = \frac{1}{4\pi}\int_0^{2\pi} \sin(2\theta)\, d\theta = 0$$

4. Conclude:

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0$$

   Thus $X$ and $Y$ are independent.

## 5.2 Correlation

A deficiency of covariance is that it depends on the units of measurement. The units of covariance $\text{Cov}(X, Y)$ are "units of $X$ times units of $Y$". This makes it hard to compare covariances: if the scales changes then the covariance changes as well.

The correlation coefficient is found by dividing the covariance of the two variables by the product of their standard deviations.

- The values of the correlation coefficient can range from -1 to +1. The closer it is to +1 or -1, the more closely are the two variables are related.
- The positive sign signifies the direction of the correlation i.e. if one of the variables increases, the other variable is also supposed to increase.

**Correlation coefficient** (of variables $X$ and $Y$): Measures the **strength of the linear relationship** between the variables $X$ and $Y$.

$$\rho = \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\text{Cov}(X, Y)}{\sigma_x \sigma_y}$$

### 5.2.1 Lemma-Cauchy-Schwartz's inequality

The Cauchy-Schwartz inequality previously encountered in linear algebra is also valid for random variables. The Cauchy-Schwartz inequality is useful for bounding expected values that are difficult to calculate.

The concept allows the splitting of $E[X1, X2]$ into an upper bound with two parts, one for each random variable:

$$E|XY| \leq \sqrt{E(X^2)E(Y^2)}$$

This inequality shows that for two random variables, $X$ and $Y$, the expected value of the square of $X$ and $Y$ multiplied together $E(XY)^2$ will **always be less than or equal** to the expected value of the product of the squares of each $E(X^2)E(Y^2)$.

**Cauchy-Schwartz Inequality**: For any two random variables $X$ and $Y$,

$$E|XY| \leq \sqrt{E(X^2)E(Y^2)}$$

where equality holds if and only if $X = \alpha Y$, for some constant $\alpha \in \Re$.

### 5.2.2 Properties of the correlation coefficient

Interpreting the correlation coefficient:

- $-1 \leq \rho \leq 1$

  The proof highlights that the correlation efficient characterises the strength of the linear relationship between the variables X and Y.

- $\rho = \pm 1$: Equality occurs only when $Y = aX + b$ for $\rho = 0$ - **a perfectly linear relationship**. In this case, $\rho = 1$ when $a > 0$, and $\rho = -1$ when $a < 0$.

- $\rho > 0$: $X$ and $Y$ evolve in the same direction.

- $\rho < 0$: $X$ and $Y$ evolve in the opposite direction.

- $\rho = 0$: $X$ and $Y$ are independent i.e. the random variables are uncorrelated (no linear relationship but a non-linear relationship may be present).

**Example 1**: Consider discrete random variables $X$ and $Y$, with joint PMF given by:

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2} & x = 3, y = 4 \\ \frac{1}{3} & x = 3, y = 6 \\ \frac{1}{6} & x = 5, y = 6 \\ 0 & \text{Otherwise} \end{cases}$$

The covariance is $\frac{1}{3}$. Find the correlation.

1. Find $Var(X)$:

$$\begin{aligned} Var(X) &= E(X^2) - E(X)^2 \\ &= \frac{1}{2}(9) + \frac{1}{3}(9) + \frac{1}{6}(25) - \left(\frac{10}{3}\right)^2 \\ &= \frac{5}{9} \end{aligned}$$

2. Find $Var(Y)$:

$$\begin{aligned}
Var(Y) &= E(Y^2) - E(Y)^2 \\
&= \frac{1}{2}(16) + \frac{1}{3}(36) + \frac{1}{6}(36) - 5^2 \\
&= 1
\end{aligned}$$

3. Find the correlation:

$$\begin{aligned}
\rho &= \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \\
&= \frac{\frac{1}{3}}{\sqrt{\frac{5}{9}}} \\
&\approx 0.447
\end{aligned}$$

## 5.3 Covariance and Correlation matrices

By stacking up $X$ and $Y$ into a vector, it results in **covariance matrices**.

$$\begin{aligned}
\mathbf{R} &= E\left[\begin{bmatrix} X - E(X) \\ Y - E(X) \end{bmatrix} \begin{bmatrix} X - E(X) & Y - E(Y) \end{bmatrix}\right] \\
&= \begin{bmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X,Y) & \mathrm{Var}(Y) \end{bmatrix}
\end{aligned}$$

The **Correlation Matrix** is defined as:

$$\mathbf{C} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

## 5.4 Joint Normal (Gaussian) Distribution

Commonly known as Multivariate normal distribution. It is a generalization of the one-dimensional normal distribution to higher dimensions. The multivariate normal distribution is often used to describe, at least approximately, any set of possibly correlated real-valued random variables each of which clusters around a mean value.

The probability density function of a vector $[XY]$:

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \times \exp^{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{2\rho(x-\mu_X)(u-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)}$$

where:

- $-\infty < x < +\infty$
- $-\infty < y < +\infty$

- $|\rho| < 1$

The marginals alone do not tell us everything about the joint PDF, except when $X$, $Y$ are independent.

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}}\ N(\mu_X, \sigma_X^2)$$
$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy = \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-\frac{(x-\mu_Y)^2}{2\sigma_Y^2}}\ N(\mu_Y, \sigma_Y^2)$$

# 6 Moments

**Moments** of a function are quantitative measures related to the shape of the function's graph i.e. a set of statistical parameters to measure a distribution.

For $r = 1, 2, \ldots.$ the moments of random variables are defined:

$$m_r = E[X^r]$$

Four moments are commonly used:

1. $1^{st}$: **Mean** - The average
2. $2^{nd}$: **Variance**

   $$\mathrm{Var}(X) = E[X^2] - E[X]^2 = m_2 - m_1^2$$

   Standard deviation is the square root of the variance: an indication of how closely the values are spread about the mean.
3. $3^{rd}$: **Skewness** - A measure of the asymmetry of a distribution about its peak; it is a number that describes the shape of the distribution.
4. $4^{th}$: **Kurtosis** - A measure of the peakedness or flatness of a distribution.

The full sequence of moments can be obtained from a function called **moment generating function** (MGF):

$$m_X(t) = E(e^{tX}) = \begin{cases} \sum_x e^{tx} f_X(x) & X \text{ discrete} \\ \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx & X \text{ continuous} \end{cases}$$

The MGF has several uses and benefits:

- Provides full sequence of moments.
- Uniquely identifies the distribution function of random variables.
- Provides useful results for sums of random variables.

The moment generating function will exist only if the sum or integral in the above definition converges. If the moment generating function of a random variable does exist, it can be used to obtain all the origin moments of the random variable.

Let $X$ be a random variable with moment generating function $M_x(t)$:

$$m_r = \frac{d^r}{dt^r} m_X(t) = \begin{cases} \sum_x x^r e^{tx} f(x) & X \text{ discrete} \\ \int_{-\infty}^{\infty} x^r e^{tx} f(x)\, dx & X \text{ continuous} \end{cases}$$

**Example 1**: Consider the random variable $X \sim Poisson(\lambda)$, with PMF

$$f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad x = 0, 1, 2, \ldots$$

Find the MGF:

$$\begin{aligned} m_X(t) &= E(e^{tX}) \\ &= \sum_x e^{tx} f_X(x) \\ &= \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^x}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} \\ &= e^{-\lambda} e^{\lambda e^t} \\ &= \exp[\lambda(e^t - 1)] \end{aligned}$$

Compute the first and second moment:

$$m'_X(t) = E(X) = \lambda e^t \exp[\lambda(e^t - 1)] = \lambda$$
$$m''_X(t) = E(X^2) = \lambda^2 e^{2t} \exp[\lambda(e^t - 1)] + \lambda e^t \exp[\lambda(e^t - 1)] = \lambda$$

Moment generating functions have many important and useful properties. One of the most important of these is the **uniqueness property**. That is, the moment generating function of a random variable is unique when it exists, so if there are two random variables $X$ and $Y$, say, with moment generating functions for all values of t, both $X$ and $Y$ have the same probability distribution.

$$m_{X+Y}(t) = E(e^{t(X+Y)}) = E(e^{tX} e^{tY}) = E(e^{tX})E(e^{tY}) = m_X(t) m_Y(t)$$

Applying this formula when $X$ and $Y$ are both normally distributed yields the MGF of another normal distribution. This result i.e. the sum of independent normal random variables is itself a normal random variable, is very important in statistical analysis.

### 6.1 Relationship of MGF with Fourier Transform

Given $t = jw$, the MGF of a random variable $X$ is called **the characteristic function of** $X$.

$$\phi X(\omega) = m_X(j\omega) = E(e^{j\omega X}) = \int_{-\infty}^{\infty} f_X(x)e^{j\omega x}dx$$

As covered previously. the characteristic function is **nothing more than a Fourier transform with** $-\omega$ **instead of** $\omega$. Hence, the PDF could be obtained from the characteristic function

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_X(\omega)e^{-j\omega x}d\omega$$

If there are two independent random variables $X$ and $Y$, $m_{X+Y}(t) = m_X(t)m_Y(t)$

$$\phi_{X+Y}(\omega) = \phi_X(\omega)\phi_Y(\omega)$$

## 7 Sum of random variables

In statistical applications, the sums of random variables are often an aspect interested.

Given the random variables $X_1$, $X_2$,..., $X_n$ and using the properties of sums (integrals):

$$E(X_1 + X_2 + \ldots, X_n) = E(X_1) + E(X_2) + \ldots, E(X_n) = \sum_{i=1}^{n} E(X_i)$$

For **mutually independent random variables** $X_1$, $X_2$,..., $X_n$:

$$Var(X_1 + X_2 + \cdots + X_n) = Var(X_1) + Var(X_2) + \cdots + Var(X_n) = \sum_{i=1}^{n} Var(X_i)$$

Summary: For two independent random variables $X$ and $Y$:
- $f_{X,Y}(x, y) = f_X(x)f_Y(y)$
- $E(XY) = E(X)E(Y)$
- $Cov(X, Y) = 0$
- $\rho = 0$
- $Var(X + Y) = Var(X) + Var(Y)$
- $Var(X - Y) = Var(X) + Var(Y)$
- $mX + Y(t) = m_X(t)m_Y(t)$
- $E[X|Y] = E(X)$

*Refer to the EE lecture notes for examples for integration testing*