Biostatistics

林 建 甫

C.F. Jeff Lin, MD. PhD.

台北大學統計系助理教授
台北榮民總醫院生物統計顧問
美國密西根大學生物統計博士

Jointly Distributed Random Variables and Joint Probability Distributions

Discrete Jointly Distributed Random Variables

Introduction

- If X and Y are two random variables, the probability distribution that defines their simultaneous behavior is a Joint Probability Distribution.

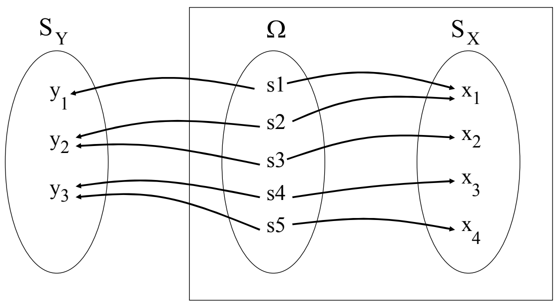
Examples:

  - Signal transmission: X is high quality signals and Y low quality signals.
  - Molding: X is the length of one dimension of molded part, Y is the length of another dimension.

- Thus, we may be interested in expressing probabilities expressed in terms of X and Y, e.g.,

  $P(2.95 < X < 3.05 \text{ and } 7.60 < Y < 7.8)$

Introduction to Multiple RVs



Two Discrete Random Variables

- **Range of random variables (X,Y)** is the set of points (x,y) in 2D space for which the probability that X = x and Y = y is positive.
- If X and Y are discrete random variables, the joint probability distribution of X and Y is a description of the set of points (x,y) in the range of (X,Y) along with the probability of each point.
- Sometimes referred to as **Bivariate probability distribution**, or **Bivariate distribution.**

Joint Probability Mass Function

Let X and Y be two discrete rv's defined on the sample space of an experiment. The *joint probability mass function* $p(x, y)$ is defined for each pair of numbers (x, y) by

$$p(x, y) = P(X = x \text{ and } Y = y)$$

Let A be the set consisting of pairs of (x, y) values, then

$$P\left[\left(X, Y\right) \in A\right] = \sum_{(x,y) \in A}\sum p(x, y)$$

Joint Probability Mass Function

- The joint probability mass function of the discrete random variables X and Y, denoted as $f_{XY}(x,y)$ satisfies:

$$(1)\ f_{XY}(x, y) \geq 0$$

$$(2) \sum_x \sum_y f_{XY}(x, y) = 1$$

$$(3)\ f_{XY}(x, y) = P(X = x, Y = y)$$

Marginal Probability Distributions

- Individual probability distribution of a random variable is referred to as its **Marginal Probability Distribution.**
- Marginal probability distribution of X can be determined from the joint probability distribution of X and other random variables.
- **Marginal probability distribution of X is found by summing the probabilities in each column**, **for Y, summation is done in each row.**

Marginal Probability Mass Functions

The *marginal probability mass functions* of X and Y, denoted $p_X(x)$ and $p_Y(y)$ are given by

$$p_X(x) = \sum_y p(x, y) \qquad p_Y(y) = \sum_x p(x, y)$$

Marginal Probability Distributions

- If X and Y are discrete random variables with joint probability mass function $f_{XY}(x,y)$, then the marginal probability mass function of X and Y are

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

$$f_Y(y) = P(Y = y) = \sum_{Ry} f_{XY}(X, Y)$$

where R_x denotes the set of all points in the range of (X, Y) for which $X = x$ and Ry denotes the set of all points in the range of (X, Y) for which $Y = y$

Multiple Discrete RVs

- Example: Consider 3 coin flips

X = Number of Heads

$$Y = \begin{cases} 1 & \text{first and third outcomes are the same} \\ 0 & \text{otherwise} \end{cases}$$

$$X(HHH) = 3, Y(HHH) = 1$$
$$X(HHT) = 2, Y(HHT) = 0$$
$$X(TTT) = 0, Y(TTT) = 1$$

- Rvs X and Y are defined on the same underlying experiment !!

Multiple Discrete RVs

- $S_{X,Y} = \{(x,y), x \in \{0,1,2,3\}, y \in \{0,1\}\}$

  $P_{X,Y}(2,0) = \mathbf{P}[\{HHT, THH\}]$

- Joint PMF of X and Y

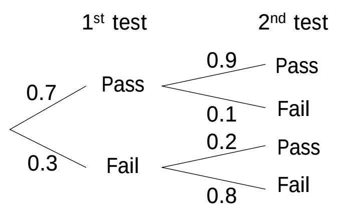
$$P_{X,Y}(x,y) = \begin{cases} 1/8 & (x,y) = (0,1),(1,1),(2,1), \text{ or } (3,1) \\ 1/4 & (x,y) = (1,0), \text{ or } (2,0) \\ 0 & \text{otherwise} \end{cases}$$

X 與 *Y* 的 聯合機率分配 與 邊際機率分配表

$X \backslash Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	y_m	$f_x(x_i)$
x_1	$f(x_1,y_1)$	$f(x_1,y_2)$	$\cdots$	$f(x_1,y_j)$	$\cdots$	$f(x_1,y_m)$	$f_x(x_1)$
x_2	$f(x_2,y_1)$	$f(x_2,y_2)$	$\cdots$	$f(x_2,y_j)$	$\cdots$	$f(x_2,y_m)$	$f_x(x_2)$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
x_i	$f(x_i,y_1)$	$f(x_i,y_2)$	$\cdots$	$f(x_i,y_j)$	$\cdots$	$f(x_i,y_m)$	$f_x(x_i)$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
x_n	$f(x_n,y_1)$	$f(x_n,y_2)$	$\cdots$	$f(x_n,y_j)$	$\cdots$	$f(x_n,y_m)$	$f_x(x_n)$
$f_y(y_j)$	$f_y(y_1)$	$f_y(y_2)$	$\cdots$	$f_y(y_j)$	$\cdots$	$f_y(y_m)$	1

Multiple Discrete RVs Representation of Joint PMF

- Example: Joe gets pulled over by a cop and is administered two tests. Suppose that Joe has to pay $200 for each test he fails. Let X be the amount he has to pay for the 1st test and Y the amount he pays for the 2nd test.



Multiple Discrete RVs Representation of Joint PMF

1. List

$$P_{X,Y}(x,y) = \begin{cases} 0.63 & x=0, y=0 \\ 0.07 & x=0, y=200 \\ 0.06 & x=200, y=0 \\ 0.24 & x=200, y=200 \\ 0 & \text{otherwise} \end{cases}$$

2. Matrix

$P_{X,Y}(x,y)$	$y=0$	$y=200$
$x=0$	0.63	0.07
$x=200$	0.06	0.24

Multiple Discrete RVs

Ex

A fair coin is thrown two times and a fair die is thrown one time simultaneously. Find the probability of getting not more than one head on the coin and two spots shown on the die by letting two-demensional random variables.

Ans:

Let outcomes be *e*, and the random variables be (*X*, *Y*), where *X* be the numbers of heads showing on the coin, and *Y* be the numbers of spots showing on the die,

$$X \in \{0,1,2\} \quad Y \in \{1,2,\cdots,6\}$$

For example:

$$e = (HH,3) \Rightarrow X(e) = 2, Y(e) = 3$$

$P(X, Y)$	Y=1	Y=2	Y=3	Y=4	Y=5	Y=6
X=0	1/24	1/24	1/24	1/24	1/24	1/24
X=1	1/12	1/12	1/12	1/12	1/12	1/12
X=2	1/24	1/24	1/24	1/24	1/24	1/24

The cumulative distribution function:

$$\begin{aligned} F_{XY}(1,2) &= P(X(e) \le 1, Y(e) \le 2) \\ &= P(\{(TT,1),(TH,1),(HT,1),(TT,2),(TH,2),(HT,2)\}) \\ &= 6/24 \\ &= 1/4 \end{aligned}$$

Multiple Discrete RVs

Three balls are drawn from a box containing 4 red balls, 3 yellow balls, 5 blue balls. If we let *X*, *Y* denote the numbers of red and yellow balls chosen respectively, find the joint probability mass function of *X* and *Y*.

Ans:

Let joint probability mass function be $p_{XY}(x, y)$

$p_{XY}(0,0) = {}_5C_3 / {}_{12}C_3 = 10/220$

$p_{XY}(0,1) = ({}_3C_1)({}_5C_2) / {}_{12}C_3 = 30/220$

$p_{XY}(0,2) = ({}_3C_2)({}_5C_1) / {}_{12}C_3 = 15/220$

$p_{XY}(0,3) = {}_3C_3 / {}_{12}C_3 = 1/220$

$p_{XY}(1,0) = ({}_4C_1)({}_5C_2) / {}_{12}C_3 = 40/220$

$p_{XY}(1,1) = ({}_4C_1)({}_3C_1)({}_5C_1) / {}_{12}C_3 = 60/220$

$p_{XY}(1,2) = ({}_4C_1)({}_3C_2) / {}_{12}C_3 = 12/220$

$p_{XY}(2,0) = ({}_4C_2)({}_5C_1) / {}_{12}C_3 = 30/220$

$p_{XY}(2,1) = ({}_4C_2)({}_3C_1) / {}_{12}C_3 = 18/220$

$p_{XY}(3,0) = ({}_4C_3)/({}_{12}C_3) = 4/220$

Otherwise : $p_{XY}(x, y) = 0$

A supermarket has two express lines. Let *X* and *Y* denote the numbers of customers in the first and in the second respectively at any given time. Find $P(|x-y|=1)$ by using the joint probability mass function below.

$p_{XY}(i, j)$	i=0	i= 1	i= 2	i= 3
i=0	0.1	0.2	0	0
i=1	0.2	0.25	0.05	0
i=2	0	0.05	0.05	0.025
i=3	0	0	0.025	0.05

Multiple Discrete RVs

- A supermarket has two express lines. Let *X* and *Y* denote the numbers of customers in the first and in the second respectively at any given time. Find $P(|x-y|=1)$ by using the joint probability mass function below.

$p_{XY}(i, j)$	i=0	i= 1	i= 2	i= 3
i=0	0.1	0.2	0	0
i=1	0.2	0.25	0.05	0
i=2	0	0.05	0.05	0.025
i=3	0	0	0.025	0.05

A supermarket has two express lines. Let *X* and *Y* denote the numbers of customers in the first and in the second respectively at any given time. Find $P(|x-y|=1)$ by using the joint probability mass function below.

$p_{XY}(i, j)$	i=0	i= 1	i= 2	i= 3
i=0	0.1	0.2	0	0
i=1	0.2	0.25	0.05	0
i=2	0	0.05	0.05	0.025
i=3	0	0	0.025	0.05

Ans:

$$P(|x-y|=1) = \sum_{|i-j|=1} p_{XY}(i, j)$$
$$= p_{XY}(0,1) + p_{XY}(1,0) + p_{XY}(1,2) + p_{XY}(2,1) + p_{XY}(3,2) + p_{XY}(2,3)$$
$$= 0.2 + 0.2 + 0.05 + 0.05 + 0.025 + 0.025$$
$$= 0.55$$

Therefore, the probability is 0.55

Ans:

$$P(|x-y|=1)$$
$$= \sum_{|i-j|=1} p_{XY}(i, j)$$
$$= p_{XY}(0,\ 1) + p_{XY}(1,\ 0) + p_{XY}(1,\ 2) + p_{XY}(2,\ 1) + p_{XY}(3,\ 2) + p_{XY}(2,\ 3)$$
$$= 0.2 + 0.2 + 0.05 + 0.05 + 0.025 + 0.025$$
$$= 0.55$$

Therefore, the probability is 0.55

Continuous Jointly Distributed Random Variables

Two Continuous Random Variables

- Analogous to the probability density function of a single continuous random variable, a Joint probability density function can be defined over two-dimensional space.

Joint Probability Density Function

Let X and Y be continuous rv's. Then $f(x, y)$ is a *joint probability density function* for X and Y if for any two-dimensional set A

$$P\left[\left(X,Y\right)\in A\right]=\iint_A f(x,y)dxdy$$

If A is the two-dimensional rectangle $\left\{(x,y): a \le x \le b, c \le y \le d\right\}$,

$$P\left[\left(X,Y\right)\in A\right]=\int_a^b\int_c^d f(x,y)dydx$$

Marginal Probability Density Functions

The *marginal probability density functions* of X and Y, denoted $f_X(x)$ and $f_Y(y)$, are given by

$$f_X(x)=\int_{-\infty}^{\infty} f(x,y)dy \quad \text{for } -\infty < x < \infty$$

$$f_Y(y)=\int_{-\infty}^{\infty} f(x,y)dx \quad \text{for } -\infty < y < \infty$$

Marginal Probability Distribution

- If the joint probability density function of continuous random variables X and Y is $f_{XY}(x,y)$, the marginal probability density function of X and Y are

$$f_X(x)=\int_{R_x} f_{XY}(x,y)dy \quad \text{and} \quad f_Y(y)=\int_{R_{xy}} f_{XY}(x,y)dx$$

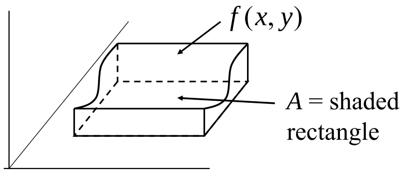
where R_x denotes the set of all points in the range of (X,Y) for which X = x and R_y denotes the set of all points in the range of (X,Y) for which Y = y.

Marginal Probability Distribution

- A probability involving only one random variable, e.g., $P(a < X < b)$, can be found from the marginal probability of X or from the joint probability distribution of X and Y.
- **For example:**

$P(a < x < b) = P(a < x < b, -\infty < Y < \infty)=$

$$\int_a^b\int_{R_x} f_{XY}(x,y)dxdy=\int_a^b\left(\int_{R_x} f_{XY}(x,y)dy\right)dx=\int_a^b f_{XY}(x)dx$$

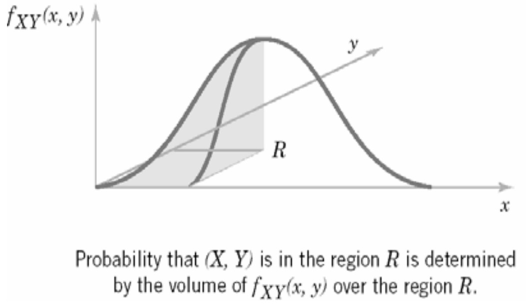


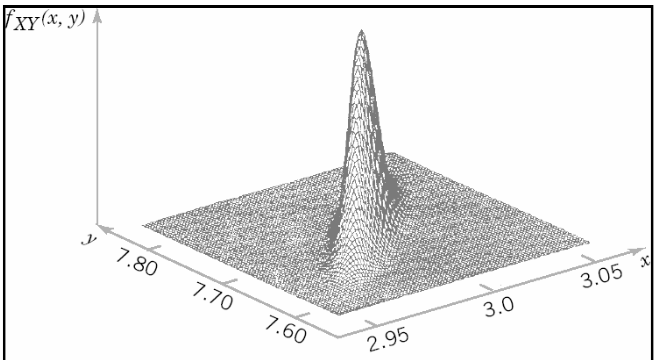
$$P\left[\left(X,Y\right)\in A\right]$$

= Volume under density surface above A

Joint Probability Distribution

- The probability that (X,Y) assumes a value in the region R equals the volume of the shaded region.





Joint probability density function for the lengths of different dimensions of an injection-molded part;

P(2.95 < X < 3.05,7.60 < Y < 7.80)

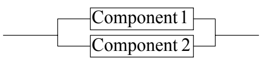
Two Continuous Random Variables

Consider an electronic system containing two components, one for backup. Suppose the two components have identical performance characteristics. Let X and Y be random variables denoting their life spans, with :

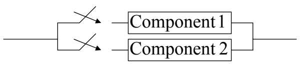
$$f_{XY}(x,y)=\begin{cases}\lambda^2 e^{-\lambda(x+y)} & \text{if } x\ge 0,\ y\ge 0,\ \lambda>0\\ 0 & \text{otherwise}\end{cases}$$

Later, the system itself is modified so that one component is kept on reserve and activated only when the other needs replacing. Find the probability that the system fails to last far more than 1000 hours before and after modification. And determine is the system become more reliable after modification.

Before modification:



After modification:



Ans:

Step 1 —To represent the failure event in term of *X* and *Y*

Step 2 —To find *R* in the two-dimensional range space R_{XY}

Step 3 —To calculate the probability by $P(\text{failure})=\iint_R f_{XY}(x,y)\,dx\,dy$

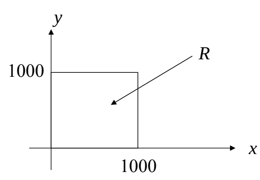
The system fails only if both component fail.

Before modification:

Let *E* be the event that the system fails before modification.

$$E=\{\text{"}X\le 1000\text{"}\cap\text{"}Y\le 1000\text{"}\}$$

$$\begin{aligned}P(E)&=P(X\le 1000, Y\le 1000)\\&=\iint_R f(x,y)\,dx\,dy\\&=\int_0^{1000}\int_0^{1000}\lambda^2 e^{-\lambda(x+y)}\,dx\,dy\\&=\int_0^{1000}\lambda e^{-\lambda x}\,dx\cdot\int_0^{1000}\lambda e^{-\lambda y}\,dy\\&=(1-e^{-1000\lambda})^2\end{aligned}$$



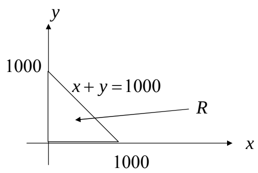
Ans:

After modification:

Let F be the event that the system fails after modification.

$F = \{"X + Y \le 1000"\}$

$$
\begin{aligned}
P(F) &= P(X + Y \le 1000) \\
&= \iint_R f(x, y)\, dx\, dy \\
&= \int_0^{1000} \left[\int_0^{1000-y} \lambda^2 e^{-\lambda(x+y)}\, dx \right] dy \\
&= \int_0^{1000} \lambda e^{-\lambda y} \left[\int_0^{1000-y} \lambda e^{-\lambda x}\, dx \right] dy \\
&= \int_0^{1000} \lambda e^{-\lambda y} \left[-e^{-\lambda x} \right]_0^{1000-y} dy \\
&= \int_0^{1000} \lambda e^{-\lambda y} \left[1 - e^{-\lambda(1000-y)} \right] dy \\
&= 1 - e^{-1000\lambda} - 1000\lambda e^{-1000\lambda}
\end{aligned}
$$



Therefore, the system is more reliable after modification.

Jointly Distributed

More Than Two Random Variables

More Than Two Random Variables

If $X_1, X_2, \ldots, X_n$ are all discrete random variables, the joint pmf of the variables is the function

$$p(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$$

If the variables are continuous, the joint pdf is the function f such that for any n intervals $[a_1, b_1], \ldots, [a_n, b_n]$, $P(a_1 \le X_1 \le b_1, \ldots, a_n \le X_n \le b_n)$

$$= \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1, \ldots, x_n)\, dx_n \ldots dx_1$$

Multiple Discrete Random Variables
Multinomial Probability Distribution

Suppose a random experiment consists of a series of n trials. Assume that

(1) The result of each trial is classified into one of k classes.

(2) The probability of a trial generating a result in class 1, class 2, …, class k is constant over the trials and equal to $p_1, p_2, \ldots, p_k$, respectively.

(3) The trials are independent.

The random variables $X_1, X_2, \ldots, X_k$ that denote the number of trials that result in class 1, class 2, …, class k, respectively, have a **multinomial distribution** and the joint probability mass function is

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{n!}{x_1! x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k} \qquad (5\text{-}13)$$

for $x_1 + x_2 + \cdots + x_k = n$ and $p_1 + p_2 + \cdots + p_k = 1$.

Multiple Discrete Random Variables
Multinomial Probability Distribution

Each trial in a multinomial random experiment can be regarded as either generating or not generating a result in class i, for each $i = 1, 2, \ldots, k$. Because the random variable X_i is the number of trials that result in class i, X_i has a binomial distribution.

If $X_1, X_2, \ldots, X_k$ have a multinomial distribution, the marginal probability distribution of X_i is binomial with

$$E(X_i) = np_i \quad \text{and} \quad V(X_i) = np_i(1 - p_i) \qquad (5\text{-}14)$$

Multiple Continuous Random Variables

Definition

If the joint probability density function of continuous random variables $X_1, X_2, \ldots, X_p$ is $f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p)$ the **marginal probability density function of X_i** is

$$f_{X_i}(x_i) = \iint_{R_{x_i}} \cdots \int f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_1\, dx_2 \ldots dx_{i-1}\, dx_{i+1} \ldots dx_p \qquad (5\text{-}23)$$

where R_{x_i} denotes the set of all points in the range of $X_1, X_2, \ldots, X_p$ for which $X_i = x_i$.

Independent Random Variables

Independent Random Variables

Two random variables X and Y are said to be *independent* if for every pair of x and y values

$$p(x, y) = p_X(x) \cdot p_Y(y)$$

when X and Y are discrete or

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

when X and Y are continuous. If the conditions are not satisfied for all (x, y) then X and Y are *dependent*.

Independence – More Than Two Random Variables

The random variables $X_1, X_2, \ldots, X_n$ are *independent* if for every subset $X_{i_1}, X_{i_2}, \ldots, X_{i_n}$ of the variables, the joint pmf or pdf of the subset is equal to the product of the marginal pmf's or pdf's.

Independence – More Than Two Random Variables

In the case that $x_1, x_2, x_3, \ldots x_n$ is a independent sample from $f(x)$ then the joint density of $x_1, x_2, x_3, \ldots x_n$ is:

$$f(x_1, x_2, x_3, \ldots x_n) = \prod_{i=1}^{n} f(x_i)$$

Example:

Suppose that $x_1, x_2, x_3, \ldots x_n$ is a sample from the Normal distribution with mean μ and standard deviation σ.

$$f(x_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}}$$

The joint density of $x_1, x_2, x_3, \ldots x_n$ is:

$$f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}}$$

$$= \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}}$$

Example:

Suppose that x_1, x_2, x_3, … x_n is a sample from the Exponential distribution with parameter λ.

$$f(x_i) = \begin{cases} \lambda e^{-\lambda x_i} & x_i \geq 0 \\ 0 & x_i < 0 \end{cases}$$

The joint density of x_1, x_2, x_3, … x_n is:

$$f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i)$$

$$= \begin{cases} \prod_{i=1}^{n} \lambda e^{-\lambda x_i} & x_1, x_2, \ldots, x_n \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i} & x_1, x_2, \ldots, x_n \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Conditional Probability Function

Conditional Probability Function

Let X and Y be two continuous rv's with joint pdf $f(x, y)$ and marginal X pdf $f_X(x)$. Then for any X value x for which $f_X(x) > 0$, *the conditional probability density function of Y given that X = x* is

$$f_{Y|X}(y \mid x) = \frac{f(x, y)}{f_X(x)} \quad -\infty < y < \infty$$

If X and Y are discrete, replacing pdf's by pmf's gives the *conditional probability mass function of Y when X = x.*

Conditional Distribution

Definition:

Let (X, Y) be a discrete r.v. with ***jpmf***, $p(x_i, y_j)$.The conditional **pmf's** are defined by

$$P(X = x_i Y = y_j) = P_{X|Y}(x_i \mid y_j) = p(x_i, y_j) / p_Y(y_j)$$

$$\ni \; p_Y(y_j) \; (\neq 0)$$

Conditional Probability

- Because a conditional probability mass function $f_{Y|x}(y)$ is a probability mass function for all y in R_x, the following properties are satisfied:

(1) $f_{Y|x}(y) \geq 0$

(2) $\sum_{R_x} f_{Y|x}(y) = 1$

(3) $P(Y=y|X=x) = f_{Y|x}(y)$

Conditional Probability & Independence

- For discrete random variables X and Y, if any one of the following properties is true, the others are also true, and X and Y are independent.
  (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) $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for **any** sets A and B in the range of X and Y respectively.
- **If we find one pair of x and y in which the equality fails, X and Y are not independent.**

大慶通訊最新型手機的銷售資料

X（有否購買大哥大手機）	Y（是否看過型錄）0（未看過）	1（看過）	合計
0（不買）	1,200	480	1,680
1（購買）	240	480	720
合　計	1,440	960	2,400

X 與 *Y* 的聯合機率分配表

X（是否購買大哥大手機）	Y（是否看過型錄）0（未看過）	1（看過）	合計 $f_x(x)$
0（不買）	0.50	0.20	0.7
1（購買）	0.10	0.20	0.3
合計 $f_y(y)$	0.60	0.40	1.0

F(x | y) 的條件機率

X \ Y	$Y = 0$（未看過）	$Y = 1$（看過）
$X = 0$（不買）	0.83	0.5
$X = 1$（購買）	0.17	0.5

F(y | x) 的條件機率

Y \ X	$X = 0$（不買）	$X = 1$（購買）
$Y = 0$（未看過）	0.71	0.33
$Y = 1$（看過）	0.29	0.67

Expected Values, Covariance, and Correlation

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 $E[h(X, Y)]$ or $\mu_{h(X,Y)}$

is

$$= \begin{cases} \sum_x \sum_y h(x,y) \cdot p(x,y) & \text{discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y) \cdot f(x,y)dxdy & \text{continuous} \end{cases}$$

Covariance and Correlation

- When two or more random variables are defined on a probability space, it is useful to **describe how they vary together**, i.e., measure the relationship between the variables.
- A common measure of the relationship between two random variables is the covariance**.**

Covariance

The *covariance* between two rv's X and Y is

$$\text{Cov}(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right]$$

$$= \begin{cases} \sum_x \sum_y (x-\mu_X)(y-\mu_Y)p(x,y) & \text{discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y)f(x,y)dxdy & \text{continuous} \end{cases}$$

Covariance

- The covariance between the random variables X and Y, denoted as cov(X,Y) or σ_{XY}, is

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

- **Covariance is a measure of the linear relationship between random variables.**

Short-cut Formula for Covariance

$$\text{Cov}(X,Y) = E(XY) - \mu_X \cdot \mu_Y$$

Covariance

- If the points in the joint probability distribution of X and Y that receive positive probability tend to fall along a line of positive (or negative) slope, σ_{XY} is positive (or negative).
- If the relationship between the random variables is nonlinear, the covariance might not be sensitive to the relationship.

Correlation

- The correlation between random variables X and Y, denoted as ρ_{XY}, is

$$\rho_{XY} = \frac{\operatorname{cov}(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

- The correlation scales the covariance by the standard deviation of each variable.
- It is dimensionless quantity that can be sued to **compare the linear relationship between pairs of variables in different units**.

Correlation

- Because $\sigma_X > 0$ and $\sigma_Y > 0$, if the covariance between X and Y is positive, negative, or zero, the correlation between X and Y is positive, negative, or zero respectively.
- If the points in the joint probability distribution of X and Y that receive positive probability tend to fall along a line of positive (or negative) slope, ρ_{XY} is near +1 (or -1).

Correlation

- For any two random variables X and Y

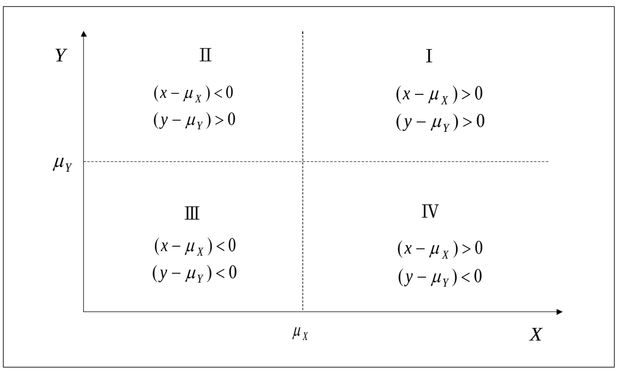
$$-1 \le \rho_{XY} \le +1$$

- If ρ_{XY} **equals +1 or -1**, the points in the joint probability distribution that receive positive probability **fall exactly along a straight line**.
- **Two random variables with nonzero correlation are said to be correlated.**

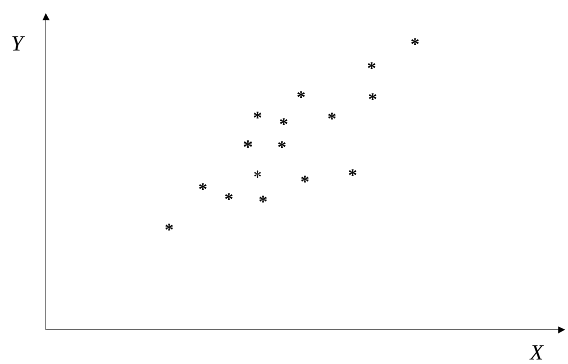
Correlation Proposition

1. If X and Y are independent, then $\rho = 0$, but $\rho = 0$ does not imply independence.
2. $\rho = 1$ or -1 iff $Y = aX + b$ for some numbers a and b with $a \neq 0$.
3. If a and c are either both positive or both negative, Corr($aX + b$, $cY + d$) = Corr(X, Y)

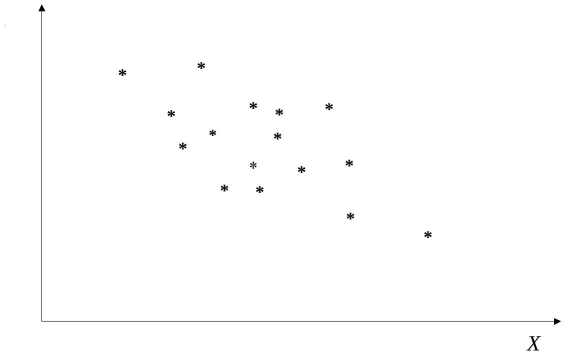
COV(*X*,*Y*)的符號



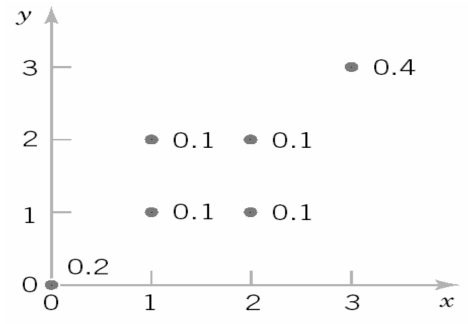
正向共變



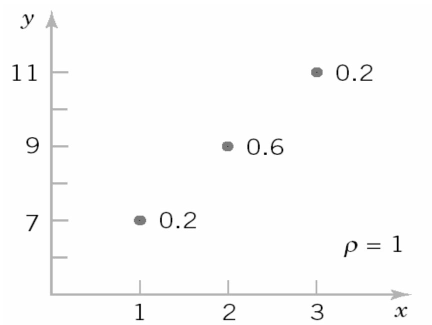
反向共變



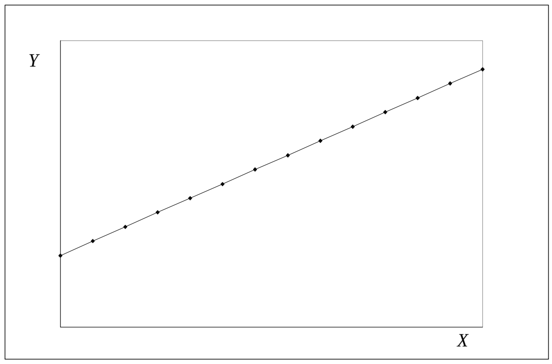
An example for a correlation that is close to unity



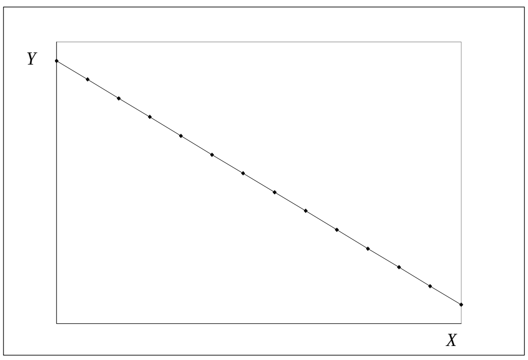
An example of unity correlation



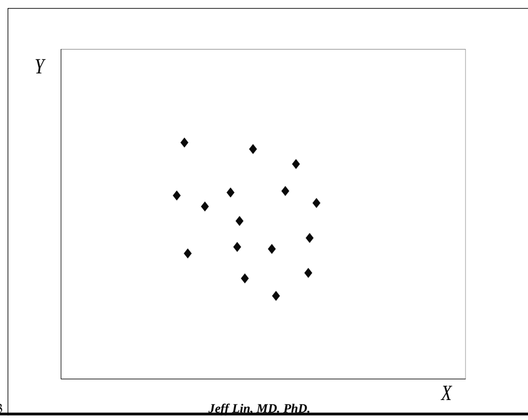
相關係數 ρ = +1



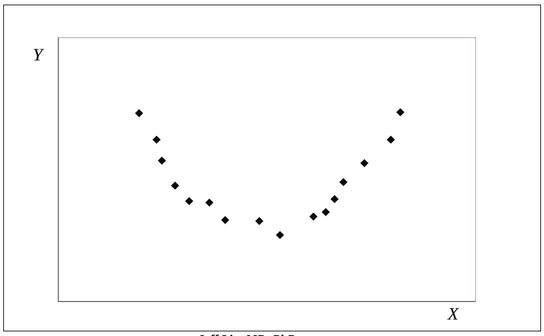
相關係數 ρ = -1



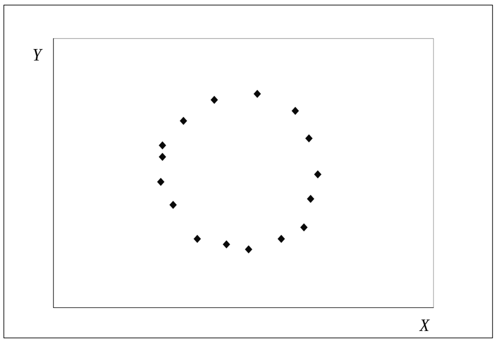
X, Y 無相關 ρ = 0



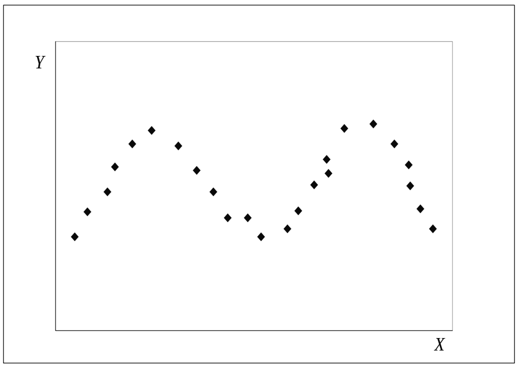
X, Y 爲抛物線關係



X, Y 爲圓形關係, ρ = 0



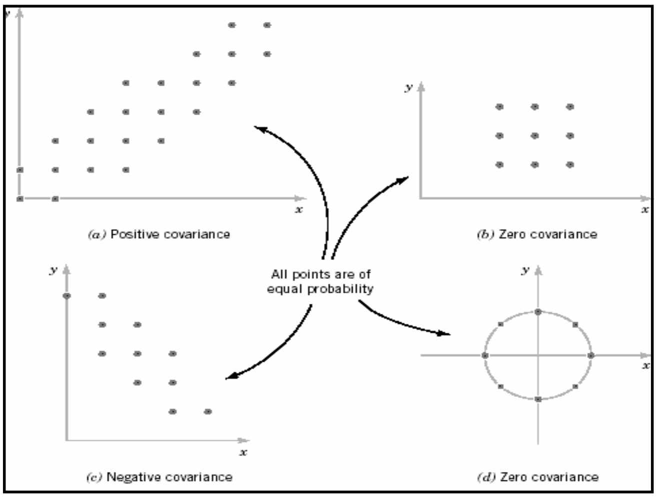
X, Y爲非線性關係



Correlation and Independence

- If X and Y are **independent random variables,**

  $\sigma_{XY} = \rho_{XY} = 0$



Conditional Expected Values, Covariance, and Correlation

Conditional Probability

- Let R_x denote the set of all points in the range of (X,Y) for which X=x. The conditional mean of Y given X = x, denoted as E(Y|x) or $\mu_{Y|x}$, is

$$E(Y \mid x) = \sum_{R_x} y f_{Y|x}(y)$$

- And the conditional variance of Y given X = x, denoted as V(Y|x) or $\sigma^2_{Y|x}$ is

$$V(Y \mid x) = \sum_{R_x} (y - \mu_{Y|x})^2 f_{Y|x}(y) = \sum_{R_x} y^2 f_{Y|x}(y) - \mu^2_{Y|x}$$

Conditional Expectation and Variance

- The conditional expectation of a r.v. X given that Y=y is defined in the discrete case by

$$\mathbf{E(X \mid Y = y_k) = \sum_{\forall x_i} x_i \, p_{X|Y}(x \mid y_k)}$$

- and in the continuous case by

$$\mathbf{E(X \mid Y = y) = \int_{-\infty}^{\infty} x \, f_{X|Y}(x \mid y) \, dx}$$

Introduction to Multiple RVs

- Example: Consider 3 coin flips

X = Number of Heads

$$Y = \begin{cases} 1 & \text{first and third outcomes are the same} \\ 0 & \text{otherwise} \end{cases}$$

$X(HHH) = 3, Y(HHH) = 1$

$X(HHT) = 2, Y(HHT) = 0$

$X(TTT) = 0, Y(TTT) = 1$

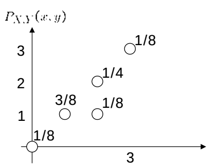
- Rvs X and Y are defined on the same underlying experiment !!

Conditional Expected Value

- Example: Toss a coin 3 times
  X = number of heads in 3 independent tosses
  Y = maximum number of consecutive heads

  Compute $E[X|Y=y]$, y = 0, 1, 2, and 3

$$E[X|Y=y] = \begin{cases} 0 & y=0 \\ 5/4 & y=1 \\ 2 & y=2 \\ 3 & y=3 \end{cases}$$



Conditional Expected Value

$P_{X,Y}(x,y)$	$y=0$	$y=1$	$y=2$	$P_X(x)$
$x=0$	0.01	0	0	0.01
$x=1$	0.09	0.09	0	0.18
$x=2$	0	0	0.81	0.81
$P_Y(y)$	0.10	0.09	0.81	

$$\begin{aligned} \gamma_{X,Y} &= E[XY] \\ &= \sum_{x=0}^{2}\sum_{y=0}^{2} xyP_{X,Y}(x,y) \\ &= (1)(1)0.09 + (2)(2)0.81 = 3.33 \\ \rho_{X,Y} &= \gamma_{X,Y} - E[X]\,E[Y] \\ &= 3.33 - 1.8 \cdot 1.71 = 0.252 \end{aligned}$$

Bivariate Normal Random Variable And Bivariate Normal Distribution

Bivariate Normal Distribution

Definition

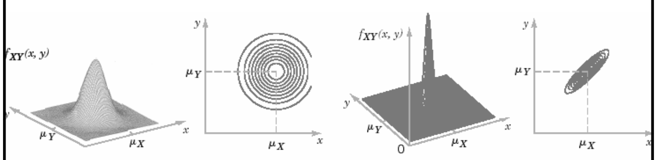
The probability density function of a **bivariate normal distribution** is

$$f_{XY}(x, y; \sigma_X, \sigma_Y, \mu_X, \mu_Y, \rho) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]\right\} \quad (5\text{-}32)$$

for $-\infty < x < \infty$ and $-\infty < y < \infty$, with parameters $\sigma_X > 0$, $\sigma_Y > 0$, $-\infty < \mu_X < \infty$, $-\infty < \mu_Y < \infty$, and $-1 < \rho < 1$.

5-6 Bivariate Normal Distribution

Examples of bivariate normal distributions.



Ex :

A bivariate normal random variable (X, Y) has the joint probability density function:

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

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

(1) Find $f_X(x)$ and $f_Y(y)$

(2) Show that ρ_{XY} is the correlation coefficient of (X, Y)

(3) Determine $E(Y \mid x)$

Ans:

(1) The probability density function of X is

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y)\, dy$$

The probability density function of Y is

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

Ans:

$$u = \frac{x-\mu_X}{\sigma_X}, \qquad v = \frac{y-\mu_Y}{\sigma_Y}$$

$$-\frac{u^2 - 2\rho_{XY}uv + v^2}{2(1-\rho_{XY}^2)} = -\frac{(v-\rho_{XY}u)^2 + (1-\rho_{XY}^2)u^2}{2(1-\rho_{XY}^2)} = -\frac{(v-\rho_{XY}u)^2}{2(1-\rho_{XY}^2)} - \frac{u^2}{2}$$

$$f_X(x) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_X} \cdot \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho_{XY}^2}} \exp\left[-\frac{\left(\frac{y-\mu_Y}{\sigma_Y} - \rho_{XY}\frac{x-\mu_X}{\sigma_X}\right)^2}{2\left(1-\rho_{XY}^2\right)} - \frac{(x-\mu_X)^2}{2\sigma_X^2}\right] dy$$

$$= \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left[-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right] \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho_{XY}^2}} \exp\left[\frac{\left(y - \left(\mu_Y + \rho_{XY}\cdot\frac{\sigma_Y}{\sigma_X}(x-\mu_X)\right)\right)^2}{2\sigma_Y^2(1-\rho_{XY}^2)}\right] dy$$

$$= \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left[-\frac{(x-\mu_X)^2}{2\sigma_x^2}\right]$$

Ans:

Similarly,

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

Therefore, $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$

(2) The correlation coefficient is:

$$\frac{\text{Cov}(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{E[(X-\mu_X)(Y-\mu_y)]}{\sigma_X\sigma_Y}$$

$$= E\left[\left(\frac{X-\mu_X}{\sigma_X}\right)\left(\frac{Y-\mu_Y}{\sigma_Y}\right)\right]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_y}{\sigma_Y}\right) f_{XY}(x,y)\,dx\,dy$$

Ans:

Let $u = \frac{x-\mu_X}{\sigma_X}$, $v = \frac{y-\mu_Y}{\sigma_Y}$,

$\therefore$ $dx = \sigma_X du$, $dy = \sigma_Y dv$

$$\frac{\text{Cov}(X,Y)}{\sqrt{V(X)V(Y)}} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u\cdot v \cdot \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}} \exp\left[-\frac{u^2 - 2\rho_{XY}uv + v^2}{2(1-\rho_{XY}^2)}\right]\sigma_X\sigma_Y\,du\,dv$$

$$= \int_{-\infty}^{\infty}\frac{v}{\sqrt{2\pi}}\left\{\int_{-\infty}^{\infty} u\frac{1}{\sqrt{2\pi}\sqrt{1-\rho_{XY}^2}}\exp\left[-\frac{(u-\rho_{XY}v)^2}{2(1-\rho_{XY}^2)}\right]du\right\}\cdot\exp\left[-\frac{v^2}{2}\right]dv$$

$$= \int_{-\infty}^{\infty}\frac{v}{\sqrt{2\pi}}\rho_{XY}v\cdot\exp\left[-\frac{v^2}{2}\right]dv$$

$$= \rho_{XY}\int_{-\infty}^{\infty}\frac{v^2}{\sqrt{2\pi}}\exp\left[-\frac{v^2}{2}\right]dv$$

$$= \rho_{XY}$$

Therefore, ρ_{XY} is the correlation coefficient of (X, Y)

Ans:

The conditional probability density function is:

$$f_{Y|X}(y\,|\,x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho_{XY}^2}}\exp\left[-\frac{\left(\left(\frac{y-\mu_Y}{\sigma_Y}\right)-\rho_{XY}\cdot\left(\frac{x-\mu_X}{\sigma_X}\right)\right)^2}{2(1-\rho_{xy}^2)}\right]$$

$$= \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho_{XY}^2}}\exp\left[\frac{\left(y-\left(\mu_Y+\rho_{XY}\cdot\frac{\sigma_Y}{\sigma_X}(x-\mu_X)\right)\right)^2}{2\sigma_Y^2(1-\rho_{XY}^2)}\right]$$

This is the probability density function of $N\left(\mu_Y + \rho_{XY}\cdot\frac{\sigma_Y}{\sigma_X}(x-\mu_X),\quad \sigma_Y^2(1-\rho_{XY}^2)\right)$

Therefore, $E[Y\,|\,x] = \mu_Y + \rho_{XY}\cdot\frac{\sigma_Y}{\sigma_X}(x-\mu_X)$

Distribution of Functions of Random Variables

Distribution of Functions of Random Variables

- Let X is a random variable, then we can transform X into another form, such as
- Y=log (X), or Y=1/X, or Y=-X, or Y=X^2, ... etc
- Y is still a random variable
- Distribution of Y, (a function of X), is called the distribution of function of random variable, X.

Methods for Determining the Distribution of Functions of Random Variables

1. Distribution function method
2. Moment generating function method
3. Transformation method

Transformations

Theorem

Let $x_1, x_2, \ldots, x_n$ denote random variables with joint probability density function

$$f(x_1, x_2, \ldots, x_n)$$

Let $u_1 = h_1(x_1, x_2, \ldots, x_n)$.
$u_2 = h_2(x_1, x_2, \ldots, x_n)$.
⋮
$u_n = h_n(x_1, x_2, \ldots, x_n)$.

define an invertible transformation from the x's to the u's

Then the joint probability density function of $u_1, u_2, \ldots, u_n$ is given by:

$$g(u_1, \cdots, u_n) = f(x_1, \cdots, x_n)\left|\frac{d(x_1, \cdots, x_n)}{d(u_1, \cdots, u_n)}\right| = f(x_1, \cdots, x_n)|J|$$

where $$J = \frac{d(x_1, \cdots, x_n)}{d(u_1, \cdots, u_n)} = \det\begin{bmatrix} \frac{dx_1}{du_1} & \cdots & \frac{dx_1}{du_n} \\ \vdots & & \vdots \\ \frac{dx_n}{du_1} & \cdots & \frac{dx_n}{du_n} \end{bmatrix}$$

Jacobian of the transformation

The Distribution of a Function of a Random Variable

- Here, the probability density function of a random variable *X* is known. How do we determine the probability density function of some **functions**, say *g*(*X*) of it.

Method 1: Express the event that $g(X) \le y$ in terms of *X* being in some set.

Method 2: If *g*(X) is a **strictly monotone**(either increasing or decreasing) and **differentiable** function of *X* , use the below formula to find the probability density function of random variable *Y* defined by *Y* = *g*(*X*).

$$f_Y(y) = \begin{cases} f_X[g^{-1}(y)]\left|\frac{d}{dy}g^{-1}(y)\right| & \text{if } y = g(x) \text{ for some } x \\ 0 & \text{if } y \ne g(x) \text{ for all } x \end{cases}$$

Note: $y = g(x), x = g^{-1}(y)$

Distribution of Functions of Random Variables

- Ex:
- X ~ N(μ , σ^2), then
- Z = (X- μ)/ σ ~ N(0,1) (Standard Normal)

Distribution of Functions of Random Variables

Let X have a normal distribution with mean 0, and variance 1. (standard normal distribution)

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Let $W = X^2$.
Find the distribution of *W*.

Thus if *X* has a standard Normal distribution then
$W = X^2$
has density

$$g(w) = \frac{1}{\sqrt{2\pi}} w^{-\frac{1}{2}} e^{-\frac{w}{2}} \quad \text{if } w \ge 0.$$

This distribution is the Gamma distribution with α = ½ and λ = ½. (Known as **chi-squared distribution**.)

This distribution is also the **χ^2 distribution with ν = 1** degree of freedom.
(**Chi-squared distribution with 1 degree of freedoom**)

The Joint Probability Distribution of a Function of Random Variables

- Here, the joint probability density function of two jointly continuous random variable X_1, X_2 are **known**. How do we determine the joint probability density function of the random variables Y_1 and Y_2 which are the functions of $X_1$and X_2 .

  where $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$

  It should be assumed that the functions g_1 and g_2 satisfy the following conditions:

1. g_1 and g_2 are two **one-to-one-functions**.
   So there exist **unique** h_1 and h_2 such that:
   $x_1 = h_1(y_1, y_2), x_2 = h_2(y_1, y_2)$

The Joint Probability Distribution of a Function of Random Variables

2. g_1 and g_2 have **continuous partial derivatives** at all points (x_1, x_2).
   Hence:

$$J(x_1, x_2) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{vmatrix} = \frac{\partial g_1}{\partial x_1}\frac{\partial g_2}{\partial x_2} - \frac{\partial g_1}{\partial x_2}\frac{\partial g_2}{\partial x_1} \neq 0 \qquad \text{all} x_1, x_2$$

- Under the previous two conditions, the **joint probability density function** of the random variables Y_1 and Y_2 is given by:

$$f_{Y_1Y_2}(y_1, y_2) = f_{X_1X_2}(x_1, x_2)\left|J(x_1, x_2)\right|^{-1}$$

Ex :
X and *Y* is jointly continuous random variable with the joint probability density function f_{XY}. Find the joint probability density function of *A* and *B* by given that *X* and *Y* are independent and uniformly distributed in the interval (0, 1), also: $A = X + Y, \quad B = X - Y$

Ans:

Let $g_1(x, y) = x + y, \quad g_2(x, y) = x - y$

$$J(x, y) = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2$$

$$f_{AB}(a, b) = \left|J(x, y)\right|^{-1} f_{XY}\left(\frac{a+b}{2}, \frac{a-b}{2}\right) = -\frac{1}{2} f_{XY}\left(\frac{a+b}{2}, \frac{a-b}{2}\right)$$

$$f_{AB}(a, b) = \begin{cases} \frac{1}{2} & 0 \le a+b \le 2, \quad -1 \le a-b \le 1 \\ 0 & \text{otherwise} \end{cases}$$

The Joint Probability Distribution of a Function of Random Variables

Suppose that *X* and *Y* are independent random variables each having an exponential distribution with parameter λ (mean $1/\lambda$)

$$f_1(x) = \lambda e^{-\lambda x} \text{ for } x \ge 0$$

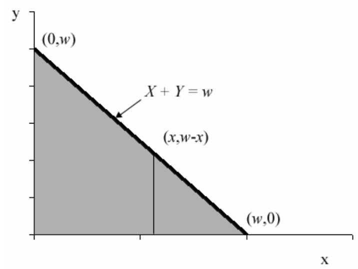
$$f_2(y) = \lambda e^{-\lambda y} \text{ for } y \ge 0$$

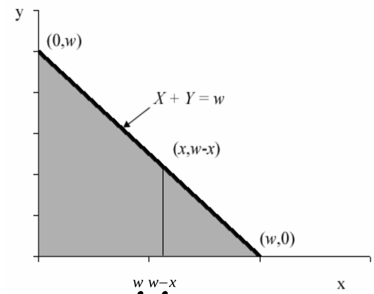
$$f(x, y) = f_1(x) f_2(y) = \lambda^2 e^{-\lambda(x+y)} \text{ for } x \ge 0, \ y \ge 0$$

Let $W = X + Y$.
Find the distribution of *W*.

First step
Find the distribution function of $W = X + Y$

$$G(w) = P[W \le w] = P[X + Y \le w]$$





$$P[X + Y \le w] = \int_0^w \int_0^{w-x} f_1(x) f_2(y)\, dy dx = \int_0^w \int_0^{w-x} \lambda^2 e^{-\lambda(x+y)}\, dy dx$$

$$P[X+Y\le w]=\int_0^w\int_0^{w-x} f_1(x)\,f_2(y)\,dydx$$
$$=\int_0^w\int_0^{w-x}\lambda^2 e^{-\lambda(x+y)}dydx$$
$$=\lambda^2\int_0^w e^{-\lambda x}\left[\int_0^{w-x} e^{-\lambda y}dy\right]dx$$
$$=\lambda^2\int_0^w e^{-\lambda x}\left[\frac{e^{-\lambda y}}{-\lambda}\right]_0^{w-x}dx$$
$$=\lambda^2\int_0^w e^{-\lambda x}\left[\frac{e^{-\lambda(w-x)}-e^0}{-\lambda}\right]dx$$

2005/10/23 *Jeff Lin, MD. PhD.* 115

$$P[X+Y\le w]=\lambda^2\int_0^w e^{-\lambda x}\left[\frac{e^{-\lambda(w-x)}-e^0}{-\lambda}\right]dx$$
$$=\lambda\int_0^w\left[e^{-\lambda x}-e^{-\lambda w}\right]dx$$
$$=\lambda\left[\frac{e^{-\lambda x}}{-\lambda}-xe^{-\lambda w}\right]_0^w$$
$$=\lambda\left[\left(\frac{e^{-\lambda w}}{-\lambda}-we^{-\lambda w}\right)-\left(\frac{e^{-0}}{-\lambda}\right)\right]$$
$$=\left[1-e^{-\lambda w}-\lambda we^{-\lambda w}\right]$$

2005/10/23 *Jeff Lin, MD. PhD.* 116

Second step

Find the density function of W

$g(w) = G'(w)$.

$$=\frac{d}{dw}\left[1-e^{-\lambda w}-\lambda we^{-\lambda w}\right]$$
$$=\left[\lambda e^{-\lambda w}-\lambda\left(\frac{dw}{dw}e^{-\lambda w}+w\frac{de^{-\lambda w}}{dw}\right)\right]$$
$$=\left[\lambda e^{-\lambda w}-\lambda e^{-\lambda w}+\lambda^2 we^{-\lambda w}\right]$$
$$=\lambda^2 we^{-\lambda w}\quad\text{for } w\ge 0$$

2005/10/23 *Jeff Lin, MD. PhD.* 117

Hence if X and Y are independent random variables each having an exponential distribution with parameter λ then W has density

$$g(w)=\lambda^2 we^{-\lambda w}\quad\text{for } w\ge 0$$

This distribution can be recognized to be the ***Gamma*** distribution with parameters α = 2 and

2005/10/23 *Jeff Lin, MD. PhD.* 118

Example: Student's t Distribution

Let Z and U be two independent random variables with:

1. Z having a Standard Normal distribution and
2. U having a χ^2 distribution with ν degrees of freedom

Find the distribution of $t=\dfrac{Z}{\sqrt{U/\nu}}$

2005/10/23 *Jeff Lin, MD. PhD.* 119

The density of Z is:

$$f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$$

The density of U is:

$$h(u)=\frac{\left(\frac{1}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}u^{\frac{\nu}{2}-1}e^{-\frac{u}{2}}$$

2005/10/23 *Jeff Lin, MD. PhD.* 120

Therefore the joint density of Z and U is:

$$f(z,u) = f(z)h(u) = \frac{\left(\frac{1}{2}\right)^{\frac{\nu}{2}}}{\sqrt{2\pi}\,\Gamma\left(\frac{\nu}{2}\right)} u^{\frac{\nu}{2}-1} e^{-\frac{z^2+u}{2}}$$

The distribution function of T is:

$$G(t) = P\left[T \le t\right] = P\left[\frac{Z}{\sqrt{U/\nu}} \le t\right] = P\left[Z \le \frac{t}{\sqrt{\nu}}\sqrt{U}\right]$$

Therefore:

$$G(t) = P\left[T \le t\right] = P\left[Z \le \frac{t}{\sqrt{\nu}}\sqrt{U}\right] =$$

$$= \int_0^\infty \int_{-\infty}^{\frac{t}{\sqrt{\nu}}\sqrt{U}} \frac{\left(\frac{1}{2}\right)^{\frac{\nu}{2}}}{\sqrt{2\pi}\,\Gamma\left(\frac{\nu}{2}\right)} u^{\frac{\nu}{2}-1} e^{-\frac{z^2+u}{2}}\, dz\,du$$

and

$$g(t) = \frac{\left(\frac{1}{2}\right)^{\frac{\nu}{2}} 2^{\frac{\nu+1}{2}}\, \Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{2\pi}\,\sqrt{\nu}\,\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{t^2}{\nu}+1\right)^{-\frac{\nu+1}{2}}$$

or

$$g(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{t^2}{\nu}+1\right)^{-\frac{\nu+1}{2}} = K\left(\frac{t^2}{\nu}+1\right)^{-\frac{\nu+1}{2}}$$

where

$$K = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}$$

Student's t Distribution

$$g(t) = K\left(\frac{t^2}{\nu}+1\right)^{-\frac{\nu+1}{2}}$$

where

$$K = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}$$

Student – W.W. Gosset

Worked for a distillery

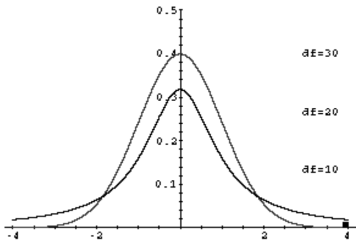
Not allowed to publish

Published under the pseudonym "Student

'Student' in 1908

t distribution

standard normal distribution



Distribution of the Max and Min Statistics

Let $x_1, x_2, \ldots, x_n$ denote a sample of size n from the density $f(x)$.

Let $M = \max(x_i)$ then determine the distribution of M.

Repeat this computation for $m = \min(x_i)$

The Probability Integral Transformation

This transformation allows one to convert observations that come from a uniform distribution from 0 to 1 to observations that come from an arbitrary distribution.

Let U denote an observation having a uniform distribution from 0 to 1.

$$g(u) = \begin{cases} 1 & 0 \le u \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

Let $f(x)$ denote an arbitrary density function and $F(x)$ its corresponding cumulative distribution function.

Let $X = F^{-1}(U)$

Find the distribution of X.

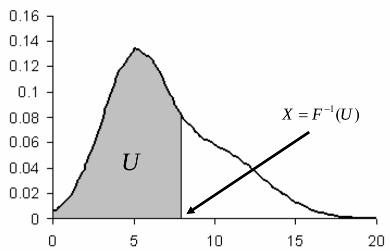
$$G(x) = P\left[X \le x\right] = P\left[F^{-1}(U) \le x\right]$$
$$= P\left[U \le F(x)\right]$$
$$= F(x)$$

Hence. $g(x) = G'(x) = F'(x) = f(x)$

Thus if U has a uniform distribution from 0 to 1. Then

$$X = F^{-1}(U)$$

has density $f(x)$.



The Distribution of a Linear Combination

Joint Distributions

- Sums and differences
  - The expectation of sums is the sum of expectations

$$E(X+Y)=E(X)+E(Y)$$

$$E(a+bX+cY)=a+bE(X)+cE(Y)$$

Joint Distributions

- Sums and differences
  - The variance of a sum is the sum of variances, plus twice the covariance

$$V(X+Y)=V(X)+V(Y)+2Cov(X,Y)$$

$$V(a+bX+cY)=b^2V(X)+c^2V(Y)+2bcCov(X,Y)$$

Joint Distributions

- Sums and differences
  - **If two variables are uncorrelated (covariance is 0),** the variance of a sum is the sum of variances

$$V(X+Y)=V(X)+V(Y)$$

$$V(a+bX+cY)=b^2V(X)+c^2V(Y)$$

Difference Between Two Random Variables

$$E(X_1-X_2)=E(X_1)-E(X_2)$$

and, if X_1 and X_2 are independent,

$$V(X_1-X_2)=V(X_1)+V(X_2)$$

Joint Distributions

- Sums and differences
  - The variance of a sum is the sum of variances plus twice the sum of every possible covariance:

$$V(a+bX+cY+dZ)=b^2V(X)+c^2V(Y)+d^2V(Z)$$
$$+2bcCov(X,Y)+2bdCov(X,Z)+2cdCov(Y,Z)$$

Joint Distributions

- Sums and differences
  - The covariance of a sum with a third variable is the sum of covariances:

$$Cov(X+Y,Z)=Cov(X,Z)+Cov(Y,Z)$$

$$Cov(a+bX+cY,Z)=bCov(X,Z)+cCov(Y,Z)$$

Sums of Random Variables and Long-Term Average

SUMS OF RANDOM VAREABLES

Let $X_1, X_2, ..., X_n$ be a sequence of random variables, and let S_n be their sum:

$$S_n = X_1 + X_2 + ... + X_n \quad (5.1)$$

In this section, we find the mean and variance of S_n, as well as the pdf of S_n in the important special case where the $X_j's$ are independent random variables.

Mean and variance of Sums of Random variables

In Section 4.7, it was shown that regardless of statistical dependence, *the expected value of a sum of n random variables is equal to the sum of the expected values:*

$$E[X_1 + X_2 + ... + X_n] = E[X_1] + ... + E[X_n] \quad (5.2)$$

Linear Combination

Given a collection of n random variables $X_1, \ldots, X_n$ and n numerical constants $a_1, \ldots, a_n$, the rv

$$Y = a_1 X_1 + ... + a_n X_n = \sum_{i=1}^{n} a_i X_i$$

is called a *linear combination* of the X_i's.

Expected Value of a Linear Combination

Let $X_1, \ldots, X_n$ have mean values $\mu_1, \mu_2, ..., \mu_n$ and variances of $\sigma_1^2, \sigma_2^2, ..., \sigma_n^2$, respectively

Whether or not the X_i's are independent,

$$E\left(a_1 X_1 + ... + a_n X_n\right) = a_1 E\left(X_1\right) + ... + a_n E\left(X_n\right)$$
$$= a_1 \mu_1 + ... + a_n \mu_n$$

Variance of a Linear Combination

If $X_1, \ldots, X_n$ are **independent,**

$$V\left(a_1 X_1 + ... + a_n X_n\right) = a_1^2 V\left(X_1\right) + ... + a_n^2 V\left(X_n\right)$$
$$= a_1^2 \sigma_1^2 + ... + a_n^2 \sigma_n^2$$

and

$$\sigma_{a_1 X_1 + ... + a_n X_n} = \sqrt{a_1^2 \sigma_1^2 + ... + a_n^2 \sigma_n^2}$$

Variance of a Linear Combination

For any $X_1, \ldots, X_n$,

$$V\left(a_1 X_1 + ... + a_n X_n\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \mathrm{Cov}\left(X_i, X_j\right)$$

Covariance of Random Variables

- Relation between variance and covariance:

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right)$$
$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{Cov}\left(X_i, X_j\right)$$
$$= \sum_{i=1}^{n} \mathrm{Var}\left(X_i\right) + \sum \sum_{i \neq j} \mathrm{Cov}\left(X_i, X_j\right)$$
$$= \sum_{i=1}^{n} \mathrm{Var}\left(X_i\right) + 2 \sum \sum_{i < j} \mathrm{Cov}\left(X_i, X_j\right)$$

If X_i and X_j are **independent**, then :

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}\left(X_i\right)$$

Covariance of Random Variables

- Properties of covariance:

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
2. $\text{Cov}(X, X) = \text{Var}(X)$
3. $\text{Cov}(aX, Y) = a\,\text{Cov}(X, Y)$
4. $\text{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} X_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \text{Cov}(X_i, Y_j)$

Difference Between Normal Random Variables

If $X_1, X_2, \ldots X_n$ are independent, normally distributed rv's, then any linear combination of the X_i's also has a normal distribution. The difference $X_1 - X_2$ between two independent, normally distributed variables is itself normally distributed.

Mean and Variance of an Average

- If $\overline{X} = (X_1 + X_2 + \ldots + X_p) / p$ with $E(X_i) = \mu$ for i= 1, 2, …, p

$$E(\overline{X}) = \mu$$

- If $X_1, X_2, .., X_p$ are also independent with V(Xi) = σ^2 for i= 1, 2, …, p,

$$V(\overline{X}) = \sigma^2 / p$$

Some Important Results

Multiple Discrete Random Variables
Multinomial Probability Distribution

Suppose a random experiment consists of a series of n trials. Assume that

(1) The result of each trial is classified into one of k classes.

(2) The probability of a trial generating a result in class 1, class 2, … , class k is constant over the trials and equal to $p_1, p_2, \ldots, p_k$, respectively.

(3) The trials are independent.

The random variables $X_1, X_2, \ldots, X_k$ that denote the number of trials that result in class 1, class 2, … , class k, respectively, have a **multinomial distribution** and the joint probability mass function is

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{n!}{x_1!x_2!\cdots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k} \quad (5\text{-}13)$$

for $x_1 + x_2 + \cdots + x_k = n$ and $p_1 + p_2 + \cdots + p_k = 1$.

Multiple Discrete Random Variables
Multinomial Probability Distribution

Each trial in a multinomial random experiment can be regarded as either generating or not generating a result in class i, for each $i = 1, 2, \ldots, k$. Because the random variable X_i is the number of trials that result in class i, X_i has a binomial distribution.

If $X_1, X_2, \ldots, X_k$ have a multinomial distribution, the marginal probability distribution of X_i is binomial with

$$E(X_i) = np_i \quad \text{and} \quad V(X_i) = np_i(1 - p_i) \quad (5\text{-}14)$$

Some Important Results

- X ~ N(μ , σ^2), then
  Z = (X- μ)/ σ ~ N(0,1) (Standard Normal)
- X ~ N(μ , σ^2), then
  $W = Z^2$ is χ^2 ν=1 (chi-squared dist. With 1.d.f.)
- Z ~ N(0,1), $U \sim \chi^2$ ν=m degrees of freedom, then
- $t = \frac{Z}{\sqrt{U/\nu}}$ is Student's t distribution with m d.f.
- $U \sim \chi^2$ ν=m, $V \sim \chi^2$ ν=n, then
  (U/m) / (V/n) ~ F (m, n) d.f.
- T^2 = F (1,m)

Some Important Results

- Let $X_1, X_2, \dots , X_n$ denote a independent identical sample of size n from the density $f(x)$.
- Let Y = $\sum X_1 + X_2 + \dots + X_n$
- Sum of Bernoulli random variables is a Binomial random variable
- Sum of Poisson (λ) is a random variable with Poisson (n λ) .
- Sum of Exponential (θ) is a random variable with Gamma (n, θ)

Some Important Results

- Let $X_1, X_2, \dots , X_n$ denote a independent identical sample of size n from the density $f(x)$.
- Let Y = $\sum X_1 + X_2 + \dots + X_n$
- Sum of Normal N(μ , σ^2), is a random variable with N(n μ ,n σ^2).
- $\overline{X}$ is a random variable with N(μ , σ^2 /n).

Expected Value of a Linear Combination

Let $X_1,\dots, X_n$ have mean values $\mu_1, \mu_2, ..., \mu_n$ and variances of $\sigma_1^2, \sigma_2^2, ..., \sigma_n^2$, respectively

Whether or not the X_i's are independent,

$$E\left(a_1 X_1 + ... + a_n X_n\right) = a_1 E\left(X_1\right) + ... + a_n E\left(X_n\right)$$
$$= a_1 \mu_1 + ... + a_n \mu_n$$

Variance of a Linear Combination

For any $X_1,\dots, X_n$,

$$V\left(a_1 X_1 + ... + a_n X_n\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \mathrm{Cov}\left(X_i, X_j\right)$$

Thanks !