743- Regression and Time Series

Mamikon S. Ginovyan

Probability Review

2. Random Variables

- **Def.1**. A – real-valued function whose domain is the sample space Ω is called a Random Variable (RV).

  Thus RV is any function

$$X = X(w),\ w \in \Omega$$

  that assigns a numerical value to each possible outcome $w \in \Omega$.

  Symbolically,
$$X(w) : \Omega \to R.$$

- RV's are classified according to the number of values they can assume, that is, according to the range of the function $X(w)$:

$$D = \{x : x = X(w), w \in \Omega\}.$$

Random Variables

Def.2. If the range ***D*** of $X(w)$ contains either a **finite** or **countable infinite** number of values, the RV is called **discrete** RV.

Def.3. If the range ***D*** includes an interval *(a, b)* of real numbers, bounded or unbounded, the RV $X(w)$ is called **continuous** RV.

3. Description of RV's. CDF and properties

- Any RV $X(w)$ (discrete or continuous) is **completely described** by **Cumulative Distribution Function (cdf)** $F(x)$ defined by

$$F(x) = F_X(x) = P(X \le x) = P\{w : X(w) \le x\}.$$

- **It is easy to check that any *cdf* *F(x)* satisfies the following conditions.**

3. Description of RV's. CDF and properties

Theorem 1. (Properties of cdf)

1) $F(x)$ is a non decreasing function, that is, for all $x_1, x_2 \in R$, if

$$x_1 < x_2, F(x_1) \le F(x_2)$$

2) $\lim_{x \to -\infty} F(x) = 0$ (the lower limit of ***F*** is 0).

3) $\lim_{x \to +\infty} F(x) = 1$ (the upper limit of ***F*** is 1).

4) $\lim_{x \downarrow x_0} F(x) = F(x_0)$ (***F*** is right continuous).

The converse is also true.

Description of RV's. CDF and properties

Theorem 2. (Kolmogorov). If $F(x)$, $x \in R$, is any real-valued function satisfying conditions (1)-(4), then there exist (it can be constructed)

a) a probability space $(\Omega, \Im, \mathrm{P})$ and

b) a RV $X = X(w), w \in \Omega$, with cdf $G_X(x)$ such that

$$G_X(x) = F(x).$$

The next theorem is helpful in computation of probabilities using cdf's:

Theorem 3. If $X(w)$ is a RV with cdf $F(x)$, then for $a<b$,

$$P\{a < X \leq b\} = F_X(b) - F_X(a).$$

Discrete Random Variables (DRV)

- The probabilistic structure (probability distribution) of any DRV $\boldsymbol{X}$ is can be described by **Probability Mass Function (*pmf*)**.

- Def. Let $X(w)$ be a DRV with range $\boldsymbol{D}$. The function

$$f(x) = f_X(x) = P\{w : X(w) = x\}, x \in D$$

  is called ***pmf*** of $\boldsymbol{X}$.

- **Properties of pmf.**
  Any *pmf* satisfies the conditions:

  a) $0 \le f(x) \le 1$ for all $x \in D$.

  b) $\sum_{x \in D} f(x) = 1.$

The converse is also true.

Discrete Random Variables (DRV)

- **Probability distribution of DRV's**
- a) **Tabular form**

X	x_1	x_2	…	x_k	…
$f(x)$	$f(x_1)$	$f(x_2)$	…	$f(x_k)$	…

- **Example**. Let X be a DRV with probability distribution

X	x_1	x_2	x_3	x_4	x_5
$f(x)$	1/9	p_2	1/3	p_4	1/9

Assuming that the distribution is **symmetric**,

find p_2 and p_4. **Answer:** $p_2 = p_4 = 2/9.$

Discrete Random Variables (DRV)

- **The relationship between *cdf F(x)* and *pmf f(x)*.**

Let $X(w)$ be a DRV with cdf $F(x)$ and pmf $f(x)$. Then

a) $f(x) = P(X = x) = F(x) - F(x-)$

b) $$F(x) = \sum_{k:\, x_k \le x} f(x_k)$$

c) For any two numbers a and b, $a \le b$

$$P(a \le X \le b) = F(b) - F(a-),$$

where

$$F(a-) = \lim_{x \uparrow a} F(x).$$

Discrete Random Variables (DRV)

Remark. If all possible values of a DRV ***X*** are integers: $x_k = k$, and a and b are integers, then

$$P(a \le X \le b) = P(X = a \quad or \quad a+1 \quad or.... \quad b)$$

$$= F(b) - F(a-1)$$

In particular, taking $a = b$ we get

$$f(a) = P(X = a) = F(a) - F(a-1)$$

The graph of *cdf* *F(x)* is a **step** function

Discrete Random Variables (DRV)

- **Examples of DRV.**

Ex. 1. Degenerate (Singular) RV.

A DRV X is called **Degenerate (or Singular)** RV if its constant with probability 1 :

$$P(X = m) = 1.$$

So the distribution of X is

X	m
$f(x)$	1

Examples of DRV

- **Ex. 2. Bernoulli RV.**

  A DRV $\boldsymbol{X}$ is called **Bernoulli** RV with probability of success p:

$$X \sim Ber(p),$$

if its pmf $f(x)$ is given by

$$f(x) = p^x (1-p)^{1-x}, x = 0,1.$$

The distribution of $\boldsymbol{X}$ is given by

X	**0**	**1**
$f(x)$	P	$q=1-p$

Examples of DRV

∨ **Ex. 3. Binomial RV.** A DRV ***X*** is called **Binomial** with parameters ***n*** = the number of trials and ***p*** = the probability of success: $X \sim Bin(n, p)$, if its pmf *f(x)* is given by

$$f(x) = b(x, n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0, 1, \ldots, n \\ 0, & otherwise. \end{cases}$$

For cdf *F(x)* we have

$$F(x) = B(x, n, p) = \sum_{k=0}^{x} b(k, n, p)$$

•**Remark**. The **Indicator function** make convenient to write the pmf (or pdf).

Examples of DRV

Def. Given a set $A \subset R$, we denote by $I_A = I_A(x)$ the indicator function of ***A*** to be

$$I_A(x) \;=\; \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A. \end{cases}$$

For example, if $A = \{0,1,\ldots,n\} = S_X$ = the support of $X \sim Bin(n, p)$, then the pmf $f(x) = b(x, n, p)$, we can write as

$$f(x) = b(x, n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0,1,\ldots,n \\ 0, & otherwise \end{cases}$$

$$= \binom{n}{x} p^x (1-p)^{n-x} I_A(x).$$

Examples of DRV

- **Ex. 4. Poisson RV**.

  A DRV $\boldsymbol{X}$ is called **Poisson** RV with parameter λ $(\lambda > 0)$ if the pmf of $\boldsymbol{X}$ is given by

$$f(x) = f(x, \lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!} I_A(x),$$

  where $A = S_X = \{0, 1, 2, \ldots, \}$.

5. Continuous RV's

- **Def.** A RV X is called **continuous** RV if its *cdf* $F(x) = P(X \le x)$ is a continuous function for all $x \in R$.

- If ***F(x)*** is absolutely continuous, that is,

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

for some function ***f(t)***, then ***f(t)*** is called **probability density function** *(pdf)* of RV X.
By the Fundamental Theorem of Calculus we have

$$f(x) = F'(x) = \frac{dF(x)}{dx}.$$

Continuous RV's

- **Properties of *pdf's*** .

1. $f(x) \geq 0$ for all $x \in R$
   (*follows from non-decreasing property of cdf* $F(x)$);

2. $\int_{-\infty}^{\infty} f(x)dx = 1$

   (*follows from* $F(+\infty) = 1$).

- **Remark.** It can be shown that any function ***g(x)***, $x \in R$ satisfying 1) and 2) is a *pdf* of some CRV ***X*** .

Continuous RV's

- **Property.** For any $a, b \in R, (a \le b)$,

$$P(a \le X \le b) = F(b) - F(a) = \int_a^b f(x)dx.$$

- **Important Remark**. For any CRV X the probability of any single point is equal to 0:

$$P(X = a) = 0.$$

This is the principal difference between discrete and continuous RV-s.

Therefore, $P(a \le X \le b) = P(a \le X < b) = P(a < X \le b)$

$= P(a < X < b).$

The Normal Distribution

The normal distribution is the cornerstone of modern statistics.

<u>**Definition.**</u> The **normal RV** ***X*** is a CRV with **probability density function (pdf)** *f(x)* given by formula

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty,$$

$\boldsymbol{\mu}$ = E[***X***]= the mean of RV ***X***

$\sigma^2 = Var(X) = E(X-\mu)^2$ = the variance of ***X***

$\boldsymbol{\sigma}$ = the standard deviation of ***X***

$\boldsymbol{\pi}$ = 3.1416…

e = 2.71828 …

Relationship between General and Standard Normal RV's.

Result: Let $X \sim N(m, s^2)$ be an **arbitrary normal** RV, and let $Z \sim N(0,1)$ be the **standard normal** RV. Then the relationship between X and Z is given by

$$\boxed{Z = \frac{X - m}{s}} \qquad \Leftrightarrow \qquad \boxed{X = s\,Z + m}$$

Thus, $X \sim N(m, s^2)$ **if and only if** $Z \sim N(0,1)$

Result: If $\boldsymbol{a}$ and $\boldsymbol{b}$ are any values of $\boldsymbol{X}$, with $\boldsymbol{a} < \boldsymbol{b}$. Then

$$\boxed{P\left(a \le X \le b\right) = P\left(\frac{a - m}{s} \le Z \le \frac{b - m}{s}\right)}$$

6. Numerical Characteristics of RV's

- **The Expected Value.**

$$m_X = E(X) = \sum_{k=1}^{\infty} x_k f(x_k),$$

$$\text{if} \quad \sum_{k=1}^{\infty} |x_k| f(x_k) < \infty \quad \text{and } \boldsymbol{X} \text{ is a DRV;}$$

$$m_X = E(X) = \int_{-\infty}^{\infty} x f(x) dx,$$

$$\text{if} \quad \int_{-\infty}^{\infty} |x| f(x) dx < \infty \quad \text{and } \boldsymbol{X} \text{ is a CRV.}$$

Numerical Characteristics of RV's

Theorem 1. (Expected value of a function of a RV).

Let $\boldsymbol{X}$ be a RV with *pdf* (or *pmf*) $f_X(x)$, and let $y = g(x)$, $x \in R$ be some function. Then

(a) $E[Y] = E[g(X)] = \sum_{k=1}^{\infty} g(x_k) f_X(x_k)$

if $\sum_{k=1}^{\infty} |g(x_k)| f_X(x_k) < \infty$ and $\boldsymbol{X}$ is a DRV;

(b) $E[Y] = E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$

if $\int_{-\infty}^{\infty} |g(x)| f(x) dx < \infty$ and $\boldsymbol{X}$ is a CRV.

Numerical Characteristics of RV's

Theorem 2.

Let $\boldsymbol{X}$ be a RV and let $g_k(x)$, $k=1,\dots,n$, be some functions such that $E[g_k(X)]$ exist for all $k=1,\dots,n$.

Then for any constants $c_1, c_2, \dots c_n$, the expectation of the RV $\sum_{k=1}^{n} c_k g_k(X)$ exists and is given by

$$E\left[\sum_{k=1}^{n} c_k g_k(X)\right] = \sum_{k=1}^{n} c_k E\left[g_k(X)\right].$$

Numerical Characteristics of RV's

- **The Variance**
- **Def.** Let X be a RV with finite mean $m = E(X)$ and such that the expectation $E\left[(X-m)^2\right]$ is finite.

Then the **Variance** of $\boldsymbol{X}$ is defined to be

$$s^2 = Var(X) = E(X-m)^2 = \begin{cases} \sum_{k=1}^{\infty}(x_k - m)^2 f(x_k) \\ \int (x-m)^2 f(x)dx. \end{cases}$$

Numerical Characteristics of RV's

- **Properties**.

1. Let $\boldsymbol{X}$ be a RV s.t. $E[X^2]$ exists, then

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

2. The variance of a constant is equal to 0: $Var(C) = 0.$

3. Let $\boldsymbol{X}$ be a RV s.t. $E[X^2]$ exists, then for any constants a and b

$$Var(aX + b) = a^2 Var(X).$$

7. Joint Distributions of RV's.

Consider two RV's $X_1 = X_1(w)$ and $X_2 = X_2(w)$ defined on $\Omega (w \in \Omega)$, which assign to each possible outcome $w \in \Omega$ **one and only one** ordered pair of numbers

$$X_1(w) = x_1,\ X_2(w) = x_2.$$

The **joint distribution** of RV's X_1 and X_2 is completely described by the joint **Cumulative Distribution Function (*cdf*)** $F_{X_1,X_2}(x_1,x_2)$ defined by

$$F(x_1,x_2) = F_{X_1,X_2}(x_1,x_2)$$
$$= P\{w : X_1(w) \le x_1,\ X_2(w) \le x_2\} \quad \text{for all} \quad (x_1,x_2) \in R^2.$$

Joint Distributions of RV's.

- **DRV's:** If X_1 and X_2 are DRV's, then the function

$$f_{X_1,X_2}(x_1,x_2) = P(X_1 = x_1,\ X_2 = x_2)$$

is called **joint *pmf*** of X_1 and X_2 .

- A *pmf* $f(x_1,x_2)$ is characterized by the two **properties**

$$a)\ \ 0 \le f(x_1,x_2) \le 1$$

$$b)\ \ \sum_{x_1}\sum_{x_2} f(x_1,x_2) = 1.$$

Joint Distributions of RV's

- The joint *pmf* $f(x_1,x_2)$ uniquely defines the joint *cdf* $F(x_1,x_2)$:

$$F(x_1,x_2)=\sum_{t_1\le x_1}\sum_{t_2\le x_2} f(t_1,t_2)$$

- If the possible values of X_1 and X_2 are

$X_1: \quad x_{11},\ldots,x_{1M}$ (generally $M \neq N$ and $M=\infty, N=\infty$)

$X_2: \quad x_{21},\ldots,x_{2N}$

and $P_{kj}=f(x_{1k},x_{2j})=P(X_1=x_{1k}\ ,\ \ X_2=x_{2j})$,

then the probability distribution of $X=(X_1,X_2)$ is given by

the following **joint probability table.**

Joint Distributions of RV's

X_1 \ X_2	x_{11}		$\cdots$	x_{1k}	$\cdots$	
x_{21}	P_{11}		$\cdots$	P_{k1}	$\cdots$	
x_{22}	P_{12}		$\cdots$	P_{k2}	$\cdots$	
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_{2j}	P_{1j}		$\cdots$	P_{kj}	$\cdots$	
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_{2N}	P_{1N}		$\cdots$	P_{kN}	$\cdots$	

Marginal pmf's

- Let X_1 and X_2 be DRV's with joint *pmf* $f_X(x_1, x_2)$. Then the **marginal pmf's** of X_1 and X_2, respectively, are given by

$$f_{X_1}(x_1) = \sum_{x_2} f(x_1, x_2) \qquad \text{and} \qquad f_{X_2}(x_2) = \sum_{x_1} f(x_1, x_2).$$

- **Example**. Assume that the distribution of $X=(X_1, X_2)$ is given by

	0	**1**
-1	.1	.2
0	.2	.3
1	0	.2

Find the marginal distributions of X_1 and X_2.

Example:

- **Answer**.

X_1	**0**	**1**
$f_{X_1}(x_1)$	.3	.7

X_2	-1	0	1
$f_{X_2}(x_2)$	.3	.5	.2

Joint Distributions of RV's

CRV's: Let X_1 and X_2 be CRV's with joint *cdf* $F(x_1, x_2)$.

- If there exists a function $f(x_1, x_2)$ such that

$$F(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(t_1, t_2) dt_1 dt_2 \quad \text{for all} \quad x_1, x_2 \in R,$$

then X_1 and X_2 are called jointly continuous RV's. Observe that

$$f(x_1, x_2) = \frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2}$$

The function $f(x_1, x_2)$ is called a **joint pdf** of X_1 and X_2.

- Any **joint pdf** $f(x_1, x_2)$ satisfies the following conditions:

a) $f(x_1, x_2) \geq 0$ for all $x_1, x_2 \in R$; b) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) dx_1 dx_2 = 1$.

Joint Distributions of RV's (CRV's)

- The **marginal** ***pdf***'s of X_1 and X_2 are given by

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)dx_2, x_1 \in R$$

$$f_{X_2}(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2)dx_1, x_2 \in R$$

v **Example 1.** Let X_1 and X_2 have the joint *pdf*

$$f(x_1, x_2) = \begin{cases} 4x_1, x_2, & 0 \le x_1 \le 1,\ 0 \le x_2 \le 1 \\ 0, \ elsewhere . \end{cases}$$

Find the marginal *pdf*'s $f_1(x_1)$ and $f_2(x_2)$.

Joint Distributions of RV's (CRV's)

- **Solution.** We have

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1 x_2)\, dx_2 = \int_0^1 4 x_1 x_2\, dx_2 = 4x_1 \left(\frac{x_2^2}{2} \right) \Bigg|_0^1$$

$$= 2x_1,\ 0 \le x_1 \le 1.$$

Similarly, we find

$$f_2(x_2) = 2x_2, 0 \le x_2 \le 1.$$

8. Independent Random Variables

- **Def**. Let X_1 and X_2 be two RV's with joint *cdf* $F(x_1, x_2)$ and marginal *cdf*'s $F_1(x_1)$ and $F_2(x_2)$. Then X_1 and X_2 are called **independent** *iff*

$$F(x_1, x_2) = F_1(x_1) \cdot F_2(x_2),$$

otherwise they called **dependent**.

- In terms of *pmf*'s and *pdf*'s we have

$$f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2).$$

- **Theorem 1.** The RV's X_1 and X_2 are independent if and only if for any constants $a_1 < b_1$ and $a_2 < b_2$,

$$P(a_1 < X_1 < b_1, a_2 < X_2 < b_2) = P(a_1 < X_1 < b_1) \cdot P(a_2 < X_2 < b_2).$$

Independent Random Variables

- **Example 2.**

  Show that the RV's X_1 and X_2 in **Example 1** are independent.

  Indeed, in Example 1 we have

$$f(x_1, x_2) = \begin{cases} 4x_1x_2, & 0 \le x_1, x_2 \le 1 \\ 0 & elswhere; \end{cases} \qquad f_1(x_1) = 2x_1; \quad f_2(x_2) = 2x_2.$$

  So $f(x_1,x_2)=f_1(x_1)\cdot f_2(x_2)$, and X_1 and X_2 are independent.

- **Remark:** The following result makes the verification of independence easier.

Independent Random Variables

Theorem 2.

Let (X_1, X_2) be a random vector with joint *pdf* $f(x_1, x_2)$.

Then X_1 and X_2 are independent if and only if there exist functions $t_1(\cdot)$ and $t_2(\cdot)$ such that

$$f(x_1, x_2) = t_1(x_1) \cdot t_2(x_2).$$

Theorem 3. Suppose that X_1 and X_2 are independent, and *g(t)* and *h(t)* are functions such that $E[g(X_1)]$ and $E[h(X_2)]$ exist.

Then,

$$E[g(X_1)h(X_2)] = E[g(X_1)] \cdot E[h(X_2)].$$

9. Measuring the Dependence between two RV's: The Covariance and Correlation.

We will distinguish two types of dependence between two RV's X_1 and X_2:

a) **Deterministic (or functional)**, when X_1 and X_2 are connected by a formula: $X_2 = g(X_1)$. **For example**, $X_2 = aX_1 + b$, $X_2 = X_1^2$.

b) **Non-deterministic (or stochastic).** A measure of stochastic dependence between X_1 and X_2 is the **covariance** of X_1 and X_2.

The Covariance and Correlation.

- **Remark.** Observe that $a) \Rightarrow b)$.
- **Functional drpendence implies stochastic drpendence.**
- **The converse is not true.**

- **Def.** The **covariance** of RV's X_1 and X_2 is defined by

$$Cov(\mathrm{X}_1, X_2) = E\left[\left(\mathrm{X}_1 - E[\mathrm{X}_1]\right)\left(X_2 - E[\mathrm{X}_2]\right)\right].$$

The Covariance and Correlation

Properties of **Covariance** $Cov(X_1, X_2)$:

1) $Cov(X_1, X_2) = E[X_1 X_2] - E[X_1] \cdot E[X_2]$

2) $Cov(X_1, X_2) = Cov(X_2, X_1)$

3) $Cov(X_1, X_1) = Var(X_1)$

4) $Cov(X_1 + a, X_2 + b) = Cov(X_1, X_2)$ for all constants a and b.

5) $Cov(aX_1 + bX_2, Y) = a \cdot Cov(X_1, Y) + bCov(X_2, Y)$

6) $Cov\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j Cov(X_i, Y_j)$

7) $Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 Var(X_i) + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{m} a_i a_j Cov(X_i, X_j)$

8) For independent $X_1, \dots, X_n$, $Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 Var(X_i)$.

Correlation versus Independence

The covariance of two random variables is important as an indicator of the relationship between them.

To explain the situation, consider the case where X and Y are the indicator variables for events $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively:

$$X = I_A(w) = \begin{cases} 1, & \text{if } w \in A \\ 0, & \text{if } w \in A^c \end{cases} = \begin{cases} 1 & \text{if } A \text{ occur} \\ 0 & \text{if } A \text{ does not occur} \end{cases}$$

and

$$Y = I_B(w) = \begin{cases} 1, & \text{if } w \in B \\ 0, & \text{if } w \in B^c \end{cases} = \begin{cases} 1 & \text{if B occur} \\ 0 & \text{if B does not occur.} \end{cases}$$

Correlation vs Independence

and note that

$$XY = I_{A\cap B}(w) = \begin{cases} 1 & \text{if} \quad w \in A\cap B, \\ 0 & \text{if} \; w \in (A\cap B)^c \end{cases} = \begin{cases} 1 & \text{,if} \quad X=1 \;\; and \;\; Y=1 \\ 0, & \qquad\qquad\quad \text{otherwise.} \end{cases}$$

•Observe that $I_A(w)$ is a DRV with probability distribution

I	*1*	*0*
P	*P(A)*	$P(A^c)$

Correlation vs Independence

Taking into account that

$$E[X]=E[I_A]=P(A), \qquad E[Y]=E[I_B]=P(B)$$

and $E[XY]=E[I_{A\cap B}]=P(A\cap B)$, we obtain

$$\begin{aligned}
Cov(X,Y) &= E[XY]-E[X]E[Y] \\
&= P(A\cap B)-P(A)P(B) \\
&= P(X=1,Y=1)-P(X=1)P(Y=1)
\end{aligned}$$

From this we see that

$$\begin{aligned}
&Cov(X,Y)>0 \Leftrightarrow P(X=1,Y=1)>P(X=1)P(Y=1) \\
&\Leftrightarrow \frac{P(X=1,Y=1)}{P(X=1)}>P(Y=1) \\
&\Leftrightarrow P(Y=1\,|\,X=1)>P(Y=1).
\end{aligned}$$

Correlation vs Independence

Thus:

- The covariance of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is **positive** if and only if the outcome $\boldsymbol{X=1}$ makes it **more likely** that $\boldsymbol{Y=1}$.

- Also, note that $Cov(X,Y)=0$ if and only if $\boldsymbol{A}$ and $\boldsymbol{B}$ are **independent.**

Correlation vs Independence

- **In general,** it can be shown that
- **a positive** value of $Cov(X,Y)$ is an indication that ***Y*** tends to **increase** as ***X*** does, whereas
- **a negative** value of $Cov(X,Y)$ indicates that ***Y*** tends to **decrease** as ***X*** increases.

- The strength of the relationship between ***X*** and ***Y*** is indicated by the **correlation** between ***X*** and ***Y*** .

Correlation vs Independence

- **Def. 1.** The number

$$r(X,Y) = Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

is called **correlation coefficient** between ***X*** and ***Y***.

- **Def. 2.** The RV's ***X*** and ***Y*** are called **uncorrelated** if

$$r(X,Y) = 0$$

- (or which is the same, $Cov(X,Y) = 0$).

Correlation vs Independence

- **Theorem 1.**

  (a) If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent, then they also are uncorrelated.

  (b) The converse, generally, is not true.

  (c) If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are **bivariate normal** RV's,

  then the converse is true.

Correlation vs Independence

***Proof* (a).** For simplicity, we prove for DRV's.

Assume that ***X*** and ***Y*** are independent, that is,

$$P\{X = x, Y = y\} = P\{X = x\} \cdot P\{Y = y\}.$$

We need to prove that $Cov(X,Y) = 0,$ or $E[XY] = E[X]E[Y].$

We have

$$\begin{aligned} E[XY] &= \sum_j \sum_i x_i y_j P\{X = x_i, Y = y_j\} \\ &= \sum_j \sum_i x_i y_j P\{X = x_i\} P\{Y = y_j\} \qquad (\textit{by independence}) \\ &= \sum_y y_j P\{Y = y_j\} \sum_i x_i P\{X = x_i\} \\ &= E[Y]E[X] \end{aligned}$$

Correlation vs Independence

***Proof* (b): Example showing (b).**

Consider the sample space

$$\Omega = \{(-2,4),\ \ (-1,1),\ \ (0,0),\ \ (1,1),\ \ (2,4)\},$$

where each point is assumed to be equally likely, that is,

$$P(-2,4) = P(-1,1) = \ldots = P(2,4) = 1/5.$$

Define the RV $\boldsymbol{X}$ to be the first component of a sample point, and $\boldsymbol{Y}$, the second, that is,

$\boldsymbol{X(-2, 4) = -2; \quad Y(-2, 4) = 4,}$ and so on.

Now, we show that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are **uncorrelated,** while they are **dependent.**

Example Showing (b)

Solution: 1) First we construct the joint distribution of ***(X,Y)*** .

We have

Y \ X	-2	-1	0	1	2	$f_Y(y)$
0	0	0	1/5	0	0	**1/5**
1	0	1/5	0	1/5	0	**2/5**
4	1/5	0	0	0	1/5	**2/5**
$f_X(x)$	**1/5**	**1/5**	**1/5**	**1/5**	**1/5**	**1**

Example (Showing (b))

To compute the marginal distributions (pmf's) $f_X(x)$ and $f_Y(y)$

we used the formulas

$$f_X(x) = \sum_y f(x,y); \quad f_Y(y) = \sum_x f(x,y).$$

2) We show that

$$Cov(X,Y) = EXY - X \cdot EY = 0$$

We will use the following properties of the expectations:

a) $E[aX + bY] = aE[X] + bE[Y]$

b) $E[a] = a,$ $\quad$ ***a*** is a constant.

Example Showing (b)

3) We compute

$$E(XY) = \sum_{x,y} xyf(x,y) = \sum_{i=1}^{5}\sum_{j=1}^{3} x_i y_j f(x_i, y_j)$$

$$= 1(-1)\frac{1}{5} + 1(1)\frac{1}{5} + 4(-2)\frac{1}{5} + 4(2)\frac{1}{5}$$

$$= [(-8) + (-1) + 0 + 1 + 8]\cdot\frac{1}{5} = 0$$

$$E(X) = \sum_{x} xf_X(x) = [(-2) + (-1) + 0 + 1 + 2]\cdot\frac{1}{5} = 0$$

$$E(Y) = \sum_{y} yf_Y(y) = [4 + 1 + 0 + 1 + 4]\cdot\frac{1}{5} = 2.$$

Example (Showing (b))

Thus,

$$Cov(X,Y) = EXY - X \cdot EY = 0 - 0(2) = 0$$

and so ***X*** and ***Y*** are **uncorrelated**.

4) We have $P(X=1,\ Y=1) = \frac{1}{5}$, while

$$P(X=1) \cdot P(Y=1) = \frac{1}{5} \cdot \frac{2}{5} = \frac{2}{25} \neq \frac{1}{5}.$$

Thus, ***X*** and ***Y*** **are not independent**.

Theorem 1

***Proof* (c).** Suppose that $(\boldsymbol{X},\boldsymbol{Y})$ has a Bivariate normal distribution

$$N(m_1, m_2, s_1^{\,2}, s_2^{\,2}, r)$$

that is, the *pdf* *f(x,y)* of $(\boldsymbol{X},\boldsymbol{Y})$ is given by

$$f(x,y)=\frac{1}{2p\,s_1 s_2\sqrt{1-r^2}}\times$$

$$\exp\left\{-\frac{1}{2(1-r^2)}\left[\left(\frac{x-m_1}{s_1}\right)^2-2r\left(\frac{x_1-m_1}{s_1}\right)\left(\frac{y-m_2}{s_2}\right)+\left(\frac{y-m_2}{s_2}\right)^2\right]\right\} \qquad (1)$$

where $m_1=E(X), m_2=E(Y), s_1^{\,2}=Var(X), s_2^{\,2}=Var(Y),$ and $r=r(X,Y).$

Theorem 1

Assuming that ***X*** and ***Y*** are uncorrelated, we have $r = 0$.

Hence, from (1) we obtain

$$f(x,y) = \frac{1}{2p s_1 s_2} \exp\left\{-\frac{1}{2}\left[\left(\frac{x-m_1}{s_1}\right)^2 + \left(\frac{y-m_2}{s_2}\right)^2\right]\right\}$$

$$= \frac{1}{\sqrt{2p}\, s_1} \exp\left\{-\frac{1}{2}\left(\frac{x-m_1}{s_1}\right)^2\right\} \times \frac{1}{\sqrt{2p}\, s_2} \exp\left\{-\frac{1}{2}\left(\frac{y-m_2}{s_2}\right)^2\right\} = f_X(x) f_Y(y).$$

Thus, $f_{X,Y}(x,y) = f_X(x) f_Y(y)$

where $f_X(x) \sim X \sim N(m_1, s_1{}^2)$ and $f_Y(y) \sim Y \sim N(m_2, s_2{}^2)$.

Theorem 1

Therefore, two normally distributed RV's ***X*** and ***Y*** are independent if and only if they are uncorrelated $r(X,Y)=0.$

The next result shows that the correlation coefficient $r(X_1,X_2)$ provides a **measure of linear dependence** between RV's X_1 and X_2.

Correlation vs Independence

- **Theorem 2**

The **correlation coefficient** $r(X_1, X_2)$ satisfies the following properties.

a) $|r(X_1, X_2)| \le 1$ (**Correlation inequality**).

b) $|r(X_1, X_2)| = 1$, if and only if $X_2 = aX_1 + b$, where a and b are constants.

Moreover,
$$r(X_1, X_2) = \begin{cases} 1, & a > 0 \\ -1, & a < 0. \end{cases}$$