INTRODUCTION

TO

PROBABILITY THEORY

by Victor Ohanyan

Yerevan-2014

L E C T U R E 1

§1. INTRODUCTION

Today we begin to study probability theory which is a branch of mathematics. Every person has some intuitive interpretation of the meaning of the concept of probability. We may ask such questions as, "what is the probability that it will rain tomorrow", "what is the probability that the waiting time for a bus is less than 10 minutes", and so on. When speaking about probability, the usual picture formed in one's mind is that of dice throwing, card playing, or any other games of chance.

The birth of Probability Theory is considered to be in the XVII century and is connected with the combinatorial problems of gambling games. It is difficult to consider gambling games a serious business. But the problems arising from them couldn't be solved within the framework of the mathematical models that existed at that time. For understanding probability models which we will consider in this course, we will often return to different examples from gambling games.

In the beginning of the last century from natural sciences more serious problems have been arisen. These problems led to the development of a vast part of mathematics which is probability theory. This field of knowledge up to the present is in a state of intensive development. Today, probability theory finds applications in a large and growing list of areas. It forms the basis of the Mendelian theory of heredity, and hence has played a major part in the development of the science of genetics. Modern theories in physics concerning atomic particles make use of probability models. The spread of an infectious disease through a population is studied in the theory of epidemics, a branch of probability theory. Queuing theory uses probability models to investigate customer waiting times under the provision of various levels and types of service.

Any realistic model of real-world phenomenon must take into account the possibility of randomness. Uncertainties are unavoidable in the design and planning of engineering systems. Therefore, the tools of engineering analysis should include methods and concepts for evaluating the significance of uncertainty on system performance and design. In this regard, the principles of probability offer the mathematical basis for modeling uncertainty and the analysis of its effects on engineering design.

The idea of attaching a number to a set or to an object is familiar to everybody. We may talk about the length of a segment, the area of a triangle, the volume of a ball, or the mass of a physical body, its temperature and so on. All these facts can be expressed in numbers. Probability is also expressed in terms of numbers attached to events. The way of doing this is very much analogous to that of length, area and volume.

Probability theory is one of the most beautiful branches of mathematics. The problems that it can address and the answers that it provides are often strikingly structured and beautiful. At the same time, probability theory is one the most applicable branches of mathematics. It is used as the primary tool for analyzing statistical methodologies.

As in the cases of all parts of mathematics, probability theory is constructed by means of an axiomatic method. In 1933, A. N. Kolmogorov provided an axiomatic basis for probability theory, and it is now the universally accepted model. Nowadays the following approach is accepted in probability theory.

§2. SAMPLE SPACE AND EVENTS

Below we introduce the concept of the probability of an event and then show how these probabilities can be computed in certain situations. As a preliminary, however, we need the concept of the *sample space* and *events* of an experiment. To each experiment corresponds some nonempty set Ω. This set consists of all possible outcomes of the experiment. This Ω is called the *set of all possible outcomes* or *sample space* and a generic element of Ω is denoted by ω. We generally denote a sample space by the capital Greek letter Ω. Let's consider some examples.

Example 1. If the outcome of an experiment consists of the determination of the sex of a newborn child, then

$$\Omega = \{\omega_1, \omega_2\},$$

where the outcome ω_1 means that the child is a girl and ω_2 that it is a boy.

Example 2. If an experiment consists of tossing a coin on a smooth surface then Ω is the same as in Example 1, that is

$$\Omega = \{\omega_1, \omega_2\},$$

where ω_1 corresponds to a head and ω_2 corresponds to a tail.

Example 3. If we toss a symmetrical (or a fair) die on a smooth surface then

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\},$$

where ω_i corresponds to the outcome when on the opened lateral face appear i points, $i = 1, 2, ..., 6$.

Example 4. Now an experiment consists of tossing a coin twice. In this case we take as the set of possible outcomes (i.e. as the sample space) the 4–element set

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\},$$

where the outcomes are the following
$\omega_1 = hh$ is the outcome: heads on the first toss and heads on the second;
$\omega_2 = ht$ is the outcome: heads on the first toss and tails on the second;
$\omega_3 = th$ is the outcome: tails on the first toss and heads on the second;
$\omega_4 = tt$ is the outcome: tails on the first toss and tails on the second.

Example 5. Now two symmetrical dice are thrown on a smooth surface then

$$\Omega = \{\omega_1 = (1,1), \quad \omega_2 = (1,2), \quad \omega_3 = (2,1), ..., \quad \omega_{36} = (6,6)\},$$

where (i, j) is the outcome that the first die equals i and the second die equals j, i. e. Ω is a 36 element set.

Example 6. A card is selected at random from a deck of 52 playing cards, then Ω is 52 element set.

Remark 1. It should be pointed out that the sample space of an experiment can be defined in more than one way. Observers with different conceptions of what could possibly be observed will arrive at different sample spaces. For instance, in Example 2, the sample space Ω might consist of three elements, if we desired to include the possibility that the coin might stand on its edge or rim. Then

$$\Omega = \{\omega_1, \omega_2, \omega_3\},$$

where ω_1, ω_2 are as in Example 2 and the outcome ω_3 represents the possibility of the coin standing on its rim.

There is yet a fourth possibility; the coin might be lost by rolling away when it lands. The sample space Ω would then be

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$$

in which the outcome ω_4 denotes the possibility of loss.

In these examples Ω is finite. However, there exist experiments for which the number of elements of Ω is infinite.

Example 7. Assume we toss a coin until a tail opens. Then

$$\Omega = \{\omega_1, \omega_2, ..., \omega_n, ...\},$$

where ω_i corresponds to the outcome for which a coin rolls exactly i times, i. e. we will have to toss the coin i times until a tail opens.

Example 8. If we choose a point at random in the bounded subset D of n–dimensional Euclidean space $\mathbb{R}^n$ then

$$\Omega = D.$$

In Example 7 Ω is a countable set and in Example 8 Ω is a continuum set.

In general, a sample space is said to be *discrete* if it has finitely many or a countable infinity of elements. If the points (outcomes) of a sample space constitute a continuum, for example, all the points on a line, all the points on a line segment, or all the points in a plane, the sample space is said to be *continuous*.

Definition 1. A subset A of the sample space is called an *event*.
If the outcome of the experiment is contained in A then we say that A has occurred. The notation $\omega \in A(\omega \notin A)$ will mean that ω is (is not) an element of A.

Some examples of events are the following.
In Example 1 if $A = \{\omega_2\}$ then A is the event that the child is a boy.
In Example 2 if $A = \{\omega_1\}$, then A is the event that a head appears on the flip of the coin.
In Example 3 if $A = \{\omega_2, \omega_4, \omega_6\}$ then A is the event when appears even face.
In Example 4 if $A = \{\omega_1, \omega_2, \omega_3\}$ then A is the event that at least one head appears.
In Example 7 if $A = \{\omega_1, \omega_2, ..., \omega_7\}$ then A is the event that the number of the throwing of a coin does not exceed 7.

Example 9. Consider an experiment that consists of counting the number of traffic accidents at a given intersection during a specified time interval. The sample space is the set

$$\Omega = \{0, 1, 2, 3, ...\}.$$

The statement "the number of accidents is less than or equal to seven" describes the event $\{0, 1, ..., 7\}$. The event $A = \{5, 6, 7, ...\}$ occurs if and only if the number of accidents is greater than or equal to 5.
In particular, the sample space Ω is a subset of itself and is thus an event. We call the sample space Ω the *certain event*, since by the method of construction of Ω it will always occur.
For any two events A and B of a sample space Ω we define the new event

$$A \cup B$$

called the *union* of the events A and B, consisting of all outcomes which belong to at least one, either A or B.
Similarly, for any two events A and B,

$$A \cap B$$

is called the *intersection* of A and B and consists of all outcomes which belong to both A and B.

For instance, in Example 1 if $A = \{\omega_1\}, B = \{\omega_2\}$, then $A \cup B = \Omega$, that is $A \cup B$ would be the certain event Ω, and $A \cap B$ does not contain any outcomes and hence cannot occur. To give such an event a name we introduce the notion of the *impossible event* and denote it by $\emptyset$. Thus $\emptyset$ means the event consists of no outcome .

If $A \cap B = \emptyset$ implying that A and B cannot both occur, then A and B are said to be *mutually exclusive*.

For any event A we define the event $\overline{A}$ referred to as the *complement* of A and consists of all outcomes in Ω that are not in A.

In Example 3 if $A = \{\omega_2, \omega_4, \omega_6\}$ then $\overline{A} = \{\omega_1, \omega_3, \omega_5\}$ is the event when appears odd face.

In Example 4 the events $A = \{\omega_1\}$ and $B = \{\omega_4\}$ are mutually exclusive.

For any two events A and B, if all of the outcomes in A are also in B, then we say that A is a subevent of B and write $A \subset B$. If $A \subset B$ and $B \subset A$ then we say that A and B are equal and we write $A = B$.

We can define unions and intersections of more than two events. The *union* of the events $A_1, A_2, ..., A_n$ denoted by

$$\bigcup_{k=1}^{n} A_k$$

is defined to be the event consisting of all outcomes that are in A_i for at least one $i = 1, ..., n$. Similarly, the *intersection* of the events A_i denoted by

$$\bigcap_{k=1}^{n} A_k$$

is defined to be the event consisting of those outcomes that are in all of the events A_i, $i = 1, 2, ..., n$. In other words, the union of the A_i occurs when at least one of the events A_i occurs; while the intersection occurs when all of the events A_i occurs.

Similarly we can define

$$\bigcup_{n=1}^{\infty} A_n \qquad \text{and} \qquad \bigcap_{n=1}^{\infty} A_n$$

for any sequence of events $\{A_n\}_{n=1}^{\infty}$.

The following useful relations between the three basic operations (unions, intersections and complements of events) are known as De Morgan's law.

$$\overline{\bigcup_i A_i} = \bigcap_i \overline{A_i}, \qquad \overline{\bigcap_i A_i} = \bigcup_i \overline{A_i} \tag{1}$$

for any family of events $\{A_i\}$.
In particular, we have

$$\overline{A \cup B} = \overline{A} \cap \overline{B} \quad \text{and} \quad \overline{A \cap B} = \overline{A} \cup \overline{B}$$

Exercise 1. Prove De Morgan's law.
Hint: Relations stating the equality of two sets is proved by the following way: we take a point belonging to the left–hand side of the equality and show that the point belongs to the right–hand side of the equality and vice versa. Proof of such type of assertions I will always omit.

§3. AXIOMS OF PROBABILITY

Let us consider an experiment and let Ω be the sample space of the experiment. A *probability* on Ω is a real function P which is defined on events of Ω and satisfies the following three axioms.

Axiom 1. $P(A) \geq 0$, for any event A.
Axiom 2. $P(\Omega) = 1$.
Axiom 3. For any sequence of mutually exclusive events A_1, A_2, ... (that is, events for which $A_i \cap A_j = \emptyset$ if $i \neq j$)

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n). \tag{2}$$

We call $P(A)$ the probability of event A.
Thus, Axiom 1 states that the probability that outcome of the experiment is contained in A is some nonnegative number. Axiom 2 states that the probability of the certain event is always equal to one. Axiom 3 states that for any sequence of mutually exclusive events $\{A_n\}_{n=1}^{\infty}$ the probability that at least one of these events occurs is equal to the sum of their respective probabilities.

LECTURE 2

PROPERTIES OF PROBABILITY

Example 10. In Example 2, if we assume that a head is equally likely to appear as a tail, then we would have

$$P(\{\omega_1\}) = P(\{\omega_2\}) = \frac{1}{2}.$$

On the other hand, if we had a biased coin and felt that a head was twice as likely to appear as a tail, then we would have

$$P(\{\omega_1\}) = \frac{2}{3} \qquad P(\{\omega_2\}) = \frac{1}{3}.$$

In Example 3, if we supposed that all six outcomes were equally likely to appear, then we would have

$$P(\{\omega_1\}) = P(\{\omega_2\}) = ... = P(\{\omega_6\}) = \frac{1}{6}.$$

From Axiom 3 follows that the probability of getting an even number would equal

$$P(\{\omega_2, \omega_4, \omega_6\}) = P(\{\omega_2\}) + P(\{\omega_4\}) + P(\{\omega_6\}) = \frac{1}{2}.$$

These axioms will now be used to prove the simplest properties concerning probabilities.

Property 1. $P(\emptyset) = 0.$

That is, the impossible event has probability 0 of occurring.

The proof of Property 1 you can find in the Appendix of the present lecture.

It should also be noted that it follows that for any finite number of mutually exclusive events $A_1, ..., A_n$

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k). \tag{3}$$

In particular, for any two mutually exclusive events A and B

$$P(A \cup B) = P(A) + P(B). \tag{4}$$

The proof of (3) you can find in the Appendix of the present lecture.

Therefore Axiom 3 is valid both for finite number of events (see (3) and (4)) and for countable number of events.

Property 2. For any event A

$$P(\overline{A}) = 1 - P(A).$$

Proof of Property 2: We first note that A and $\overline{A}$ are always mutually exclusive and since $A \cup \overline{A} = \Omega$ we have by Axiom 3 that

$$P(\Omega) = P(A \cup \overline{A}) = P(A) + P(\overline{A}) \quad \text{and by Axiom 2} \quad P(A) + P(\overline{A}) = 1.$$

The proof is complete.

As a special case we find that $P(\emptyset) = 1 - P(\Omega) = 0$, since the impossible event is the complement of Ω.

Property 3. For any two events A and B

$$P(B \setminus A) = P(B) - P(A \cap B). \tag{5}$$

Proof: The events $A \cap B$ and $B \cap \overline{A}$ are mutually exclusive, and their union is B. Therefore, by Axiom 3, $P(B) = P(A \cap B) + P(B \cap \overline{A})$, from which (5) follows immediately because $B \setminus A = B \cap \overline{A}$.

Property 4. If $A \subset B$ then

$$P(B \setminus A) = P(B) - P(A).$$

Proof: Property 4 is a corollary of Property 3.

Property 5. If $A \subset B$ then $P(A) \leq P(B)$, that is probability P is nondecreasing function.

Proof of Property 5: As $P(B \setminus A) \geq 0$, then Property 5 implies from Property 4.

Property 6. For any event A

$$P(A) \leq 1.$$

Property 6 immediately follows from both Property 5 where we substitute $B = \Omega$ and from Axiom 2 (i. e. any event A is a subevent of the certain event).

Therefore Axiom 1 and Property 6 state that the probability that the outcome of the experiment is contained in A is some number between 0 and 1, i. e.

$$0 \leq P(A) \leq 1.$$

Property 7. For any two events A and B

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{6}$$

The proof of Property 7 you can find in the Appendix of the present lecture.

Example 11. A card is selected at random from a deck of 52 playing cards. We will win if the card is either a club or a king. What is the probability that we will win?

Solution: Denote by A the event that the card is clubs and by B that it is a king. The desired probability is equal to $P(A \cup B)$. It follows from Property 7 that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

As

$$P(A) = \frac{1}{4}, \quad P(B) = \frac{4}{52} \quad \text{and} \quad P(A \cap B) = \frac{1}{52}$$

we obtain

$$P(A \cup B) = \frac{1}{4} + \frac{4}{52} - \frac{1}{52} = \frac{4}{13}.$$

Property 8 (Inclusion–Exclusion Principle) . For any events A_1, A_2, ..., A_n we have

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k) - \sum_{k<j} P(A_k \cap A_j) + \sum_{k<j<i} P(A_k \cap A_j \cap A_i) - ... + (-1)^{n-1} P\left(\bigcap_{k=1}^{n} A_k\right). \tag{7}$$

In words, formula (7) states that the probability of the union of n events equals the sum of the probabilities of these events taken one at a time minus the sum of the probabilities of these events taken two at a time plus the sum of the probabilities of these events taken three at a time, and so on.

Exercise 2. Prove Property 8.

Hint: We note that (6) is a special case of (7) when $n = 2$. For finishing the proof we have to apply the method of mathematical induction[1].

Property 9. For any two events A and B the inequality

$$P(A \cup B) \leq P(A) + P(B) \tag{8}$$

holds.

The proof follows from (6).

Property 10 (Boole's inequality). For any sequence of events A_1, ..., A_n, ...

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n). \tag{9}$$

The proof of Boole's inequality you can find in the Appendix.

Properties 9 and 10 state that for any events the probability that at least one of these events occurs is less than or equal to the sum of their respective probabilities.

Definition 2. A pair (Ω, P) is called a ***probability space***, where Ω is a sample space on which a probability P (satisfying Axioms 1, 2 and 3) has been defined.

§4. FINITE SAMPLE SPACES

Definition 3. A sample space Ω is called finite if the number of possible outcomes of the experiment is finite, i. e.

$$\Omega = \{\omega_1, \omega_2, ..., \omega_n\}.$$

In other words, a sample space Ω is defined as finite if it is of finite size which is to say that the random experiment under consideration possesses only a finite number of possible outcomes.

---

[1]The principle of mathematical induction states that a proposition $p(n)$ which depends on an integer n is true for $n = 1, 2, ...$ if one shows that (i) it is true for $n = 1$ and (ii) it satisfies the implication: $p(n)$ implies $p(n + 1)$.

Definition 4. A single–member event is an event that contains exactly one outcome. If an event A has as its only member the outcome ω_i, this fact may be expressed in symbols by writing $A = \{\omega_i\}$. Thus $\{\omega_i\}$ is the event that occurs if and only if the random situation being observed has outcome ω_i.

Let Ω be a finite sample space. Let us take numbers p_i so that $p_i \geq 0$, $i = 1, ..., n$ and $\sum\limits_{i=1}^{n} p_i = 1$. We set

$$P(A) = \sum_{i:\omega_i \in A} p_i. \tag{10}$$

Exercise 3. Prove that the function defined by (10) is a probability.

Hint: For that Axioms 1, 2 and 3 must be checked.

Exercise 4. Show that if P is a probability on $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$ then there exist numbers $p_i \geq 0$, $i = 1, ..., n$ and $\sum\limits_{i=1}^{n} p_i = 1$ such that the relation (10) is satisfied for $p_i = P(\{\omega_i\})$.

Hint: There are 2^n possible events on a sample space of finite size n. Let A be an event. We can represent $A = \{\omega_{i_1}, ..., \omega_{i_k}\}$, $k < \infty$. A probability $P(\cdot)$ defined on Ω can be specified by giving its values $P(\{\omega_i\})$ on the single–member events $\{\omega_i\}$ which correspond to the elements of Ω. Its value $P(A)$ on an event A may then be computed by the formula

$$P(A) = \sum_{j=1}^{k} P(\{\omega_{i_j}\}).$$

Example 12. Let Ω be 2 element sample space, i. e. $\Omega = \{\omega_1, \omega_2\}$. Assume that $p_1 = p$ and $p_2 = 1 - p$, where $0 \leq p \leq 1$. This example describes all experiments which have two outcomes. If our experiment consists of tossing a coin and if we assume that a head is as likely to appear as a tail, then we would have $p = 1/2$; on the other hand, if the coin were biased and a head were twice as likely to appear as a tail, then we would have $p = 2/3$ (see also Examples 2 and 10).

§5. SAMPLE SPACES HAVING EQUALLY LIKELY OUTCOMES

In many probability situations in which finite sample spaces arise it may be assumed that all outcomes are equally likely; that is, all outcomes in Ω have equal probabilities of occurring. More precisely, we define the sample space $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$ as having equally likely outcomes if all the single–member events of Ω have equal probabilities, so that

$$P(\{\omega_1\}) = P(\{\omega_2\}) = ... = P(\{\omega_n\}) = p.$$

Now it follows from Axioms 2 and 3 that

$$1 = P(\Omega) = \sum_{i=1}^{n} P(\{\omega_i\}) = p \cdot n$$

what shows that

$$P(\{\omega_i\}) = p = \frac{1}{n} \qquad \text{for any} \quad i = 1, 2, ..., n.$$

It should be clear that each of the single–member events $\{\omega_i\}$ has probability $1/n$, since there are n such events, each of which has equal probability, and the sum of their probabilities must equal 1, the probability of the certain event.

It follows from Axiom 3 that for any event A

$$P(A) = \sum_{i:\omega_i \in A} P(\{\omega_i\}) = \frac{\text{Number of outcomes in } A}{n}.$$

Therefore the calculation of the probability of an event defined on a sample space with equally likely outcomes can be reduced to the calculation of the size of the event. By (10) probability of A is equal to $1/n$, multiplied by the number of outcomes in A. In other words, *the probability of A is equal to the ratio of the size of A to the size of Ω.*

If, for an event A of finite size, we define by $N(A)$ the size of A (the number of outcomes of A), then the foregoing conclusions can be summed up in a formula

$$P(A) = \frac{N(A)}{N(\Omega)} = \frac{\text{size of } A}{\text{size of } \Omega} = \frac{\text{the number of outcomes favorable to } A}{\text{total number of outcomes}}.$$

This is a precise formulation of the classical definition of the probability of an event, first explicitly formulated by Laplace in 1812. For several centuries, the theory of probability was based on the classical definition. This concept is used today to determine probabilistic data and as a working hypothesis. It is important to note, however, that the significance of the numbers $N(\Omega)$ and $N(A)$ is not always clear.

§6. GEOMETRIC PROBABILITIES

Let Ω be the same as in Example 8 i. e. we choose a point at random in the bounded subset D of n–dimensional Euclidean space $\mathbb{R}^n$. Therefore $\Omega = D$ and events are subsets of D. We suppose that volume of D does not equal 0, $V(D) \neq 0$ (V stands for volume). We define $P(A)$ by the following formula

$$P(A) = \frac{V(A)}{V(\Omega)}. \tag{15}$$

In particular, when $n = 2$

$$P(A) = \frac{S(A)}{S(\Omega)}, \tag{16}$$

in which S stands for area.

Exercise 5. Prove that P which is defined by formula (15) is a probability.

Remark 2. You know from school that $0+0 = 0$, i. e. the sum of finite number of zeros is equal to zero. Later we learned that the sum of countable number of zeros is zero, i. e. $\sum\limits_{k=0}^{\infty} 0 = 0$ (see Lemma 2 in the Appendix of the present lecture). But for the sum of uncountable number of events it is not true. In order to be convinced let us consider the following example.

Example 13. Let $\Omega = [0,1] \times [0,1]$.

$$P(A) = \frac{S(A)}{S(\Omega)} = S(A) \qquad (S(\Omega) = 1).$$

Denote by

$$A_y = \{(x,y) \colon \quad y \quad \text{is fixed}\}.$$

$$P(A_y) = 0 \quad \text{for any} \quad y \in [0,1] \quad \text{and} \quad \Omega = \bigcup_{y \in [0,1]} A_y, \quad P(\Omega) = 1.$$

APPENDIX 1:

§7. PROBABILITY AS A CONTINUOUS SET FUNCTION

A sequence $\{A_n,\ n \geq 1\}$ is called an increasing sequence if

$$A_1 \subset A_2 \subset ... \subset A_n \subset A_{n+1} \subset ...$$

and decreasing sequence if

$$A_1 \supset A_2 \supset ... \supset A_n \supset A_{n+1} \supset ....$$

Definition 5. **We will say that a sequence $\{A_n\}$ monotone increasing tends to the event A and will denote it by**

$$\lim_{n\to\infty} \uparrow A_n = A \qquad \text{or} \quad A_n \uparrow A$$

if the following two conditions are satisfied:

i) $\{A_n\}$ **is an increasing sequence of events (i. e. $A_n \subset A_{n+1}$, for any $n = 1, 2, ...$)**

ii) $A = \bigcup\limits_{n=1}^{\infty} A_n$.

If we have an increasing sequence then the limit is naturally determined as the union of all events A_n.

Definition 6. **We will say that a sequence $\{A_n\}$ monotone decreasing tends to the event A and will denote it by**

$$\lim_{n\to\infty} \downarrow A_n = A \qquad \text{or} \quad A_n \downarrow A$$

if the following two conditions are satisfied:

i) $\{A_n\}$ **is an decreasing sequence of events (i. e. $A_n \supset A_{n+1}$, for any natural n)**

ii) $A = \bigcap\limits_{n=1}^{\infty} A_n$.

If we have a decreasing sequence then the limit is naturally determined as the intersection of all events A_n.

Therefore, there exists a limit for any monotone sequence $\{A_n\}$.

In particular, $A_n \downarrow \emptyset$, if $\{A_n, \quad n \geq 1\}$ is a decreasing sequence and

$$\bigcap_{n=1}^{\infty} A_n = \emptyset.$$

Definition 7. **A function P defined on the set of events is said to be continuous at $\emptyset$ if for any sequence $\{A_n\}$ of events such as $A_n \downarrow \emptyset$ implies**

$$P(A_n) \downarrow 0.$$

Theorem 1. *Any probability P is continuous at $\emptyset$.*

Property 11. **If $A_n \downarrow A$ then $P(A_n) \downarrow P(A)$.**

Property 12. **If $A_n \uparrow A$ then $P(A_n) \uparrow P(A)$.**

Properties 11 and 12 state that any probability is continuous under monotone increasing or decreasing sequences of events.

APPENDIX 2:
Proofs of (3), Theorem 1 and Properties 1, 7, 10 — 12

Proof of Property 1: **If we consider a sequence of events $A_1, A_2, ...,$ where $A_1 = \Omega$, $A_k = \emptyset$ for $k > 1$ then, as the events are mutually exclusive and as $\Omega = \bigcup\limits_{n=1}^{\infty} A_n$, we have from Axiom 3 that**

$$P(\Omega) = \sum_{n=1}^{\infty} P(A_n) = P(\Omega) + \sum_{n=2}^{\infty} P(A_n)$$

and by Axiom 2 $P(\Omega) = 1$ we obtain

$$\sum_{n=2}^{\infty} P(\emptyset) = 0$$

implying that

$$P(\emptyset) = 0.$$

Proof of (3): **It follows from Axiom 3 by defining A_i to be the impossible event for all values of i greater than n. Indeed,**

$$P\left(\bigcup_{i=1}^{n} A_i \bigcup \left[\bigcup_{i=n+1}^{\infty} \emptyset\right]\right) = \sum_{i=1}^{n} P(A_i) + \sum_{i=n+1}^{\infty} P(\emptyset).$$

As $P(\emptyset) = 0$ we obtain (3).

Proof of Property 7: **It is not difficult to prove the following identity:**

$$A \cup B = A \cup (B \cap \overline{A}),$$

where A and $B \cap \overline{A}$ are mutually exclusive.
By Axiom 3 we get

$$P(A \cup B) = P(A) + P(B \cap \overline{A}). \tag{11}$$

Because $B \cap \overline{A} = B \setminus A$, by Property 3 we obtain

$$P(B \cap \overline{A}) = P(B) - P(A \cap B). \tag{12}$$

Substituting (12) into (11) we have (6). The Property is proved.

Proof of Property 10: **To prove (9) we represent $\bigcup\limits_{n=1}^{\infty} A_n$ in the form**

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} (A_n \cap B_n),$$

where

$$B_n = \overline{\bigcup_{k=1}^{n-1} A_k} \quad (B_1 = \Omega,\ B_2 = \overline{A_1},\ B_3 = \overline{A_1 \cup A_2} \quad ... \quad).$$

Therefore

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty} (A_n \cap B_n)\right).$$

It is not difficult to verify that the events $\{A_n \cap B_n\}$ are mutually exclusive. Hence we have

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n \cap B_n).$$

Since $P(A_n \cap B_n) \leq P(A_n)$ (by Property 5) the proof is complete.

Proof of Theorem 1: **Let $A_n \downarrow \emptyset$, i. e. $A_{n+1} \subset A_n$ for any n and**

$$\bigcap_{n=1}^{\infty} A_n = \emptyset. \tag{13}$$

We have to prove that $P(A_n) \downarrow \emptyset$.
It is not difficult to verify that for any decreasing sequence $\{A_n\}$ we have

$$A_n = \bigcup_{k=n}^{\infty} (A_k \setminus A_{k+1}) \bigcup \left(\bigcap_{n=1}^{\infty} A_n \right).$$

We must prove that if an outcome belongs to the left–hand side of the identity then the outcome belongs to the right–hand side and vice versa (the proof of this assertion is left to the reader).

It follows from (13) that

$$A_n = \bigcup_{k=n}^{\infty} (A_k \setminus A_{k+1})$$

and $\{A_k \setminus A_{k+1}\}$ are mutually exclusive. Thus

$$P(A_n) = \sum_{k=n}^{\infty} P(A_k \setminus A_{k+1}). \tag{14}$$

Let us write this equation for $n = 1$, we have

$$P(A_1) = \sum_{k=1}^{\infty} P(A_k \setminus A_{k+1}).$$

Since by Axiom 1 and Property 6 $0 \le P(A) \le 1$ we come to conclusion that the series

$$\sum_{k=1}^{\infty} P(A_k \setminus A_{k+1})$$

is convergent. As we know that the remainder $\sum_{k=N}^{\infty} a_k$ of the convergent series $\sum_{k=1}^{\infty} a_k$ tends to zero as $N \to \infty$, therefore by (14) $P(A_n) \downarrow 0$.

Proof of Property 11: $A_n \downarrow A$ **if and only if** $(A_n \setminus A) \downarrow \emptyset$. **Therefore by Theorem 1** $P(A_n \setminus A) \downarrow \emptyset$. **By Property 4 we get** $P(A_n) - P(A) \downarrow \emptyset$. **The proof is complete.**

Proof of Property 12: $A_n \uparrow A$ **if and only if** $(A \setminus A_n) \downarrow \emptyset$. **Therefore by Theorem 1** $P(A \setminus A_n) \downarrow \emptyset$. **By Property 4 we get** $P(A) - P(A_n) \downarrow \emptyset$. **Therefore** $P(A_n) \uparrow P(A)$. **The proof is complete.**

Exercise 6.
For given Ω we define a function μ on subsets of Ω in the following way

$$\mu(A) = \begin{cases} 0 & \textbf{if} \quad A \neq \Omega \\ 1 & \textbf{if} \quad A = \Omega. \end{cases}$$

Is μ a probability or not?

Remark 3. We have supposed that $P(A)$ is defined for all events A of the sample space. Actually, when the sample space is an uncountably infinite set, $P(A)$ is defined for only the so–called measurable events. However, this restriction need not concern us as all events of any practical interest are measurable.

Exercise 7. Give counterexamples to the following assertions:

i) if $P(A) = 0$ then $A = \emptyset$,
ii) if $P(A) = 1$ then $A = \Omega$.

Lemma 1. If each event of finite or infinite sequence $A_1, A_2, ..., A_n, ...$ has probability equal to one (i. e. $P(A_k) = 1$ for any $k = 1, 2, ...$) then

$$P\left(\bigcap_{k=1}^{\infty} A_k\right) = 1.$$

Lemma 2. If each event of finite or infinite sequence $A_1, A_2, ..., A_n, ...$ has probability equal to zero (i. e. $P(A_k) = 0$ for any $k = 1, 2, ...$) then

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = 0.$$

Exercise 8. Prove Lemmas 1 and 2.

LECTURE 3

§8. COMBINATORIAL ANALYSIS

Many of the basic concepts of probability theory, as well as a large number of important problems of applied probability theory, may be considered in the context of finite sample spaces. Therefore, many problems in probability theory require that we count the number of ways that a particular event can occur.

The mathematical theory of counting is formally known as *Combinatorial Analysis*. Let us briefly give the main notions of Combinatorial Analysis.

1. Basic Principle of Counting.

Suppose that two experiments are to be performed. Then if experiment 1 can result in any one of m possible outcomes and if, for each outcome of experiment 1, there are n possible outcomes of experiment 2, then together there are $m \cdot n$ possible outcomes of the two experiments.

Proof: The basic principle can be proved by enumerating all the possible outcomes of the two experiments as follows:

$$\begin{array}{cccc} (1,1), & (1,2), & ..., & (1,n) \\ (2,1), & (2,2), & ..., & (2,n) \\ .... & .... & ..., & .... \\ (m,1), & (m,2), & ..., & (m,n) \end{array}$$

where we say that the outcome is (i,j) if experiment 1 results in its ith possible outcome and experiment 2 then results in the jth of its possible outcomes. Hence, the set of possible outcomes consists of m rows, each row containing n elements, which proves the result.

2. Generalized Basic Principle of Counting.

If k experiments that are to be performed are such that the first one may result in any of n_1 possible outcomes, and if for each of these n_1 possible outcomes there are n_2 possible *outcomes of the second experiment, and if for each of the possible outcomes of the first two experiments there are n_3 possible outcomes of the third experiment, and if ...., then there are a total of $n_1 \cdot n_2 \cdot ... \cdot n_k$ possible outcomes of the k experiments.*

3. Ordered Sequences or Permutations. Briefly, a permutation is an ordered arrangement of objects. In general, if k objects are chosen from a set of n distinct objects, any particular arrangement, or order, of these objects is called a *permutation*. How many different permutations of k objects selected from a set of n distinct objects are possible? There are

$$(n)_k = n \cdot (n-1) \cdot (n-2) \cdot ... \cdot (n-k+1)$$

different arrangements of n objects taken k.

This result can be proved by using the basic principle of counting, since the first object in the permutation can be any of the n objects, the second object in the permutation can then be chosen from the remaining $(n-1)$, the third object in the permutation is then chosen from the remaining $(n-2)$ etc. and the final object in the permutation is chosen from the remaining $n-k+1$. Thus there are $n \cdot (n-1) \cdot ... \cdot (n-k+1)$ possible permutations of n objects taken k at a time.

$(n)_k$ read, permutations of n taken k. In particular

$$(n)_n = n \cdot (n-1) \cdot ... \cdot 2 \cdot 1$$

different permutations of n objects taken n.

Since products of consecutive positive integers arise in many problems relating to permutations or other kinds of special selections, it will be convenient to introduce here the *factorial* notation, where

$$n! = n \cdot (n-1) \cdot ... \cdot 2 \cdot 1 \quad \text{(read, } n \text{ factorial).}$$

In particular, to make various formulae more generally applicable, we let $0! = 1$ by definition.

To express the formula for $(n)_k$ in terms of factorials, we multiply and divide by $(n-k)!$ getting

$$(n)_k = \frac{n!}{(n-k)!}$$

and

$$(n)_n = n!.$$

Example 14. All possible rearrangements of the numbers $A = \{1, 2, 3\}$ (or permutations of 3 taken 3) are

$$123 \qquad 132 \qquad 213 \qquad 231 \qquad 312 \qquad 321.$$

Suppose that from the set $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$ we choose k elements and list them in order. How many ways can we do it? The answer depends on whether we are allowed to duplicate items in the list. If no duplication is allowed, we are sampling *without replacement*. If duplication is allowed, we are sampling *with replacement*. We can think of the problem as that of taking labeled balls from an urn. In the first type of sampling, we are not allowed to put a ball back before choosing the next one, but in the second, we are. In either case, when we are done choosing, we have a list of k balls ordered in the sequence in which they are drawn.

The generalized principle of counting can be used to count the number of different samples possible for a set of n elements. First, suppose that sampling is done with replacement. The first ball can be chosen in any of n ways, the second in any of n ways, etc., so that there are $n \times n \times ... \times n = n^k$ samples. Next, suppose that sampling is done without replacement. There are n choices for the first ball, $n-1$ choices for the second ball, $n-2$ for the third, ..., and $n-k+1$ for the kth. We have just proved the following lemma.

Lemma 3. For a set of size n and a sample of size k, there are n^k different ordered samples with replacement and $(n)_k$ different ordered samples without replacement.

Example 15. How many ways five children be lined up?

Solution: This corresponds to sampling without replacement. According to the formula for permutation, we obtain

$$(5)_5 = 5! = 120.$$

Example 16. Suppose that from ten children, five are to be chosen and lined up. How many different lines are possible?

Solution: From Lemma 3, there are

$$(10)_5 = 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 = 30,240 \qquad \text{different lines.}$$

Example 17. Suppose that a room contains n people. What is the probability that at least two of them have a common birthday?

Solution: Assume that every day of the year is equally likely to be a birthday, disregard leap years, and denote by A the event that there are at least two people with a common birthday. As is sometimes the case, it is easier to find $P(\overline{A})$ than to find $P(A)$. This is because A can happen in many ways, whereas $\overline{A}$ is much simpler. There are 365^n possible outcomes, and $\overline{A}$ can happen in $365 \times 364 \times ... \times (365 - n + 1)$ ways. Thus,

$$P(\overline{A}) = \frac{365 \times 364 \times ... \times (365 - n + 1)}{365^n}$$

and

$$P(A) = 1 - P(\overline{A}) = 1 - \frac{365 \times 364 \times ... \times (365 - n + 1)}{365^n}.$$

The following table exhibits the latter probabilities for various values of n:

Table

n	4	16	23	32	40	56
P(A)	0.016	0.284	0.507	0.753	0.891	0.988

From the Table, we see that if there are only 23 people, the probability of at least one match exceeds 0.5.

We now shift our attention from counting permutations to counting combinations. Here we are no longer interested in ordered samples.

4. Combinations. Consider n element set. How many possible subsets of size k ($k \le n$) can be formed?

The number of subsets of size k that may be formed from the members of a set of size n is

$$\binom{n}{k} = \frac{n!}{(n-k)!\ k!}, \qquad \text{(read, } n \text{ choose } k\text{)} \tag{17}$$

Therefore, $\binom{n}{k}$ represents the number of possible combinations of n objects taken k at a time, or the number of different groups of size k that could be selected from a set of n objects when the order of selection is not considered relevant.

Remark 4. In taking samples sequentially from a discrete sample space, the sampling may be done either *with replacements* or *without replacements*; that is when an element is drawn from the set it is either returned or not returned to the set before another element is drawn. In a set of n elements, the number of ordered samples of size k then is n^k for sampling with replacements, whereas in sampling without replacements, the corresponding number of ordered samples is $(n)_k$.

Now we ask the following question: If k objects are taken from the set of n objects without replacement and disregarding order, how many different samples are possible? From the general principle of counting, the number of ordered samples equals the number of unordered samples multiplied by the number of ways to order each sample. Since the number of ordered samples is $(n)_k$ and since a sample of size k can be ordered in $k!$ ways, the number of unordered samples is (17) and therefore, coincides with $\binom{n}{k}$.

5. Unordered samples with replacements. Suppose that a sample of size k is to be chosen from a set with n elements. In sampling without replacement, no elements may be chosen more than once, so that the k items in the sample will all be different. In sampling with replacement, a member may be chosen more than once, so that not all of the k items in the sample need be different. Indeed it is possible that the same item might be chosen every time, in which case the sample would consist of a single item repeated k times.

Lemma 4. The number of ways to form (unordered) combinations of length k from a set of

n distinct objects, replacements allowed, is equal to

$$\binom{n+k-1}{k}$$

Proof: Each k–combination of a set with n elements when repetition is allowed can be represented by a list of $n-1$ bars and k stars. The $n-1$ bars are used to mark off n different cells, with the ith cell containing a star for each time the ith element of the set occurs in the combination. For example, a 6–combination of a set with four elements is represented with three bars and six stars. Here

$$* * \mid * \mid \mid * * *$$

represents the combination containing exactly two of the first element, one of the second element, none of the third element, and three of the fourth element of the set.
As we have seen, each different list containing $n-1$ bars and k stars corresponds to an k–combination of the set with n elements, when repetition is allowed. The number of such lists is $\binom{n+k-1}{k}$, since each list corresponds to a choice of the k positions to place the k stars from the $n+k-1$ positions that contain k stars and $n-1$ bars.

Let us consider examples.

Example 18. How many people must you ask in order to have a $50:50$ chance of finding someone who shares your birthday?

Solution: Suppose that you ask n people. Let A denote the event that someone's birthday is the same as yours. Again, it is easier to work with $\overline{A}$. The total number of outcomes is 365^n, and the number of ways that $\overline{A}$ can happen is 364^n ways. Thus,

$$P(\overline{A}) = \frac{364^n}{365^n}$$

and

$$P(A) = 1 - P(\overline{A}) = 1 - \frac{364^n}{365^n}.$$

In order for the latter probability to be 0.5, n should be 253.

Example 19. Two balls are drawn from a bowl containing six balls, of which four are white and two are red. Find the probability that

i) both balls will be white;

ii) both balls will be the same color;

iii) at least one of the balls will be white.

Solution: **I.** Let us first consider that the balls are drawn without replacement.

i) Denote by A_1 the event that both balls will be white. By classical definition of probability

$$P(A_1) = \frac{\text{size of } A_1}{\text{size of } \Omega} = \frac{\binom{4}{2}}{\binom{6}{2}} = \frac{2}{5}.$$

ii) Denote by A_2 the event that both balls will be red, and A is the event that both balls will be the same color. We have

$$P(A) = P(A_1 \cup A_2) = P(A_1) + P(A_2) = \frac{2}{5} + \frac{1}{15} = \frac{7}{15}.$$

iii) Let B be the event that at least one of the balls will be white.

$$P(B) = 1 - P(\overline{B}) = 1 - P(A_2) = 1 - \frac{1}{15} = \frac{14}{15}.$$

II. Let us consider that the balls are drawn with replacement.

$$\text{i) } P(A_1) = \frac{4^2}{6^2} = \frac{4}{9}; \quad \text{ii) } P(A) = P(A_1 \cup A_2) = \frac{4}{9} + \frac{2^2}{6^2} = \frac{5}{9};$$

$$iii)\, P(B) = 1 - P(A_2) = 1 - \frac{1}{9} = \frac{8}{9}.$$

Example 20. An ordinary deck of 52 playing cards is randomly divided into 4 piles of 13 card each. Compute the probability that each pile has exactly 1 ace.

Solution: Let us use the classical definition of probability. Denote by A the event that each pile has exactly 1 ace.

$$P(A) = \frac{\text{size of } A}{\text{size of } \Omega}.$$

Ω consists of

$$\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13} = \frac{52!}{(13!)^4}$$

outcomes.

Let us count the number of outcomes in A. The first pile which contains only one ace may be chosen by $4 \cdot \binom{48}{12}$ different ways, the second pile by $3 \cdot \binom{36}{12}$ different ways, the third pile by $2 \cdot \binom{24}{12}$ and the fourth only 1. Therefore

$$P(A) = \frac{4 \cdot \binom{48}{12} \cdot 3 \cdot \binom{36}{12} \cdot 2 \cdot \binom{24}{12}}{\frac{52!}{(13!)^4}} = \frac{39 \cdot 26 \cdot 13}{51 \cdot 50 \cdot 49} = 0.105.$$

Example 21. Suppose that a cookie shop has four different kinds of cookies. How many different ways can six cookies be chosen? Assume that only the type of cookie, and not the individual cookies or the order in which they are chosen, matters.

Solution: The number of ways to choose six cookies is the number of 6–combinations of a set with four elements. From Lemma 4 this equals $\binom{4+6-1}{6} = \binom{9}{6}$. Since

$$\binom{9}{6} = \binom{9}{3} = 84,$$

there are 84 different ways to choose the six cookies.

LECTURE 4

Lemma 4 can also be used to find the number of solutions of certain linear equations where the variables are integers subject to constraints. The outcome of the experiment of distributing the k balls into n urns can be described by a vector $(x_1, x_2, ..., x_n)$, where x_i denotes the number of balls that are distributed into the ith urn. Hence the problem reduces to finding the number of distinct nonnegative integer–valued vectors $(x_1, x_2, ..., x_n)$ such that

$$x_1 + x_2 + ... + x_n = k.$$

Hence from Lemma 4, we obtain the following statement:

There are

$$\binom{n+k-1}{n-1} = \binom{n+k-1}{k}$$

distinct nonnegative integer–valued vectors $(x_1, x_2, ..., x_n)$ *satisfying* $x_1 + x_2 + ... + x_n = k$*,* $x_i \geq 0$*,* $i = 1, ..., n$*.*

This illustrated by the following example.

Example 23. How many solutions does the equation

$$x_1 + x_2 + x_3 = 11$$

have, where x_1, x_2 and x_3 are nonnegative integers?

Solution: To count the number of solutions, we note that a solution corresponds to a way of selecting 11 items from a set with three elements, so that x_1 items of type one, x_2 items of type two, and x_3 items of type three are chosen. Hence, the number of solutions is equal to the number of 11–combinations with repetition allowed from a set with three elements. From Lemma 4 it follows that there are

$$\binom{3+11-1}{11} = \binom{13}{11} = \binom{13}{2} = 78$$

solutions.

To obtain the number of positive solutions, note that the number of positive solutions of

$$x_1 + x_2 + ... + x_n = k$$

is the same as the number of nonnegative solutions of

$$y_1 + y_2 + ... + y_n = k - n$$

(seen by letting $y_i = x_i - 1$). Hence from Lemma 4, we obtain the following statement:
There are

$$\binom{k-1}{n-1}$$

distinct positive integer–valued vectors $(x_1, x_2, ..., x_n)$ *satisfying* $x_1 + x_2 + ... + x_n = k,\ x_i > 0$, $i = 1, ..., n$.

Therefore the number of positive solutions in Example 23 is

$$\binom{10}{2} = 45.$$

Remark 5. The sample is said to have been chosen *at random* and is called a *random sample* if all possible ordered sequences are equally probable. We showed that, under random sampling without replacement, the $\binom{n}{k}$ possible unordered samples are equally likely. However, for random sampling with replacement unordered samples are not equally probable. In general, the probability of an unordered sample depends upon how many repeated elements it contains and how often they appear.

For example, suppose that three digits are chosen at random with replacement from $0, 1, 2, ..., 9$. There are $10^3 = 1,000$ equally probable ordered outcomes 000, 001, ...999. The three digits $0, 1, 2$ can appear in $3!$ different arrangements, and so the unordered outcome $\{0, 1, 2\}$ has probability 0.006. However the three digits $0, 0, 1$ can be arranged in only 3 ways, and the three digits $0, 0, 0$ can be arranged in only 1 way. Hence the unordered outcomes $\{0, 0, 1\}$ and $\{0, 0, 0\}$ have probabilities 0.003 and 0.001 respectively.

In general, suppose that the ith member of the population occurs r_i times in the sample $(i = 1, 2, ..., k)$ where $\sum_{i=1}^{k} r_i = n$. The number of arrangements or permutations of the n elements in the sample is

$$\frac{n!}{r_1!\, r_2! \,....r_k!} \tag{18}$$

§9. CONDITIONAL PROBABILITY

In this section we ask and answer the following question. Suppose we assign a probability to a sample space and then learn that an event A has occurred. How should we change the probabilities of the remaining events? We call the new probability for an event B the *conditional probability of* B *given* A and denote it by $P(B/A)$.

The probability of an event may depend on the occurrence (or nonoccurrence) of another event. If this dependence exists, the associated probability is a conditional probability. In the sample space Ω, the conditional probability $P(B/A)$, means the likelihood of realizing an outcome in B assuming that it belongs to A. In other words, we are interested in the event B within the reconstituted sample space A. Hence, with the appropriate normalization, we obtain the conditional probability of B given A.

Example 24. A die is rolled. Let B be the event "a six turns up". Let A be the event "a number greater than 4 turns up". Before the experiment $P(B) = 1/6$. Now we are told that the event A has occurred. This leaves only two possible outcomes: 5 and 6. In the absence of any other information, we would still regard these outcomes to be equally likely, so the probability of B becomes $1/2$ making, $P(B/A) = 1/2$.

Definition 8. Let A and B be two events on a sample space Ω, on the subsets of which is defined a probability $P(\cdot)$. The conditional probability of the event B, given the event A, denoted by

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \qquad \text{if} \quad P(A) \neq 0 \tag{19}$$

and if $P(A) = 0$, then $P(B/A)$ is undefined.

By multiplying both sides of equation (19) by $P(A)$ we obtain

$$P(A \cap B) = P(A) \cdot P(B/A) = P(B) \cdot P(A/B). \tag{20}$$

This formula is often useful as a tool to enable us in computing the desired probabilities more easily.

Example 25. Suppose that an urn contains 8 red and 4 white balls. We draw 2 balls from the urn without replacement. What is the probability that both drawn balls are red (event A)?

Solution: Let A_1 and A_2 denote, respectively the events that the first and second ball drawn is red. By (20)

$$P(A) = P(A_1 \cap A_2) = P(A_1) \cdot P(A_2 \,/ A_1\,).$$

It is obvious that $P(A_1) = \dfrac{8}{12}$.

Now given that the first ball selected is red, there are 7 remaining red balls and 4 white balls and so $P(A_2 \,/ A_1\,) = \dfrac{7}{11}$. The desired probability is

$$P(A) = \frac{8}{12} \cdot \frac{7}{11} = \frac{14}{33}.$$

Of course, this probability could also have been computed by the classical definition

$$P(A) = \frac{\binom{8}{2}}{\binom{12}{2}}.$$

A generalization of equation (20) is sometimes referred to as the *multiplication rule*

$$P\left(\bigcap_{k=1}^{n} A_k\right) = P(A_1) \cdot P(A_2 \,/ A_1\,) \cdot P(A_3 \,/ (A_1 \cap A_2)\,) \cdot ... \cdot P(A_n \,/ (A_1 \cap A_2 \cap ... \cap A_{n-1})\,).$$

To prove the multiplication rule apply the definition of conditional probability to its right–hand side. This gives

$$P(A_1) \cdot \frac{P(A_1 \cap A_2)}{P(A_1)} \cdot \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \cdot ... \cdot \frac{P(\bigcap_{k=1}^{n} A_k)}{P(\bigcap_{k=1}^{n-1} A_k)}$$

and reducing we get $P\left(\bigcap_{k=1}^{n} A_k\right)$.

§10. INDEPENDENCE AND DEPENDENCE

The notions of independent and dependent events play a central role in probability theory. If the events A and B have the property that the conditional probability of B, given A, is equal to the unconditional probability of B, one intuitively feels that event B

is statistically independent of A, in the sense that the probability of B having occurred is not affected by the knowledge that A has occurred.
Since $P(B/A) = \dfrac{P(A\cap B)}{P(A)}$ we see that B is independent of A if

$$P(A\cap B) = P(A)\cdot P(B). \tag{21}$$

As equation (21) is symmetric in A and B, it shows that whenever B is independent of A, A is also independent of B. We thus have the following definition

Definition 9. Two events A and B are said to be independent if equation (21) holds. Two events A and B that are not independent are said to be dependent.

Example 26. Suppose that we toss 2 dice and suppose that each of the 36 possible outcomes is equally likely to occur, and hence has probability $\frac{1}{36}$. Assume further that we observe that the first die is a 4. What is the conditional probability that the sum of the 2 dice equal 8 (event B) given that the first die equals 4 (event A)?

Solution: The event A consists of 6 outcomes

$$\{(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)\}.$$

The event B consists of 5 outcomes

$$\{(2,6),(3,5),(4,4),(5,3),(6,2)\}.$$

The event $A\cap B$ consists of 1 outcome $\{(4,4)\}$, Ω is 36 element set. We obtained

$$P(A) = \frac{1}{6} \qquad P(B) = \frac{5}{36} \qquad P(A\cap B) = \frac{1}{36}$$

and

$$P(B/A) = \frac{P(A\cap B)}{P(A)} = \frac{\frac{1}{36}}{\frac{6}{36}} = \frac{1}{6}.$$

Therefore $P(A\cap B) \neq P(A)\cdot P(B)$ and the events A and B are not independent.

Example 27. A card is selected at random from an ordinary deck of 52 playing cards. If A is the event that the card is an ace and B is the event that it is a club, then A and B are independent. This follows because

$$P(A \cap B) = \frac{1}{52}, \quad P(A) = \frac{4}{52} \quad \text{and} \quad P(B) = \frac{13}{52}.$$

PROPERTIES OF INDEPENDENT EVENTS.

Property 1. If $P(A) \neq 0$ then A and B are independent if and only if

$$P(B\,/A) = P(B).$$

Property 2. If A and B are independent then so are A and $\overline{B}$.

So if A and B are independent then

$$A \text{ and } \overline{B}; \qquad \overline{A} \text{ and } B; \qquad \overline{A} \text{ and } \overline{B}$$

are also independent.

Property 3. If A is independent of B_i, $i = 1, 2$ and $B_1 \cap B_2 = \emptyset$ then A is independent of $B_1 \cup B_2$.

Let us construct an example of three events A_1, A_2, A_3 such that

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2), \qquad P(A_1 \cap A_3) = P(A_1) \cdot P(A_3), \qquad P(A_2 \cap A_3) = P(A_2) \cdot P(A_3),$$

(i. e. A_1, A_2, A_3 are pairwise independent) but

$$P(A_1 \cap A_2 \cap A_3) \neq P(A_1) \cdot P(A_2) \cdot P(A_3).$$

The proofs of Properties 1- 3 we can find in Appendix 1 of the present lecture.

Example 28. Two symmetrical dice are thrown. Let A_1 denote the event that the first die equals 4 and let A_2 be the event that the second die equals 3. A_3 denote the event that the sum of the dice is 7.

We have

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{6},$$

and

$$P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = \frac{1}{36}.$$

Therefore A_1, A_2, A_3 are pairwise independent (compare with Example 26).
It is not difficult to verify that

$$P(A_1 \cap A_2 \cap A_3) = P(A_1 \cap A_2).$$

Therefore, the concept of independence becomes more complex for the case of more than two events. We are thus led to the following definition.

Definition 10. The events $A_1, ..., A_n$ are said to be independent if, for any k $(1 \leq k \leq n)$ and for every subset $A_{i_1}, ..., A_{i_k}$ of these events

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j}).$$

APPENDIX 1.

Proofs of Properties 1 — 3.

Proof of the Property 1: **Let us prove the necessity. Let A and B be independent, therefore $P(A \cap B) = P(A) \cdot P(B)$. From here**

$$P(B\,/A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A) \cdot P(B)}{P(A)} = P(B).$$

Now we suppose that $P(B\,/A) = P(B)$. Therefore

$$P(A \cap B) = P(A) \cdot P(B\,/A) = P(A) \cdot P(B),$$

i. e. A and B are independent.

Proof of the Property 2: **As A and B are independent then $P(A \cap B) = P(A) \cdot P(B)$. It is not difficult to verify that**

$$A \cap \overline{B} = A \setminus (A \cap B).$$

Hence

$$P(A \cap \overline{B}) = P\left(A \setminus (A \cap B)\right) = P(A) - P(A \cap B) = P(A) - P(A) \cdot P(B) = P(A)[1 - P(B)] = P(A) \cdot P(\overline{B}).$$

The result is proved.

Proof of Property 3: **We have $P(A \cap B_i) = P(A) \cdot P(B_i), i = 1, 2$. Therefore**

$$\begin{aligned} &P(A \cap (B_1 \cup B_2)) = P((A \cap B_1) \cup (A \cap B_2)) = P(A \cap B_1) + P(A \cap B_2) = \\ &= P(A) \cdot P(B_1) + P(A) \cdot P(B_2) = P(A) \cdot [P(B_1) + P(B_2)] = P(A) \cdot P(B_1 \cup B_2). \end{aligned}$$

The proof is complete.

APPENDIX 2.

$P(\cdot\,/B\,)$ IS A PROBABILITY

Conditional Probabilities satisfy all of the properties of ordinary probabilities. This is proved by Theorem 2 which shows that $P(\cdot\,/B\,)$ satisfies the three axioms of a probability.

Theorem 2. *Conditional probability $P(A\,/B\,)$ as a function of event A satisfies the following conditions:*

a) $P(A\,/B\,) \geq 0$ *for any* A*;*
b) $P(\Omega\,/B\,) = 1$*;*
c) If events A_i are mutually exclusive, then

$$P\left(\bigcup_{i=1}^{\infty} A_i \Big/ B\right) = \sum_{i=1}^{\infty} P(A_i\,/B\,).$$

Proof: **Condition a) is obvious. Condition b) follows because**

$$P(\Omega\,/B\,) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

Condition c) follows since

$$P\left(\bigcup_{i=1}^{\infty} A_i \Big/ B\right) = \frac{P\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \bigcap B\right)}{P(B)} = \frac{P\left(\bigcup_{i=1}^{\infty} (A_i \bigcap B)\right)}{P(B)} = \frac{\sum_{i=1}^{\infty} P(A_i \bigcap B)}{P(B)} = \sum_{i=1}^{\infty} P(A_i\,/B\,),$$

where the next–to–last equality follows because $A_i \cap A_j = \emptyset$ implies that $(A_i \cap B) \cap (A_j \cap B) = \emptyset$. The proof is complete.

If we define $P_1(A) = P(A\,/B\,)$ (event B is fixed and $P(B) \neq 0$), then it follows from Theorem 2 that $P_1(\cdot)$ may be regarded as a probability function on the events of the sample space Ω. Hence all of the properties proved for probabilities apply to it.

L E C T U R E 5

§11. TOTAL PROBABILITY AND BAYES' FORMULAE

Sometimes the probability of an event A cannot be determined directly. However, its occurrence is always accompanied by the occurrence of other events B_i, $i \geq 1$ such that the probability of A will depend on which of the events B_i has occurred. In such a case the probability of A will be an expected probability (that is, the average probability weighted by those of B_i). Such problems require the *Theorem of Total Probability.*

Let A be a subevent of $\bigcup_{n\geq 1} B_n$ (i. e. $A \subset \bigcup_{n\geq 1} B_n$), $\{B_n\}$ be mutually exclusive and $P(B_n) \neq 0$ for any n. Then

$$P(A) = \sum_{n=1}^{\infty} P(B_n)\, P(A/B_n). \tag{22}$$

We will call formula (22) by *formula of Complete or Total Probability.*

Example 29. Urn I contains 6 white and 4 black balls. Urn II contains 5 white and 2 black balls. From urn I one ball is transferred to urn II. Then 2 balls are drawn without replacement from urn II. What is the probability that the two balls are white?

Solution: Denote by B_1 the event that from urn I a white ball is transferred to urn II, and by B_2 the event that from urn I is a black ball transferred to urn II. Denote by A the event that from urn II are selected two white balls.

By formula (22)

$$P(A) = P(B_1)\, P(A/B_1) + P(B_2)\, P(A/B_2).$$

Since

$$P(B_1) = \frac{3}{5}, \quad P(B_2) = \frac{2}{5}, \quad P(A/B_1) = \frac{15}{28}, \quad P(A/B_2) = \frac{5}{14}$$

we obtain

$$P(A) = \frac{13}{28}.$$

Thus, for any event A one may express the unconditional probability $P(A)$ of A in terms of the conditional probabilities $P(A/B_1), ..., P(A/B_n)...$ and the unconditional probabilities $P(B_1), ..., P(B_n)....$

There is an interesting consequence of the formula of complete probability. Suppose that all conditions of the previous formula are satisfied. Then the following formula

$$P(B_i\,/A\,) = \frac{P(B_i)\cdot P(A\,/B_i\,)}{\sum\limits_{n=1}^{\infty} P(B_n)\,P(A\,/B_n\,)} \qquad i = 1, 2, ... \tag{23}$$

is known as *Bayes' formula*.

If we think of the events B_n as being possible "hypotheses" about some subject matter, then Bayes' formula may be interpreted as showing us how opinions about these hypotheses held before the experiment [that is, the $P(B_n)$] should be modified by the evidence of the experiment.

Let us prove (23). Applying (20) to the events A and B_i, we have

$$P(B_i)\cdot P(A\,/B_i\,) = P(A)\cdot P(B_i\,/A\,).$$

Therefore we obtain the desired probability

$$P(B_i\,/A\,) = \frac{P(B_i)\cdot P(A\,/B_i\,)}{P(A)}. \tag{24}$$

Using the total probability formula, (24) becomes (23).

Example 30. Suppose that there are n balls in an urn. Every ball can be either white or black. The number of white and black balls in the urn is unknown. Let our aim be to find out that number. Determine the following hypotheses:

Denote by B_i the event that the urn consists of exactly i white balls (from n balls), $i = 0, 1, 2, ..., n$.

As we do not have any additional information about the contents of the urn, therefore all hypotheses are equally likely, that is

$$P(B_i) = \frac{1}{n+1}, \quad \text{for any} \quad i = 0, 1, ..., n.$$

Suppose that we have selected a ball and it is white (the event A). It is obvious that we have to modify the probabilities of B_i. For example

$$P(B_0\,/A\,) = 0.$$

What are the remained probabilities equal to? Let us use the Bayes' formula.

$$P(B_i\,/A) = \frac{P(B_i) \cdot P(A\,/B_i\,)}{\sum\limits_{k=0}^{n} P(B_k)\,P(A\,/B_k\,)} \qquad i = 0, 1, ...n.$$

Since

$$P(B_i) = \frac{1}{n+1}, \qquad P(A\,/B_i\,) = \frac{i}{n}$$

we get

$$P(B_i\,/A) = \frac{\frac{i}{n}}{\sum\limits_{k=0}^{n} \frac{k}{n}} = \frac{2i}{n(n+1)}, \qquad i = 0, 1, ...n.$$

In particular, for $n = 3$ we obtain

$$P(B_0\,/A) = 0, \quad P(B_1\,/A) = \frac{1}{6}, \quad P(B_2\,/A) = \frac{1}{3}, \quad P(B_3\,/A) = \frac{1}{2}.$$

§12. INDEPENDENT TRIALS.

It is sometimes the case that the probability experiment under consideration consists of performing a sequence of subexperiments. For instance, if the experiment consists of tossing a coin n times, we may think of each toss as being a subexperiment. In many cases it is reasonable to assume that the outcomes of any group of the subexperiments have no effect on the probabilities of the outcomes of the other subexperiments. If such is the case, we say that the subexperiments are independent.

If each subexperiment is identical — that is, if each subexperiment has the same sample space and the same probability function on its events — then the subexperiments are called *Trials*.

Many problems in Probability Theory involve independent repeated trials of an experiment whose outcomes have been classified in two categories, called "successes" (the event A) and "failures" (the event $\overline{A}$). The probability of the event A is usually denoted by p $\big(P(A) = p\big)$ and therefore $P(\overline{A}) = 1 - p$ where $0 \le p \le 1$.
Such an experiment is called a Bernoulli trial.
Consider now n independent repeated Bernoulli trials, in which the word "repeated" is meant to indicate that the probabilities of success $\big(P(A) = p\big)$ and failure $\big(P(\overline{A}) = 1 - p\big)$

remain the same throughout the trials. The sample space Ω of n independent repeated Bernoulli trials contains 2^n outcomes.

Therefore, in the problems we study in this section, we shall always make the following assumptions:

1. *There are only two possible outcomes for each trial* (arbitrarily called "success" and "failure", without inferring that a success is necessary desirable).

2. *The probability of a success is the same for each trial.*

3. *There are n trials, where n is a constant.*

4. *The n trials are independent.*

If the assumptions cannot be met, the theory we shall develop here does not apply. Frequently, the only fact about outcome of a succession of n Bernoulli trials in which we are interested is the *number of successes*.

We now compute the probability that the number of successes will be k, for any integer k from $0, 1, 2, ..., n$. The event "k successes in n trials" can happen in as many ways as k letters A may be distributed among n places; this is the same as the number of subsets of size k that may be formed from a set containing n members. Consequently, there are $\binom{n}{k}$ outcomes containing exactly k successes and $n-k$ failures (in which $k = 0, 1, ..., n$) is given by

$$P_n(k) = \binom{n}{k} p^k (1-p)^{n-k}. \tag{25}$$

The law expressed by (25) is called the *binomial law* because of the role the quantities in (25) play in the binomial theorem, which states that

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

for any real a and b. Taking $a = p$ and $b = 1-p$, it follows immediately that

$$1 = (p + (1-p))^n = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{n} P_n(k).$$

Letting $a = b = 1$, one finds that the sum of all binomial coefficients is 2^n:

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n.$$

Letting $b = 1$ and $a = -1$, we obtain

$$\sum_{k=0}^{n}(-1)^k \binom{n}{k} = 0.$$

However, (25) represents the solution to a probability problem that does not involve equally likely outcomes.

Let us compute the probability that at least one success occurs in n trials;

In order to determine the probability of at least 1 success in n trials, it is easiest to compute first the probability of the complementary event, that of no successes in n trials. We have

$$P(\text{at least 1 success}) = 1 - P_n(0) = 1 - (1-p)^n.$$

Example 31. Four fair coins are flipped. If the outcomes are assumed independent, what is the probability that two heads and two tails are obtained?

Solution: We have four repeated trials $n = 4$ with parameters $p = 0.5$ and $k = 2$. Hence, by (25) we get

$$P_4(2) = \binom{4}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = \frac{3}{8}.$$

APPENDIX 1.

Proof of TOTAL PROBABILITY formula.

Since A is a subevent of the union of B_n, then

$$A = \bigcup_{n=1}^{\infty} (A \cap B_n),$$

where $\{A \cap B_n\}$ are mutually exclusive because $\{B_n\}$ are mutually exclusive. Therefore by Axiom 3

$$P(A) = \sum_{n=1}^{\infty} P(A \cap B_n)$$

and by formula (20) $P(A \cap B_n) = P(B_n) \cdot P(A/B_n)$. The proof is complete.

APPENDIX 2:

PROBLEM. *For any two events A and B*

$$|P(A \cap B) - P(A)\,P(B)| \leq \frac{1}{4}. \tag{A1}$$

Solution: **We start with the following proposition**

Proposition 1. *For any event A*

$$P(A)\,P(\overline{A}) \leq \frac{1}{4}, \tag{A2}$$

where $\overline{A}$ is the complement of A.

Letting $P(A) = x$ and $P(\overline{A}) = 1 - x$ we have to prove that a function

$$f(x) = x\,(1 - x), \qquad x \in [0, 1]$$

has a maximum at the point $x = \frac{1}{2}$. The proof is obvious.
For any two events A_1 and A_2

$$P(A_1) = P(A_1 \cap A_2) + P(A_1 \cap \overline{A}_2) \tag{A3}$$

The proof immediately follows from the equality

$$A_1 = (A_1 \cap A_2) \bigcup (A_1 \cap \overline{A_2}).$$

Substituting $A_1 = B$ and $A_2 = A$ into the formula (A3) we get

$$P(B) = P(A \cap B) + P(\overline{A} \cap B) \leq P(A \cap B) + P(\overline{A}) \tag{A4}$$

Above we also used that P is a monotone function, that is for any $A_1 \subset A_2$ implies

$$P(A_1) \leq P(A_2). \tag{A5}$$

Since $\overline{A} \cap B \subset \overline{A}$ which implies that $P(\overline{A} \cap B) \leq P(\overline{A})$.
Multiplying both sides (A4) by $P(A)$ we get

$$P(A)\,P(B) \leq P(A)\,P(A \cap B) + P(A)\,P(\overline{A}) \leq P(A \cap B) + P(A)\,P(\overline{A})$$

here we used that $P(A) \leq 1$.
Therefore we obtain the inequality

$$P(A)\,P(B) - P(A \cap B) \leq P(A)\,P(\overline{A}).$$

It follows from (A2) that

$$P(A)\,P(B) - P(A \cap B) \leq \frac{1}{4}. \tag{A6}$$

Using Property (A5) of Probability we have

$$P(B) \geq P(A \cap B) \quad \textbf{and} \quad P(A) \geq P(A \cap B)$$

which imply that

$$P(A)\,P(B) \geq P(A \cap B)\,P(A \cap B).$$

Therefore

$$P(A)\,P(B) - P(A \cap B) \geq P(A \cap B)\,P(A \cap B) - P(A \cap B) = -P(A \cap B)\,P(\overline{A \cap B}) \geq -\frac{1}{4},$$

where the above equality is obtained by noting

$$P(A \cap B) - 1 = -P(\overline{A \cap B})$$

and (A2).
Thus we obtain two inequalities (compare with (A6))

$$P(A)\,P(B) - P(A \cap B) \leq \frac{1}{4},$$

$$P(A)\,P(B) - P(A \cap B) \geq -\frac{1}{4}.$$

Now (A1) follows from these two inequalities. The proof is complete.

APPENDIX 3:

PROBLEM. **Let A and B be mutually exclusive events of an experiment. Then, when independent trials of this experiment are performed, the event A will occur before the event B with probability**

$$\frac{P(A)}{P(A)+P(B)}.$$

Solution: **If we let C_n denote the event that no A or B appears on the first $n-1$ trials and A appears on the nth trial, then the desired probability p is**

$$p = P(\bigcup_{n=1}^{\infty} C_n) = \sum_{n=1}^{\infty} P(C_n).$$

Now, since $P(A$ on any trial$) = P(A)$ and $P(B$ on any trial$) = P(B)$, we obtain, by the independence of trials

$$P(C_n) = [1-(P(A)+P(B))]^{n-1}\, P(A),$$

and thus

$$p = P(A)\sum_{n=1}^{\infty}[1-(P(A)+P(B))]^{n-1} =$$

(using the formula of the sum of terms of geometric progression; the first term equals 1)

$$= P(A)\,\frac{1}{1-(1-(P(A)+P(B)))} = \frac{P(A)}{P(A)+P(B)}.$$

L E C T U R E 6

Example 32. A fair die is rolled five times. We shall find the probability that "six" will show twice.

Solution: In the single roll of a die $A = \{\text{six}\}$ is an event with probability $1/6$. Setting $n = 5,\ k = 2,\ p = P(A) = 1/6$ in (25), we obtain

$$P_5(2) = \binom{5}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3 = \frac{625}{3,888} = 0.160751.$$

Example 33. A pair of fair dice is rolled four times. We shall find the probability that "seven" will not show at all.

Solution: The event $A = \{$ the sum of the dice equals $7\}$ consists of the six favorable outcomes

$$\{(3,4),(4,3),(5,2),(2,5),(6,1),(1,6)\}.$$

Therefore, $P(A) = p = 1/6$. With $n = 4$ and $k = 0$, (25) yields

$$P_4(0) = \left(\frac{5}{6}\right)^4.$$

Example 34. We place at random n points in the interval $(0,T)$. What is the probability that k of these points are in the interval $(t_1,t_2),\ t_1 > 0,\ t_2 < T$?

Solution: This problem can be considered as a problem in repeated trials. The experiment is the placing of a single point in the interval $(0,T)$. In this experiment, $A = \{$ the point is in the interval $(t_1,t_2)\}$ is an event with probability

$$p = P(A) = \frac{t_2 - t_1}{T}.$$

The event $\{A$ occurs k times $\}$ means that k of the n points are in the interval (t_1,t_2). Hence, by (25)

$$P_n(k) = \binom{n}{k}\left(\frac{t_2 - t_1}{T}\right)^k\left(1 - \frac{t_2 - t_1}{T}\right)^{n-k}.$$

Exercise 9 (The behavior of the binomial probabilities). Show, that as k goes from 0 to n, probabilities $P_n(k)$ increase monotonically, then decrease monotonically, reaching their largest value when k satisfying the inequalities

$$n \cdot p - (1-p) \leq k \leq n \cdot p + p.$$

§13. GENERALIZED INDEPENDENT TRIALS

In the foregoing we considered independent trials of a random experiment with just two possible outcomes. It is natural to consider next the independent trials of an experiment with several possible outcomes, say r possible outcomes, i. e. in each trial we have A_1, A_2, ..., A_r possible events. We assume that we know nonnegative numbers p_1, p_2, ..., p_r, whose sum is 1 such that at each trial p_k represents the probability that A_k will be the outcome of that trial $(P(A_k) = p_k,\ k = 1, ..., r$ and $\sum\limits_{k=1}^{r} p_k = 1)$. For Bernoulli repeated trials $r = 2$, $A_1 = A$ (success) and $A_2 = \overline{A}$ (failure) and $p_1 + p_2 = p + 1 - p = 1$. Therefore independent repeated Bernoulli trials is a particular case for $r = 2$. Corresponding to the binomial law, we have the multinomial law: the probability that in n trials the event A_1 will occur k_1 times, the event A_2 will occur k_2 times, ..., the event A_r will occur k_r times, for any nonnegative integers k_j satisfying the condition $k_1 + k_2 + ... + k_r = n$, is given by

$$P_n(k_1, k_2, ..., k_r) = \frac{n!}{k_1!\, k_2! \cdot ... \cdot k_r!} \quad p_1^{k_1} \cdot p_2^{k_2} \cdot ... \cdot p_r^{k_r}. \tag{26}$$

To prove (26) one must note only that the number of outcomes in Ω, which contain k_1 A_1's, k_2 A_2's, ..., k_r A_r's, is equal to the number of ways a set of size n can be partitioned into r subsets of sizes k_1, k_2, ..., k_r respectively, which is equal to

$$\frac{n!}{k_1!\, k_2! \cdot ... \cdot k_r!}.$$

Each of these outcomes has probability $p_1^{k_1} \cdot p_2^{k_2} \cdot ... \cdot p_r^{k_r}$. Consequently, (26) is proved. The name, "multinomial law" derives from the role played by the expression given in the multinomial theorem

$$(a_1 + ... + a_r)^n = \sum_{\substack{0 \leq k_i \leq n \\ i=1,...,r \\ k_1+...+k_r=n}} \frac{n!}{k_1!\, k_2! \cdot ... \cdot k_r!} \quad a_1^{k_1} \cdot a_2^{k_2} \cdot ... \cdot a_r^{k_r}.$$

Example 35. A bowl contains 10 white, 6 black and 4 red balls. 7 balls are drawn with replacement. What is the probability that

i) exactly 4 white balls were selected;

ii) 4 white, 2 black and 1 red balls were selected.

Solution: i) The required probability, based on the binomial law, is

$$P_7(4) = \binom{7}{4}(0.5)^7;$$

ii) by the multinomial law:

$$n = 7,\ r = 3,\ k_1 = 4,\ k_2 = 2,\ k_3 = 1,\ p_1 = .5,\ p_2 = .3,\ p_3 = .2.$$

Example 36. 10 fair dice are thrown on the smooth surface. What is the probability that

i) exactly four "6" were appeared;

ii) four "6", three "5" and three "4" were appeared.

Solution: i) By (25) we obtain the required probability (the binomial law)

$$P_{10}(4) = \binom{10}{4}\ (1/6)^4\ (5/6)^6;$$

ii) by the multinomial law:

$$n = 10,\ r = 4,\ k_1 = 4,\ k_2 = 3,\ k_3 = 3,\ k_4 = 0,\ p_1 = \frac{1}{6},\ p_2 = \frac{1}{6},\ p_3 = \frac{1}{6},\ p_4 = 0.5.$$

Example 37. An urn contains five white and ten black balls. Eight times in succession a ball is drawn out but it is replaced before the next drawing takes place. What is the probability that the balls drawn were white on two occasions and black on six?

Solution: Since the balls are replaced before the next drawing takes place the condition of the urn is always the same just before every trial, and therefore the chance of drawing a white or a black ball is the same for each of the trials. In other words, the trials are independent. The probability of drawing a white ball is 1/3 and the probability of drawing a black ball 2/3. Hence the probability of drawing exactly two white and six black in eight trials is

$$P_8(2) = \binom{8}{2} \cdot \left(\frac{1}{3}\right)^2 \cdot \left(\frac{2}{3}\right)^6 = \frac{1,792}{6,561} = 0.273129.$$

Example 38. An urn contains five white and ten black balls. Eight of these are drawn out and placed in another urn. What is the probability that the latter then contains two white and six black balls?

Solution: This example resembles the former one that it might be very simply stated as, What is the probability of drawing exactly two white balls in eight trials? It differs in that the trials are not independent; that is, the probability of drawing a white ball on the first attempt is $\frac{5}{15}$, whereas the probability of drawing a white ball on the second attempt is either $\frac{4}{14}$ or $\frac{5}{14}$ according as the first ball was white or black. We can calculate this probability using classical definition:

$$\frac{\binom{5}{2} \cdot \binom{10}{6}}{\binom{15}{8}}$$

which we can represent in the form

$$\binom{8}{2} \cdot \frac{5}{15}\,\frac{4}{14}\,\frac{10}{13}\,\frac{9}{12}\,\frac{8}{11}\,\frac{7}{10}\,\frac{6}{9}\,\frac{5}{8} = \frac{140}{429} = 0.3263403.$$

LECTURE 7

Example 39. Suppose that 20% of all copies of a particular textbook fail a certain binding strength test. Let $\eta(\omega)$ denote the number among 15 randomly selected copies that fail the test. What is the probability that

a) at most eight fail the test;

b) exactly eight fail;

c) at least eight fail;

d) between 4 and 7, inclusive, fail.

Solution:

a) The probability that at most eight fail the test is

$$P(\eta(\omega) \leq 8) = \sum_{k=0}^{8} P_{15}(k) = \sum_{k=0}^{8} \binom{15}{k} 0.2^k \, 0.8^{15-k} = 0.999.$$

b) The probability that exactly eight fail is

$$P(\eta(\omega) = 8) = \binom{15}{8} 0.2^8 \, 0.8^7 = 0.003.$$

c) The probability that at least eight fail is

$$P(\eta(\omega) \geq 8) = 1 - P(\eta(\omega) \leq 7) = 1 - \sum_{k=0}^{7} P_{15}(k) = 1 - \sum_{k=0}^{7} \binom{15}{k} 0.2^k \, 0.8^{15-k} = 1 - 0.996 = 0.004.$$

d) Finally, the probability that between 4 and 7, inclusive, fail is

$$P(4 \leq \eta(\omega) \leq 7) = \sum_{k=4}^{7} P_{15}(k) = \sum_{k=4}^{7} \binom{15}{k} 0.2^k \, 0.8^{15-k} = 0.348.$$

§14. THE POISSON LAW

Let us return to Bernoulli trials and to the binomial distribution (see (25)). In many applications the number of trials n is large, and computation of $P_n(k)$ may be cumbersome. It is therefore of interest to see if some approximation could be found, when n becomes large.

A random phenomenon whose sample space Ω consists of all nonnegative integers that is $\Omega = \{\omega_0, \omega_1, ..., \omega_k, ...\}$, and on whose subsets a probability function $P(\cdot)$ is defined in terms of a parameter $\lambda > 0$ by

$$p_k = P(\{\omega_k\}) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad \text{for} \quad k = 0, 1, ... \tag{27}$$

is said to obey the Poisson probability law with parameter λ.
Let us show that the Poisson probability law naturally arises from the binomial law.

Theorem 3 (Poisson, 1837). *Consider n Independent repeated Bernoulli trials in which the probability of success is equal $p = P(A)$. Let the number of trials $n \to \infty$ and $\lambda = n \cdot p$ then*

$$P_n(k) = \binom{n}{k} p^k (1-p)^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda}, \quad n \to \infty \tag{28}$$

In other words, if n independent repeated trials, each of which results in a "success" with probability p, are performed, then, when n is large and p small enough to make $n \cdot p = \lambda$ moderate, the number of successes occurring is approximately a Poisson with parameter $\lambda = n \cdot p$.

Example 40. A book of 200 pages contains 100 misprints. Find the probability that a given page contains at least 2 misprints.

Solution: We have $\lambda = \frac{100}{200} = \frac{1}{2}$, hence

$$P(\text{ the number of misprints } \geq 2) = 1 - P(\text{ the number of misprints } < 2) =$$

$$= 1 - p_0 - p_1 \approx 1 - (0.6065 + 0.3033) = 1 - 0.9098 = 0.0902.$$

Exercise 10 (The behavior of the Poisson probabilities). Show that probabilities of the Poisson probability law, given by (27), increase monotonically, then decrease monotonically as k increases, and reach their maximum when k is the largest integer not exceeding λ.

§15. THE HYPERGEOMETRIC PROBABILITY MODEL

The hypergeometric distribution is closely related to the binomial distribution. The number of successes in n trials cannot be described by a binomial model unless the trials are independent, that is, the probability of a success remains constant from trial to trial. But suppose we have a finite set of N items, and M of which have some characteristic so that $(N - M)$ do not have the characteristic. When this finite population of N items is sampled *without replacement*, the trials cannot be independent because the probability of a success on the lth trial depends on the outcomes for all previous trials.

The assumptions leading to the hypergeometric distribution are:

1. The population or set to be sampled consists of N individuals, objects, or elements (a finite population).

2. Each item can be characterized as a success (S) or a failure (F), and there are M successes in the population.

3. A sample of n individuals is drawn in such a way that each subset of size n is equally likely to be chosen.

As in the Binomial law, the only fact about outcome of a succession of n trials in which we are interested is the *number of successes*.

We now compute the probability that the number of successes will be k, for any integer k.

If η is the number of S's in a random sample of size n drawn from a population consisting of M successes and $(N - M)$ failures, then the probability distribution of η, called the *hypergeometric distribution*, is given by

$$P(\eta = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

for k an integer satisfying $0 \leq k \leq \min(n, M)$.

Example 41. A bridge hand consists of any 13 cards selected from a 52 card deck without regard to order. There are $\binom{52}{13}$ different bridge hands, which works out to approximately 635 billion. Since there are 13 cards in each suit, the number of hands consisting entirely of clubs and / or spades (no red cards) is $\binom{26}{13} = 10,400,597$. One of these $\binom{26}{13}$ hands consists entirely of spades and one consists entirely of clubs, so there are

$$\left[\binom{26}{13} - 2\right]$$

hands that consist entirely of clubs and spades with both suits represented in the hand. Suppose a bridge hand is dealt from a well–shuffled deck (that is, 13 cards are randomly selected from among the 52 possibilities), and let

$$A = \{\text{the hand consists entirely of spades and clubs with both suits represented}\},$$

$$B = \{\text{the hand consists of exactly two suits}\}.$$

The $N = \binom{52}{13}$ possible outcomes are equally likely, so

$$P(A) = \frac{N(A)}{N(\Omega)} = \frac{\binom{26}{13} - 2}{\binom{52}{13}} = 0.0000164.$$

Since there are $\binom{4}{2} = 6$ combinations consisting of two suits, of which spades and clubs is one such combination, we obtain

$$P(B) = \frac{N(B)}{N(\Omega)} = \frac{6\left[\binom{26}{13} - 2\right]}{\binom{52}{13}} = 0.0000984.$$

That is, a hand consisting entirely of cards from exactly two of the four suits will occur roughly once in every 10,000 hands. If you play bridge only once a month, it is likely that you will never be dealt such a hand.

APPENDIX:

Proof of Theorem 3: **To prove (28), we need only rewrite its left–hand side in the following form (by substituting $\frac{\lambda}{n}$ instead of p)**

$$\frac{n(n-1)...(n-k+1)}{n^k}\,\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^{n-k}.$$

Since

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda},$$

we obtain (28).

LECTURE 8

§16. RANDOM VARIABLES

In engineering and the physical sciences many random phenomena of interest are associated with the numerical outcomes of some physical quantity. However, we also saw examples in which the outcomes are not in numerical terms. Outcomes of this latter type may also be identified numerically by artificially assigning numerical values to each of the possible alternative events. In other words, the possible outcomes of a random phenomenon can be identified numerically, either naturally or artificially. In any case, an outcome or event may be identified through the values of a function. In performing an experiment we are mainly interested in some function of the outcomes as opposed to the actual outcome itself. For instance, in tossing dice we are often interested in the sum of the two dice and are not really concerned about the actual outcome. That is, we may be interested in knowing that the sum is seven and not be concerned over whether the actual outcome was $(1,6)$ or $(6,1)$ or $(2,5)$ or $(5,2)$ or $(3,4)$ or $(4,3)$. The above means that to any outcome $\omega \in \Omega$ corresponds a real number $\eta(\omega)$. Briefly, real–valued functions defined on the sample space are known as random variables. Therefore we come to the following definition.

Definition 11. Let (Ω, P) be a probability space, i. e. Ω is a sample space on which a probability P has been defined. A *random variable* is a function η from Ω to the set of real numbers

$$\eta: \qquad \Omega \longmapsto \mathbb{R}^1,$$

i. e. for every outcome $\omega \in \Omega$ there is a real number, denoted by $\eta(\omega)$, which is called the value of $\eta(\cdot)$ at ω.

A SIMPLEST EXAMPLES OF RANDOM VARIABLES

Example 41. $\eta(\omega) \equiv \text{const}$ is a random variable.

Example 42. $\eta(\omega) = I_A(\omega)$, where

$$I_A(\omega) = \begin{cases} 1 & \text{if} \quad w \in A \\ 0 & \text{if} \quad \omega \notin A \end{cases}$$

is called the indicator function of the event A ($A \subset \Omega$).
Because the value of a random variable is determined by the outcome of the experiment, we may assign probabilities to the possible values of the random variable.

Example 43. Suppose that our experiment consists of tossing 3 fair coins. If we let $\eta(\omega)$ denote the number of heads appearing, then $\eta(\omega)$ is a random variable taking on one of the values $0, 1, 2, 3$ with respective probabilities

$$p_0 = P(\omega\colon\ \eta(\omega) = 0) = \frac{1}{8},$$

$$p_1 = P(\omega\colon\ \eta(\omega) = 1) = \frac{3}{8},$$

$$p_2 = P(\omega\colon\ \eta(\omega) = 2) = \frac{3}{8},$$

$$p_3 = P(\omega\colon\ \eta(\omega) = 3) = \frac{1}{8}.$$

Since $\eta(\omega)$ must take on one of the values 0 through 3, we have

$$1 = \sum_{k=0}^{3} p_k = \sum_{k=0}^{3} P(\omega\colon\ \eta(\omega) = k)$$

which, of course, is in accord with the above probabilities.

Example 44. Two balls are drawn without replacement from a bowl containing 6 balls, of which 4 are white. Let random variable $\eta(\omega)$ be the number of white drawn balls. The sample space Ω contains 15 outcomes and $\eta(\omega)$ is a random variable taking on one of the values $0, 1, 2$ with respective probabilities

$$p_0 = P(\omega\colon\ \eta(\omega) = 0) = \frac{1}{15},$$

$$p_1 = P(\omega:\ \eta(\omega) = 1) = \frac{8}{15},$$
$$p_2 = P(\omega:\ \eta(\omega) = 2) = \frac{6}{15}.$$

Since $\eta(\omega)$ must take on one of the values 0, 1, 2, we have

$$1 = \sum_{k=0}^{2} p_k = \sum_{k=0}^{2} P(\omega:\ \eta(\omega) = k)$$

which, of course, is in accord with the above probabilities.

Therefore, a random variable may be considered as a rule that maps events in a sample space into the real line. The purpose and advantages of identifying events in numerical terms should be obvious — this will then permit convenient analytical description as well as graphical display of events and their probabilities.

§17. DISTRIBUTION FUNCTIONS

The distribution function F of a random variable $\eta(\omega)$ is defined for all real numbers $x \in \mathbb{R}^1$, by the formula

$$F(x) = P(\omega:\ \eta(\omega) < x). \tag{29}$$

In words, $F(x)$ denotes the probability that the random variable $\eta(\omega)$ takes on a value that is less than x.

Some properties of the distribution function are the following:

Property 0. $0 \le F(x) \le 1$.

Property 1. F is a nondecreasing function, that is, if $x_1 \le x_2$ then $F(x_1) \le F(x_2)$.

Proof: We will present two proofs. For $x_1 \le x_2$ the event $\{\omega:\ \eta(\omega) < x_1\}$ is contained in the event $\{\omega:\ \eta(\omega) < x_2\}$ and so cannot have a larger probability (see Property 5 of Probabilities), i. e.

$$P(\omega:\ \eta(\omega) < x_1) \le P(\omega:\ \eta(\omega) < x_2).$$

Therefore, by definition of the distribution function we have $F(x_1) \le F(x_2)$.

Another proof of the property 1 is the following. Let us prove the formula

$$P(\omega:\ x_1 \le \eta(\omega) < x_2) = F(x_2) - F(x_1), \qquad \text{for all} \quad x_1 < x_2. \tag{30}$$

This can best be seen by writing the event $\{\omega\colon\ \eta(\omega) < x_2\}$ as the union of the mutually exclusive events $\{\omega\colon\ \eta(\omega) < x_1\}$ and $\{\omega\colon\ x_1 \le \eta(\omega) < x_2\}$. That is,

$$\{\omega\colon\ \eta(\omega) < x_2\} = \{\omega\colon\ \eta(\omega) < x_1\} \bigcup \{\omega\colon\ x_1 \le \eta(\omega) < x_2\}.$$

and so

$$P(\omega\colon\ \eta(\omega) < x_2) = P(\omega\colon\ \eta(\omega) < x_1) + P(\omega\colon\ x_1 \le \eta(\omega) < x_2)$$

which established equation (30). By Axiom 1 the left–hand side of (30) is nonnegative, and therefore $F(x_2) - F(x_1) \ge 0$. The proof is complete.

Property 2. $F(x) \to 1$ as $x \to +\infty$.

Property 3. $F(x) \to 0$ as $x \to -\infty$.

Property 4. $F(x)$ is left continuous. That is, for any x and any increasing sequence x_n that converges to x,

$$\lim_{n\to\infty} F(x_n) = F(x).$$

The proofs of Properties 2, 3 and 4 you can find in the Appendix of the present lecture.

Thus, Properties 1 – 4 are necessary conditions for a function $G(x)$ to be a distribution function.

However, these properties are also sufficient. This assertion follows from the following theorem which we cite without proof.

Theorem 4 (about Distribution Function). *Let a function $G(x)$, $x \in \mathbb{R}^1$ satisfy the Properties 1 — 4. Then there exist a probability space (Ω, P) and a random variable $\eta(\omega)$ for which distribution function coincides with given function $G(x)$, i. e.*

$$P(\omega\colon \quad \eta(\omega) < x) = G(x).$$

Therefore, for giving an example of a random variable we have to cite a function which satisfies the Properties 1 — 4.

We want to stress that in Theorem about distribution function a random variable $\eta(\omega)$ is determined by non–unique way.

Example 45. Let (Ω, P) be a probability space and $P(A) = P(\overline{A}) = 0.5$. We define the following two random variables

$$\eta_1(\omega) = \begin{cases} 1 & \text{if} \quad \omega \in A \\ -1 & \text{if} \quad \omega \notin A \end{cases}, \qquad \eta_2(\omega) = \begin{cases} 1 & \text{if} \quad \omega \in \overline{A} \\ -1 & \text{if} \quad \omega \in A \end{cases}.$$

It is obvious that $\{\omega: \quad \eta_1(\omega) \neq \eta_2(\omega)\} = \Omega$. However,

$$F_{\eta_1}(x) = F_{\eta_2}(x) = \begin{cases} 0 & \text{if} \quad x \leq -1 \\ 0.5 & \text{if} \quad -1 < x \leq 1 \\ 1 & \text{if} \quad x > 1 \end{cases}.$$

Definition 12. Two random variables $\eta_1(\omega)$ and $\eta_2(\omega)$ are said to be *Identically Distributed* if their distribution functions are equal, that is,

$$F_{\eta_1}(x) = F_{\eta_2}(x) \quad \text{for all} \quad x \in \mathbb{R}^1.$$

§17-1. DISTRIBUTION FUNCTIONS

We can also give the following definition of distribution function.

The distribution function F of a random variable $\eta(\omega)$ is defined for all real numbers $x \in \mathbb{R}^1$, by the formula

$$F(x) = P(\omega: \ \eta(\omega) \leq x). \tag{29}$$

In words, $F(x)$ denotes the probability that the random variable $\eta(\omega)$ takes on a value that is less than or equal to x.

Some properties of the distribution function are the following:

Property 0. $0 \leq F(x) \leq 1$.

Property 1. F is a nondecreasing function, that is, if $x_1 \leq x_2$ then $F(x_1) \leq F(x_2)$.

Proof: We will present two proofs. For $x_1 \leq x_2$ the event $\{\omega: \ \eta(\omega) \leq x_1\}$ is contained in the event $\{\omega: \ \eta(\omega) \leq x_2\}$ and so cannot have a larger probability (see Property 5 of Probabilities), i. e.

$$P(\omega: \ \eta(\omega) \leq x_1) \leq P(\omega: \ \eta(\omega) \leq x_2).$$

Therefore, by definition of the distribution function we have $F(x_1) \leq F(x_2)$.

Another proof of the property 1 is the following. Let us prove the formula

$$P(\omega:\ x_1 < \eta(\omega) \le x_2) = F(x_2) - F(x_1), \qquad \text{for all} \quad x_1 < x_2. \tag{30}$$

This can best be seen by writing the event $\{\omega:\ \eta(\omega) \le x_2\}$ as the union of the mutually exclusive events $\{\omega:\ \eta(\omega) \le x_1\}$ and $\{\omega:\ x_1 < \eta(\omega) \le x_2\}$. That is,

$$\{\omega:\ \eta(\omega) \le x_2\} = \{\omega:\ \eta(\omega) \le x_1\} \bigcup \{\omega:\ x_1 < \eta(\omega) \le x_2\}.$$

and so

$$P(\omega:\ \eta(\omega) \le x_2) = P(\omega:\ \eta(\omega) \le x_1) + P(\omega:\ x_1 < \eta(\omega) \le x_2)$$

which established equation (30). By Axiom 1 the left–hand side of (30) is nonnegative, and therefore $F(x_2) - F(x_1) \ge 0$. The proof is complete.

Property 2. $F(x) \to 1$ as $x \to +\infty$.

Property 3. $F(x) \to 0$ as $x \to -\infty$.

Property 4. $F(x)$ is right continuous. That is, for any x and any decreasing sequence x_n that converges to x,

$$\lim_{n\to\infty} F(x_n) = F(x).$$

Thus, Properties 1 – 4 are necessary conditions for a function $G(x)$ to be a distribution function.

However, these properties are also sufficient. This assertion follows from the following theorem which we cite without proof.

Theorem 4 (about Distribution Function). *Let a function $G(x)$, $x \in \mathbb{R}^1$ satisfy the Properties 1 — 4. Then there exist a probability space (Ω, P) and a random variable $\eta(\omega)$ for which distribution function coincides with given function $G(x)$, i. e.*

$$P(\omega: \quad \eta(\omega) \le x) = G(x).$$

Therefore, for giving an example of a random variable we have to cite a function which satisfies the Properties 1 — 4.

APPENDIX:

Proof of Property 2: **To prove this property we note that x_n increases to $+\infty$, then the events $A_n = \{\omega\colon\ \eta(\omega) < x_n\}$, $n \geq 1$, are increasing events, i. e. $A_n \subset A_{n+1}$ for any n and whose union is the certain event, that is**

$$A_n \uparrow \Omega.$$

Therefore, by Property 12 of Probability

$$P(A_n) \uparrow 1,$$

Since, by definition of F, $P(A_n) = F(x_n)$, the proof is complete.

Proof of Property 3: **To prove this property we note that x_n decreases to $-\infty$, then the events $A_n = \{\omega\colon\ \eta(\omega) < x_n\}$, $n \geq 1$, are decreasing events, i. e. $A_{n+1} \subset A_n$ for any n and whose intersection is the impossible event, that is**

$$A_n \downarrow \emptyset.$$

Therefore, by Property 11 of Probability

$$P(A_n) \downarrow 0,$$

Since, by definition of F, $P(A_n) = F(x_n)$, the proof is complete.

Proof of Property 4: **To prove Property 4 we note that if x_n increase to x, then $A_n = \{\omega\colon\ \eta(\omega) < x_n\}$, $n \geq 1$, are an increasing sequence of events, i. e. $A_n \subset A_{n+1}$ for any n and whose union is equal to the event $A = \{\omega\colon\ \eta(\omega) < x\}$. Hence by Property 11 of Probability**

$$P(A_n) \uparrow P(A),$$

Since, by definition of F, $P(A_n) = F(x_n)$ and $P(A) = F(x)$, the proof is complete.

L E C T U R E   9

§18. EXAMPLES OF DISTRIBUTION FUNCTIONS

Example 46. A random variable $\eta(\omega)$ is said to be *Normally distributed* if its distribution function has the following form

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{(y-a)^2}{2\sigma^2}\right) dy, \tag{31}$$

where a and σ are constants, moreover $a \in \mathbb{R}^1$ and $\sigma > 0$.
In order to show the correctness of Example 46 we have to verify that the function in the right–hand side of (31) satisfies the Properties 1 — 4. The correctness of Example 46 can be found in the Appendix I of this lecture.
The normal distribution plays a central role in probability and statistics. This distribution is also called the Gaussian distribution after Carl Friedrich Gauss, who proposed it as a model for measurement errors.

Example 47. A random variable is said to be *Uniformly distributed* on the interval (a, b) if its distribution function is given by

$$F(x) = \begin{cases} 0 & \text{if} \quad x \le a \\ \dfrac{x-a}{b-a} & \text{if} \quad a \le x \le b \\ 1 & \text{if} \quad x \ge b \end{cases}. \tag{32}$$

It is obvious that the function (32) satisfies all Properties 1 — 4.

Example 48. A random variable is said to be *Exponentially distributed* with parameter $\lambda > 0$ if its distribution function is given by

$$F(x) = \begin{cases} 0 & \text{if} \quad x \le 0 \\ 1 - e^{-\lambda x} & \text{if} \quad x \ge 0 \end{cases}. \tag{33}$$

It is obvious that the function (33) satisfies all Properties 1 — 4.
Like the Poisson distribution, the exponential distribution depends on the only parameter.

Example 49. If $\eta(\omega) \equiv c$ then corresponding distribution function has the form

$$F(x) = \begin{cases} 0 & \text{if} \quad x < c \\ 1 & \text{if} \quad x \geq c \end{cases}. \tag{34}$$

Consider the experiment of flipping a symmetrical coin once. The two possible outcomes are "heads" (outcome ω_1) and "tails" (outcome ω_2), that is, $\Omega = \{\omega_1, \omega_2\}$. Suppose $\eta(\omega)$ is defined by putting $\eta(\omega_1) = 1$ and $\eta(\omega_2) = -1$. We may think of it as earning of the player who receives or loses a dollar according as the outcome is heads or tails. Corresponding distribution function has the form

$$F(x) = \begin{cases} 0 & \text{if} \quad x < -1 \\ 1/2 & \text{if} \quad -1 \leq x < 1 \\ 1 & \text{if} \quad x \geq 1 \end{cases}.$$

Lemma 4. *Let $F(x)$ be a distribution function of a random variable $\eta(\omega)$. Then for any real number x we have*

$$P\{\omega: \quad \eta(\omega) = x\} = F(x) - F(x - 0), \tag{35}$$

where $F(x-0)$ is the left–hand limit at x.

The Proof of Lemma 4 can be found in the Appendix II.

Therefore, for continuous distribution function (as in Examples 46 — 48) we have

$$P(\omega: \quad \eta(\omega) = x) = 0 \qquad \text{for any} \quad x \in \mathbb{R}^1.$$

Example 50. The distribution function of the random variable $\eta(\omega)$ is given by

$$F(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \frac{x}{2} & \text{if} \quad 0 \leq x < 1 \\ \frac{2}{3} & \text{if} \quad 1 \leq x < 2 \\ \frac{11}{12} & \text{if} \quad 2 \leq x < 3 \\ 1 & \text{if} \quad x \geq 3 \end{cases}.$$

Compute (a) $P(\omega\colon \eta(\omega) < 3)$, (b) $P(\omega\colon \eta(\omega) = 1)$, (c) $P(\omega\colon \eta(\omega) > \frac{1}{2})$, (d) $P(\omega\colon 2 < \eta(\omega) \leq 4)$.

Solution: (a) $P(\omega\colon \eta(\omega) < 3) = F(3-0) = \frac{11}{12}$,

(b) $P(\omega: \quad \eta(\omega) = 1) = F(1) - F(1-0) = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}$,

(c) $P(\omega\colon \eta(\omega) > \frac{1}{2}) = 1 - P(\omega: \quad \eta(\omega) \leq \frac{1}{2}) = 1 - F\left(\frac{1}{2}\right) = \frac{3}{4}$,

(d) $P(\omega: \quad 2 < \eta(\omega) \leq 4) = F(4) - F(2) = 1 - \frac{11}{12} = \frac{1}{12}$.

§19. DISCRETE RANDOM VARIABLES

A random variable that can take on at most a countable number of possible values is said to be discrete.

Typical examples of discrete random variables are:

1. The number of defective bolts in a sample of 10 drawn from industrial production.
2. The number of deer killed per year in a state wildlife preserve.
3. The number of rural electrified homes in a township.

For a discrete random variable $\eta(\omega)$, we define the probability mass function $p(x)$ of $\eta(\omega)$ by

$$p(x) = P(\omega: \quad \eta(\omega) = x). \tag{37}$$

The probability mass function $p(x)$ is positive for at most a countable number of values of x. That is, if $\eta(\omega)$ must assume one of the values x_1, x_2, ..., then

$$p(x_i) > 0, \quad i = 1, 2, ... \qquad p(x) = 0 \quad \text{for all other values of } \; x.$$

Since $\eta(\omega)$ must take on one of the values x_i, we have

$$\sum_{i=1}^{\infty} p(x_i) = \sum_{\substack{\text{over all points } x \\ \text{such that } p(x)>0}} p(x) = 1. \tag{38}$$

From the distribution function $F(x)$ one may obtain the probability mass function $p(x)$ by the following formula (see Lemma 4)

$$p(x_i) = F(x_i) - F(x_i - 0)$$

and conversely, the distribution function $F(x)$ can be expressed in terms of its probability mass function by

$$F(x) = \sum_{\substack{\text{over all points } x_i < x \\ \text{such that } p(x_i)>0}} p(x_i). \tag{39}$$

Therefore, a discrete random variable $\eta(\omega)$ can be defined by its *Distribution Law*, i. e.

x_1	x_2	...	x_n	...
$p(x_1)$	$p(x_2)$	...	$p(x_n)$	...

where x_1, x_2 ... are possible values of $\eta(\omega)$ and $p(x_1), p(x_2), ...$ satisfy condition (38).

Hence, the distribution function of a discrete random variable is a step function. That is, the value of $F(x)$ is constant in the intervals $[x_{i-1}, x_i)$ and then takes a step (or jump) of size $p(x_i)$ at point x_i.

Example 51. Let $\eta(\omega)$ be a discrete random variable with the following distribution law

1	2	3	4
1/4	1/2	1/8	1/8

.

Then its distribution function is given by

$$F(x) = \begin{cases} 0 & \text{if} \quad x < 1 \\ \frac{1}{4} & \text{if} \quad 1 \leq x < 2 \\ \frac{3}{4} & \text{if} \quad 2 \leq x < 3 \\ \frac{7}{8} & \text{if} \quad 3 \leq x < 4 \\ 1 & \text{if} \quad x \geq 4 \end{cases}.$$

It is easy to note that the size of the step at any of the values $1, 2, 3, 4$ is equal to the probability that $\eta(\omega)$ assumes that particular value.
It is also noted that since probability is associated with points in the discrete case, the inclusion or exclusion of interval end points is important.

APPENDIX I:

The correctness of Example 46: **Indeed, the function (31) as a function of upper bound is continuous. Therefore the Properties 4 and 3 are satisfied. Since the integrand is positive we also have Property 1. Thus it is left to prove that**

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left(-\frac{(y-a)^2}{2\sigma^2}\right)dy=1.$$

Let us make a change of variable:

$$x=\frac{y-a}{\sigma},\quad \sigma\,dx=dy.$$

Therefore

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left(-\frac{x^2}{2}\right)dx=1.$$

To prove that $F(x)$ is indeed a distribution function, we need to show that

$$A=\int_{-\infty}^{+\infty}\exp\left(-\frac{x^2}{2}\right)dx=\sqrt{2\pi}.$$

Therefore

$$A^2=\left(\int_{-\infty}^{+\infty}\exp\left(-\frac{x^2}{2}\right)dx\right)\cdot\left(\int_{-\infty}^{+\infty}\exp\left(-\frac{y^2}{2}\right)dy\right)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\exp\left(-\frac{x^2+y^2}{2}\right)dx\,dy.$$

We now evaluate the double integral by means of a change of variables to polar coordinates. That is, let

$$x=r\cdot\cos\varphi,$$
$$y=r\cdot\sin\varphi.$$

As the area element in polar coordinates equals $r\cdot dr\,d\varphi$, therefore

$$dx\,dy=r\,dr\,d\varphi.$$

Thus

$$A^2=\int_0^{\infty}\int_0^{2\pi}\exp\left(-\frac{r^2}{2}\right)r\,dr\,d\varphi=2\pi\int_0^{\infty}r\,\exp\left(-\frac{r^2}{2}\right)dr=-2\pi\,\exp\left(-\frac{r^2}{2}\right)\Bigg|_0^{\infty}=2\pi.$$

Hence $A=\sqrt{2\pi}$ and the result is proved. Therefore, by Theorem about distribution function there exists a random variable for which its distribution function has the form (31).

APPENDIX II:

Proof of Lemma 4: **Let us prove the following equation**

$$P(\omega:\quad \eta(\omega) < x) = F(x-0), \tag{36}$$

i. e. we want to compute the probability that $\eta(\omega)$ is less than x.
It is not difficult to verify that

$$\{\omega:\quad \eta(\omega) < x\} = \bigcup_{n=1}^{\infty} A_n,$$

where $A_n = \left\{\omega:\quad \eta(\omega) \le x - \frac{1}{n}\right\}$.
A_n **is an increasing sequence and therefore, by Definition 4 tends to the event** $\{\omega:\quad \eta(\omega) < x\}$**. Thus**

$$A_n \uparrow \{\omega:\quad \eta(\omega) < x\}.$$

By Property 11 we get

$$P(A_n) \uparrow P(\omega:\quad \eta(\omega) < x).$$

Hence (36) is proved.
As

$$P(\omega:\quad \eta(\omega) = x) = P(\omega:\quad \eta(\omega) \le x) - P(\omega:\quad \eta(\omega) < x) = F(x) - F(x-0)$$

the assertion of the Lemma follows from the equation (36). The proof is complete.

LECTURE 10

Example 52. A random variable η has a *Binomial distribution* with parameters n and p if it is a discrete random variable

0	1	...	n
$p(0)$	$p(1)$	...	$p(n)$

whose probability mass function $p(x)$ is given by

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, ..., n.$$

Thus for a random variable $\eta(\omega)$, which has a binomial distribution with parameter $n = 6$ and $p = 1/3$ we have

$$P(\omega\colon\ 1 < \eta(\omega) < 2) = 0,$$

$$P(\omega\colon\ 1 < \eta(\omega) \leq 2) = P(\omega\colon\ \eta(\omega) = 2) = \binom{6}{2}(1/3)^2(2/3)^4 = 0.3292,$$

$$P(\omega\colon\ 1 \leq \eta(\omega) \leq 2) = P(\omega\colon\ \eta(\omega) = 1) + P(\omega\colon\ \eta(\omega) = 2) =$$

$$= \binom{6}{1}(1/3)(2/3)^5 + \binom{6}{2}(1/3)^2(2/3)^4 = 0.5926.$$

Example 53. A random variable η has a *Poisson distribution* with parameter $\lambda > 0$ if it is a discrete random variable

0	1	...	n	...
$p(0)$	$p(1)$	...	$p(n)$	...

whose probability mass function $p(n)$ is given by

$$p(n) = P(\eta = n) = e^{-\lambda}\,\frac{\lambda^n}{n!}, \quad n = 0, 1, ...$$

Example 54. Suppose that fifteen balls are numbered so that one ball has the number 1, two balls have the number 2, etc. An experiment is to be conducted by mixing the fifteen balls thoroughly and drawing one at random. If we let $\eta(\omega)$ be the number that occurs on the random draw, then the probability function which describes the possible outcomes and their probabilities may be given in tabular form as

1	2	3	4	5
$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{4}{15}$	$\frac{5}{15}$

Note that the function $p_k = \dfrac{k}{15}$ gives the probability for any outcome which may occur. The model is used, of course, to find the probability of any specific outcome or event of interest. The probability of obtaining a number less than 4 is

$$P(\omega\colon\ 1 \le \eta(\omega) \le 3) = \frac{6}{15} = \frac{2}{5}.$$

The probability of getting either a 4 or 5 is

$$P(\omega\colon\ 4 \le \eta(\omega) \le 5) = \frac{9}{15} = \frac{3}{5}.$$

§20. CONTINUOUS RANDOM VARIABLES

In §19 we considered discrete random variables, that is, random variables whose set of possible values is either finite or countably infinite. However, there exist random variables whose set of possible values is uncountable. Let $\eta(\omega)$ be such a random variable. We say that $\eta(\omega)$ is an ***absolutely continuous*** random variable if there exists a function $f(x)$ defined for all real numbers and the distribution function $F(x)$ of the random variable $\eta(\omega)$ is represented in the form

$$F(x) = \int_{-\infty}^{x} f(y)\,dy. \tag{40}$$

The function f is called the ***Density function*** of $\eta(\omega)$.

Typical examples of continuous random variables are:

1. The height of a human.
2. The length of life of a human cell.
3. The amount of sugar in an orange.

A function $f(x)$ must have certain properties in order to be a density function. Since $F(x) \to 1$ as $x \to +\infty$ we obtain

Property 1.

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1. \tag{41}$$

Property 2. $f(x)$ is a nonnegative function.

Proof: Differentiating both sides of (40) yields

$$f(x) = \frac{d}{dx} F(x). \tag{42}$$

That is, the density is the derivative of the distribution function. We know that the first derivative of a nondecreasing function is always nonnegative. Therefore the proof is complete as $F(x)$ is nondecreasing.

Remarkably that these two properties are also sufficient for a function $g(x)$ be a density function.

Theorem 5 (About Density Function). *Let a function $g(x)$, $x \in \mathbb{R}^1$ satisfy (41) and, in addition, satisfies the condition*

$$g(x) \geq 0 \quad \textit{for all } x \in \mathbb{R}^1.$$

Then there exist a probability space (Ω, P) and an absolutely continuous random variable $\eta(\omega)$ for which density function coincides with given function $g(x)$.

Proof: Let us define a function

$$G(x) = \int_{-\infty}^{x} g(y)\, dy.$$

It is not difficult to verify that $G(x)$ satisfies all conditions 1 — 4 for distribution function. Therefore by Theorem 4 about distribution function, there exists a random variable $\eta(\omega)$ for which distribution function coincides with $G(x)$. By definition of density function, $g(x)$ is a density function of the random variable $\eta(\omega)$. The proof is complete.

Therefore, for giving an example of an absolutely continuous random variable we have to cite a *nonnegative* function which satisfies (41).

The normally distributed random variable (see Example 46) is absolutely continuous and its density function has the form

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right), \tag{43}$$

where a and σ are constant, moreover $a \in \mathbb{R}^1$ and $\sigma > 0$.

The uniformly distributed random variable over the interval (a, b) (see Example 47) is absolutely continuous and its density function has the form

$$f(x) = \begin{cases} 0 & \text{if} \quad x \notin (a, b) \\ \dfrac{1}{b-a} & \text{if} \quad a \leq x \leq b \end{cases}. \tag{44}$$

It is obvious that the function (44) satisfies (41).

The exponentially distributed random variable with parameter $\lambda > 0$ (see Example 48) is absolutely continuous and its density function has the form

$$f(x) = \begin{cases} 0 & \text{if} \quad x \leq 0 \\ \lambda\, e^{-\lambda x} & \text{if} \quad x > 0 \end{cases}. \tag{45}$$

It is obvious that the function (45) satisfies (41).

We obtain from (30) that

$$\begin{aligned} &P(\omega\colon\ a \leq \eta(\omega) \leq b) = P(\omega\colon\ a \leq \eta(\omega) < b) = \\ &= P(\omega\colon\ a < \eta(\omega) \leq b) = P(\omega\colon\ a < \eta(\omega) < b) = \int_a^b f(x)\, dx. \end{aligned} \tag{46}$$

As the distribution function of an absolutely continuous random variable is continuous at all points thus

$$P(\omega\colon\ \eta(\omega) = x) = 0$$

for any fixed x.

Therefore this equation states that the probability that an absolutely continuous random variable will assume any fixed value is zero.

A somewhat more intuitive interpretation of the density function may be obtained from (46). If $\eta(\omega)$ is an absolutely continuous random variable having density function $f(x)$, then for small dx

$$P(\omega\colon\ x \leq \eta(\omega) \leq x + dx) = f(x)\, dx + o(dx).$$

L E C T U R E 11

§21. JOINT DISTRIBUTION FUNCTIONS

Two random variables, $\eta_1(\omega)$ and $\eta_2(\omega)$, are said to be jointly distributed if they are defined as functions on the same probability space. It is then possible to make joint probability statements about $\eta_1(\omega)$ and $\eta_2(\omega)$ (that is, probability statements about the simultaneous behavior of the two random variables). In order to deal with such probabilities, we define, for any two random variables $\eta_1(\omega)$ and $\eta_2(\omega)$, the *Joint distribution function* of $\eta_1(\omega)$ and $\eta_2(\omega)$ by the formula

$$F(x_1, x_2) = P\Big(\omega: \ \eta_1(\omega) \le x_1 \bigcap \eta_2(\omega) \le x_2\Big), \quad x_1, x_2 \in \mathbb{R}^1. \tag{47}$$

PROPERTIES OF JOINT DISTRIBUTION FUNCTION

Property 1. $F(x_1, x_2)$ is a nondecreasing function by each argument.

Property 2. $F(x_1, x_2) \to 1$ as $x_1 \to +\infty$ and $x_2 \to +\infty$.

Property 3. $F(x_1, x_2) \to 0$ as either x_1 or x_2 tends to $-\infty$.

Property 4. $F(x_1, x_2)$ is right continuous by each argument.
Verification of these properties is left as an exercise because they can be proved as the corresponding properties for distribution function of a random variable.
Therefore, Properties 1 — 4 are necessary conditions for $G(x_1, x_2)$ to be a joint distribution function.

Theorem 6. *Let $F(x_1, x_2)$ be a joint distribution function of a vector $(\eta_1(\omega), \eta_2(\omega))$. Then*

$$P\Big\{\omega: \ a_1 < \eta_1(\omega) \le b_1 \bigcap a_2 < \eta_2(\omega) \le b_2\Big\} = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2), \tag{48}$$

whenever, $a_1 < b_1$ and $a_2 < b_2$, i. e. probability that a random point $(\eta_1(\omega), \eta_2(\omega))$ belongs to rectangle $(a_1, b_1] \times (a_2, b_2]$ is equal to the algebraic sum of the values of joint distribution function at the vertices of the rectangle ($F(b_1, b_2)$ and $F(a_1, a_2)$ with plus, but $F(b_1, a_2)$ and $F(a_1, b_2)$ with minus).

The theorem is easily proved geometrically.

Let us give an example that Properties 1 — 4 are not sufficient for a function $G(x_1, x_2)$ to be a joint distribution function of a random variables $\eta_1(\omega)$ and $\eta_2(\omega)$.

Let

$$G(x_1, x_2) = \begin{cases} 0 & \text{if either } x_1 < 0 \text{ or } x_2 < 0 \\ & \text{or } x_1 \geq 0, x_2 \geq 0 \text{ and } x_1 + x_2 < 1 \\ & \\ 1 & \text{if } x_1 \geq 0, x_2 \geq 0 \text{ and } x_1 + x_2 \geq 1. \end{cases}$$

It is not difficult to verify that the function satisfies Properties 1 — 4. Let us assume that the function $G(x_1, x_2)$ is a joint distribution function for some random vector (η_1, η_2). Then

$$P\left\{\omega: \ \frac{1}{2} < \eta_1(\omega) \leq 1 \bigcap 0 < \eta_2(\omega) \leq \frac{1}{2}\right\} =$$

$$= G\left(1, \frac{1}{2}\right) - G(1, 0) - G\left(\frac{1}{2}, \frac{1}{2}\right) + G\left(\frac{1}{2}, 0\right) = 1 - 1 - 1 + 0 = -1.$$

By Axiom 1 this probability must be nonnegative.

The distribution of $\eta_1(\omega)$ can be obtained from the joint distribution function as follows:

$$F_{\eta_1}(x_1) = P(\omega: \ \eta_1(\omega) \leq x_1) = P\left(\omega: \ \bigcup_{n=1}^{\infty}[(\eta_1(\omega) \leq x_1)\bigcap(\eta_2(\omega) \leq x_{2n})]\right) =$$

$$= \lim_{x_2 \to +\infty} P\left(\omega: \ (\eta_1(\omega) \leq x_1)\bigcap(\eta_2(\omega) \leq x_2)\right) = \lim_{x_2 \to +\infty} F(x_1, x_2) \equiv F(x_1, +\infty),$$

where x_{2n} is a monotone increasing sequence which tends to $+\infty$.

Proof: The proof follows from the relation

$$\{\omega: \ \eta_1(\omega) \leq x_1\} = \bigcup_{n=1}^{+\infty}\{\omega: \ (\eta_1(\omega) \leq x_1)\bigcap(\eta_2(\omega) \leq x_{2n})\}$$

and Property 12 of probability.

Similarly, the distribution function of $\eta_2(\omega)$ is given by

$$F_{\eta_2}(x_2) = P(\omega: \ \eta_2(\omega) \leq x_2) = \lim_{x_2 \to +\infty} F(x_1, x_2) \equiv F(+\infty, x_2).$$

The distribution functions F_{η_1} and F_{η_2} are sometimes referred to as the *Marginal* distributions of $\eta_1(\omega)$ and $\eta_2(\omega)$.

In the case when $\eta_1(\omega)$ and $\eta_2(\omega)$ are both discrete random variables, it is convenient to define the *joint probability mass function* of $\eta_1(\omega)$ and $\eta_2(\omega)$ by

$$p(x,y) = P\left\{\omega:\ \eta_1(\omega) = x \bigcap \eta_2(\omega) = y\right\}.$$

The probability mass function of $\eta_1(\omega)$ can be obtained from $p(x,y)$ by

$$p_{\eta_1(\omega)}(x) = P(\omega:\ \eta_1(\omega) = x) = \sum_{y:\ p(x,y)>0} p(x,y).$$

Similarly

$$p_{\eta_2(\omega)}(y) = P(\omega:\ \eta_2(\omega) = y) = \sum_{x:\ p(x,y)>0} p(x,y).$$

We say that $\eta_1(\omega)$ and $\eta_2(\omega)$ are *jointly continuous* if there exists a function $f(x_1,x_2)$ defined for all real x_1 and x_2, having the property

$$F(x_1,x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f(x,y)\,dx\,dy. \tag{49}$$

The function $f(x_1,x_2)$ is called the *joint density function* of $\eta_1(\omega)$ and $\eta_2(\omega)$.
It follows upon differentiation, that

$$f(x_1,x_2) = \frac{\partial^2}{\partial x_1\,\partial x_2}F(x_1,x_2) \tag{50}$$

whenever the partial derivatives are defined.
If $\eta_1(\omega)$ and $\eta_2(\omega)$ are jointly continuous, they are individually continuous, and their probability density functions can be obtained as follows:

$$P\{\omega:\ \eta_1(\omega) \le x_1\} = P\left\{\omega:\ \eta_1(\omega) \le x_1 \bigcap -\infty < \eta_2(\omega) < +\infty\right\} =$$
$$= \int_{-\infty}^{x_1}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy = \int_{-\infty}^{x_1} f_{\eta_1}(x)\,dx$$

where

$$f_{\eta_1}(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy \tag{51}$$

is thus the density function of $\eta_1(\omega)$. Similarly, the density function of $\eta_2(\omega)$ is given by

$$f_{\eta_2}(x) = \int_{-\infty}^{+\infty} f(x,y)\,dx. \tag{52}$$

It follows from (48) that

$$P\left\{\omega:\ x < \eta_1(\omega) \le x + dx \bigcap y < \eta_2(\omega) \le y + dy\right\} = \int_x^{x+dx} \int_y^{y+dy} f(x_1, x_2)\, dx_1\, dx_2 =$$
$$= f(x, y)\, dx\, dy + o(dx\, dy)$$

when dx and dy are small and $f(x_1, x_2)$ is continuous at (x, y). Hence $f(x, y)$ is a measure of how likely it is that the random vector $(\eta_1(\omega), \eta_2(\omega))$ will be near (x, y).

The following theorem we cite without proof.

Theorem 7 (About Joint Distribution Function). *Let a function* $G(x_1, x_2)$, $x_1, x_2 \in \mathbb{R}^1$ *satisfy the properties 1 — 4 and, in addition, the condition*

$$G(b_1, b_2) - G(a_1, b_2) - G(b_1, a_2) + G(a_1, a_2) \ge 0, \qquad \text{for any } a_1 < b_1 \text{ and } a_2 < b_2. \tag{53}$$

Then there exist a probability space (Ω, P) *and a random vector* $(\eta_1(\omega), \eta_2(\omega))$ *for which joint distribution function coincides with given function* $G(x_1, x_2)$, *that is*

$$P(\omega:\quad \eta_1(\omega) \le x_1 \bigcap \eta_2(\omega) \le x_2) = G(x_1, x_2).$$

Therefore, for giving an example of a random vector we have to cite a function which satisfies the properties 1 — 4 and the condition (53).
It should be noted that it is not easy to verify the condition (53) for a function $G(x_1, x_2)$.

Remark 6. We can also define joint distributions for n random variables in exactly the same manner as we did for $n = 2$. For instance, the joint distribution function $F(x_1, x_2, ..., x_n)$ of the n random variables $\eta_1(\omega)$, $\eta_2(\omega)$, ..., $\eta_n(\omega)$ is defined by

$$F(x_1, x_2, ..., x_n) = P\left(\omega:\ \eta_1(\omega) \le x_1 \bigcap \eta_2(\omega) \le x_2 \bigcap ... \bigcap \eta_n(\omega) \le x_n\right).$$

§22. SOME REMARKS ABOUT JOINT DENSITY FUNCTIONS

Remark 7. We note that a joint density function has two properties

$$1)\quad f(x_1, x_2) \ge 0; \tag{54}$$

$$2)\quad \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x_1, x_2)\, dx_1\, dx_2 = 1. \tag{55}$$

These conditions can be proved in the same manner as we did for $n = 1$ (compare with Properties 1, 2 in §20).

Remark 8. It should be noted that if a function $g(x_1, x_2)$ $x_1, x_2 \in \mathbb{R}^1$ satisfies the conditions

$$g(x_1, x_2) \geq 0 \quad \text{and} \quad \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x_1, x_2)\, dx_1\, dx_2 = 1$$

then there exist a probability space (Ω, P) and two random variables $\eta_1(\omega)$ and $\eta_2(\omega)$ such that their joint density function $f(x_1, x_2)$ coincides with $g(x_1, x_2)$, that is

$$P\left\{\omega:\ \eta_1(\omega) \leq x_1 \bigcap \eta_2(\omega) \leq x_2\right\} = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} g(x, y)\, dx\, dy. \tag{56}$$

Remark 9. If $f(x_1, x_2)$ is a joint density function for a random vector $(\eta_1(\omega), \eta_2(\omega))$ then

$$P\left\{\omega:\ a_1 < \eta_1(\omega) \leq b_1 \bigcap a_2 < \eta_2(\omega) \leq b_2\right\} = P\left\{\omega:\ a_1 \leq \eta_1(\omega) \leq b_1 \bigcap a_2 \leq \eta_2(\omega) \leq b_2\right\} = \\ = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x, y)\, dx\, dy. \tag{57}$$

(57) follows from Theorem 6 and (49).

Example 55. If the joint density function of two random variables is given by

$$f(x_1, x_2) = \begin{cases} 6\ \exp(-2x_1 - 3x_2) & \text{for} \quad x_1 > 0\ x_2 > 0 \\ 0 & \text{otherwise} \end{cases}$$

find

a). the probability that the first random variable will take on a value between 1 and 2 and the second random variable will take on a value between 2 and 3;

b). the probability that the first random variable will take on a value less than 2 and the second random variable will take on a value greater than 2;

c). the joint distribution function of the two random variables;

d). the probability that the both random variables will take on values less than 1;

e). the marginal density function of the first random variable;

f). the marginal distribution functions.

Solution: Using (48), (49) and performing the necessary integrations, we get

$$\int_1^2 \int_2^3 6\ \exp(-2x_1 - 3x_2)\, dx_1\, dx_2 = (e^{-2} - e^{-4})\,(e^{-6} - e^{-9}) \approx 0.0003$$

for part a), and

$$\int_0^2 \int_2^{\infty} 6\ \exp(-2x_1 - 3x_2)\, dx_1\, dx_2 = (1 - e^{-4})\, e^{-6}$$

for part b).

c). By definition (see (49))

$$F(x_1,x_2)=\begin{cases}\displaystyle\int_0^{x_1}\int_0^{x_2} 6\,\exp(-2y_1-3y_2)\,dy_1\,dy_2 & \text{for}\quad x_1>0\ x_2>0\\ 0 & \text{otherwise}\end{cases}$$

so that

$$F(x_1,x_2)=\begin{cases}(1-e^{-2x_1})\,(1-e^{-3x_2}) & \text{for}\quad x_1>0\ x_2>0\\ 0 & \text{otherwise}\end{cases}$$

and hence, for d) we obtain

$$F(1,1)=(1-e^{-2})\,(1-e^{-3})\approx 0.8216.$$

e). Using (51) we get

$$f_{\eta_1}(x_1)=\begin{cases}\displaystyle\int_0^{\infty} 6\,\exp(-2x_1-3x_2)\,dx_2 & \text{for}\quad x_1>0\\ 0 & \text{otherwise}\end{cases}$$

or

$$f_{\eta_1}(x_1)=\begin{cases}2\,\exp(-2x_1) & \text{for}\quad x_1>0\\ 0 & \text{otherwise}\end{cases}$$

f). Now since $F_{\eta_1}(x_1)=F(x_1,+\infty)$ and $F_{\eta_2}(x_2)=F(+\infty,x_2)$ it follows that

$$F_{\eta_1}(x_1)=\begin{cases}(1-e^{-2x_1}) & \text{for}\quad x_1>0\\ 0 & \text{otherwise}\end{cases}$$

and

$$F_{\eta_2}(x_2)=\begin{cases}(1-e^{-3x_2}) & \text{for}\quad x_2>0\\ 0 & \text{otherwise.}\end{cases}$$

Thus in our Example $F(x_1,x_2)=F_{\eta_1}(x_1)\,F_{\eta_2}(x_2)$.

LECTURE 12

§23. INDEPENDENT RANDOM VARIABLES

In sections 12 and 13 we define the notion of a series of independent trials. In this section we define the notion of independent random variables. This notion plays the same role in the theory of jointly distributed random variables that the notion of independent trials plays in the theory of sample spaces consisting of n trials. We consider the case of two jointly distributed random variables.

Let $\eta_1(\omega)$ and $\eta_2(\omega)$ be jointly distributed random variables, with individual distribution functions $F_{\eta_1}(x)$, $F_{\eta_2}(x)$, respectively, and joint distribution function $F(x_1, x_2)$.

Definition 13. Two jointly distributed random variables $\eta_1(\omega)$ and $\eta_2(\omega)$ are independent if their joint distribution function $F(x_1, x_2)$ may be written as the product of their individual distribution functions $F_{\eta_1}(x)$ and $F_{\eta_2}(x)$ in the sense that, for any real numbers x_1 and x_2

$$F(x_1, x_2) = F_{\eta_1}(x_1) \cdot F_{\eta_2}(x_2). \tag{58}$$

Similarly, two jointly continuous random variables $\eta_1(\omega)$ and $\eta_2(\omega)$ are independent if their joint density function $f(x_1, x_2)$ may be written as the product of their individual density functions $f_{\eta_1}(x_1)$ and $f_{\eta_2}(x_2)$ in the sense that, for any real numbers x_1 and x_2

$$f(x_1, x_2) = f_{\eta_1}(x_1) \cdot f_{\eta_2}(x_2). \tag{59}$$

Equation (59) follows from (58) by differentiating both sides of (58) first with respect to x_1 and then with respect to x_2. Equation (58) follows from (59) by integrating both sides of (59).

Similarly, two discrete random variables $\eta_1(\omega)$ and $\eta_2(\omega)$ are independent if their joint probability mass function $p(x, y)$ may be written as the product of their individual probability mass functions $p_{\eta_1}(x)$ and $p_{\eta_2}(y)$ in the sense that, for any x and y

$$p(x, y) = p_{\eta_1}(x) \cdot p_{\eta_2}(y). \tag{60}$$

The equivalence follows because if (58) is satisfied then we obtain (60). Furthermore, if equation (60) is valid, then for any real numbers x_1 and x_2, by (38) we obtain

$$F(x_1,x_2) = \sum_{x:\ x\le x_1}\sum_{y:\ y\le x_2} p(x,y) = \sum_{x:\ x\le x_1}\sum_{y:\ y\le x_2} p_{\eta_1}(x)\cdot p_{\eta_2}(y) = \sum_{x:\ x\le x_1} p_{\eta_1}(x)\cdot \sum_{y:\ y\le x_2} p_{\eta_2}(y) = \\ = F_{\eta_1}(x_1)\cdot F_{\eta_2}(x_2)$$

and therefore $\eta_1(\omega)$ and $\eta_2(\omega)$ are independent.

Thus, loosely speaking, $\eta_1(\omega)$ and $\eta_2(\omega)$ are independent if knowing the value of one does not change the distribution of the other. Random variables that are not independent are said to be dependent or nonindependent.

Lemma 5. *For any two independent random variables $\eta_1(\omega)$ and $\eta_2(\omega)$ their joint distribution function always satisfies the additional condition (53).*

Proof: As $\eta_1(\omega)$ and $\eta_2(\omega)$ are independent we can rewrite (21.7) in the following form

$$\begin{aligned} F(b_1,b_2) - F(a_1,b_2) - F(b_1,a_2) + F(a_1,a_2) &= F_{\eta_1}(b_1)\,F_{\eta_2}(b_2) - F_{\eta_1}(a_1)\,F_{\eta_2}(b_2) - \\ - F_{\eta_1}(b_1)\,F_{\eta_2}(a_2) + F_{\eta_1}(a_1)\,F_{\eta_2}(a_2) &= [F_{\eta_1}(b_1) - F_{\eta_1}(a_1)]\cdot[F_{\eta_2}(b_2) - F_{\eta_2}(a_2)]\,. \end{aligned} \tag{61}$$

As $F_{\eta_1}(\cdot)$ and $F_{\eta_2}(\cdot)$ are nondecreasing functions we conclude that the right–hand side of (61) is nonnegative. The proof is complete.

Example 56 (Buffon's needle). A fine needle of length $2a$ is dropped at random on a board covered with parallel lines distance $2b$ apart, where $b > a$. We shall show that the probability $\mathbf{p}$ that the needle intersects one of the lines equals $\dfrac{2a}{\pi\, b}$.

Solution: In terms of random variables the above experiment can be phrased as follows: We denote by η_1 the distance from the center of the needle to the nearest line and by η_2 the angle between the needle and the direction perpendicular to the lines. We assume that the random variables η_1 and η_2 are independent, η_1 is uniform in the interval $(0, b)$, and η_2 is uniform in the interval $(0, \pi/2)$. From this it follows that the joint density function for the vector (η_1, η_2) has the following form:

$$f(x_1,x_2) = f_{\eta_1}(x_1)\,f_{\eta_2}(x_2) = \frac{1}{b}\,\frac{2}{\pi}, \qquad 0\le x_1\le b, \qquad 0\le x_2\le\frac{\pi}{2}$$

and 0 elsewhere. Hence the probability that the point (η_1, η_2) is in a region D included in the rectangle

$$R = \left\{(x_1,x_2):\ 0\le x_1\le b, \qquad 0\le x_2\le\frac{\pi}{2}\right\}$$

equals the areas of D times $\frac{2}{\pi b}$.
The needle intersects the lines if $x_1 < a \cos x_2$. Hence $\mathbf{p}$ equals

$$\mathbf{p} = P\{\eta_1 < a \cos \eta_2\} = \frac{2}{\pi b} \int_0^{\pi/2} a \cos x_2 \, dx_2 = \frac{2a}{\pi b}.$$

The above can be used to determine experimentally the number π using the relative frequency interpretation of $\mathbf{p}$: If the needle is dropped n times and it intersects the lines m times, then

$$\frac{m}{n} \approx \mathbf{p} = \frac{2a}{\pi b}$$

hence

$$\pi \approx \frac{2an}{bm}.$$

Independent random variables have the following exceedingly important property, the proof of which we leave as an exercise for the reader.

Theorem 8. *Let $\eta_1(\omega)$ and $\eta_2(\omega)$ be independent random variables and $\varphi_1(x)$ and $\varphi_2(x)$ be two continuous functions from $\mathbb{R}^1$ into $\mathbb{R}^1$. Then the random variables $\zeta_1(\omega) = \varphi_1(\eta_1(\omega))$ and $\zeta_2 = \varphi_2(\eta_2(\omega))$ are also independent.*

That is independence of the random variables $\eta_1(\omega)$ and $\eta_2(\omega)$ implies independence of the random variables $\zeta_1(\omega)$ and $\zeta_2(\omega)$.

§24. EXPECTATION OF RANDOM VARIABLES

In this section we define the notion of the expectation of a random variable and describe the significant role that this notion plays in probability theory.
Given the random variable $\eta(\omega)$, we define the *expectation* of the random variable, denoted by $E\eta$, as the mean of the probability law of η. By definition,

$$E\eta = \begin{cases} \sum\limits_{x:\ p(x)>0} x\, p(x) & \text{if} \quad \eta \quad \text{is a discrete random variable} \\ \int_{-\infty}^{+\infty} x\, f_\eta(x)\, dx & \text{if} \quad \eta \quad \text{is an absolutely continuous random variable} \end{cases} \tag{62}$$

depending on whether η is specified its density function $f_\eta(x)$, or its probability mass function $p(x)$. In words, the expected value of discrete random variable η is a weighted average of the possible values that η can take on, each value being weighted by the probability that η assumes it. For instance, if the probability mass function of η is given by

$$p(0) = p(1) = 0.5$$

then

$$E\eta = 0 \cdot 0.5 + 1 \cdot 0.5 = 0.5$$

is just the ordinary average the two possible values 0 and 1 that η can assume.
On the other hand if

$$p(0) = 1/3, \quad p(1) = 2/3$$

then

$$E\eta = 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}$$

is a weighted average of the two possible values 0 and 1, where value 1 is given twice as much weight as the value 0, since $p(1) = 2p(0)$.

Remark 10. The concept of expectation is analogous to the physical concept of the *center of gravity* of a distribution of mass. Consider a discrete random variable η having probability mass function $p(x_i)$, $i \geq 1$. If we now imagine a weightless rod in which weights mass $p(x_i)$, $i \geq 1$, are located at the points x_i, $i \geq 1$, then the point at which the rod would be in balance is known as the center of gravity. For those readers acquainted with elementary statics it is now a simple matter to show that this point is at $E\eta$.

Example 57. Find $E\eta$ where η is the outcome when we roll a fair die.

Solution: Since

$$p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6},$$

we obtain that

$$E\eta = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}.$$

The reader should note that, for this example, the expected value of η is not a value that η could possibly assume. That is, rolling a die cannot possibly lead to an outcome of $\frac{7}{2}$.

Thus, even though we call $E\eta$ the expectation of η, it should not be interpreted as the value that we expect η to have but rather as the average value of η in a large number of repetitions of the experiment. That is, if we continually roll a fair die, then after a large number of rolls the average of all the outcomes will be approximately $\frac{7}{2}$.

Properties of Expectation of Random Variables

Property 1. If $\eta(\omega) = I_A(\omega)$ is the indicator function for the event A, then

$$EI_A(\omega) = P(A).$$

Proof: Since $p(1) = P(A)$, $p(0) = 1 - P(A)$, we have that

$$EI_A(\omega) = 1 \cdot P(A) + 0 \cdot (1 - P(A)) = P(A).$$

The proof is complete.

That is, the expected value of the indicator function for the event A is equal to the probability that A occurs.

Property 2. If $\eta(\omega)$ is a discrete random variable that takes on one of the values x_i, $i \geq 1$, with respective probabilities $p(x_i)$, then for any real–valued function $g(x)$

$$E\left(g(\eta(\omega))\right) = \sum_i g(x_i) \cdot p(x_i). \tag{63}$$

Proof: If $\eta(\omega) = x_i$ then $g(\eta(\omega)) = g(x_i)$ and

$$P\left(\omega\colon\ g(\eta(\omega)) = g(x_i)\right) = P(\omega\colon\ \eta = x_i) = p(x_i).$$

Therefore, the random variable $g(\eta(\omega))$ has the following Distribution Law

$g(x_1)$	$g(x_2)$	...	$g(x_n)$	...
$p(x_1)$	$p(x_2)$	...	$p(x_n)$	...

Hence, by definition, we get (63).

Property 3. If $\eta(\omega)$ is an absolutely continuous random variable with density function $f(x)$, then for any real–valued continuous function $g(x)$

$$E\left[g(\eta(\omega)\right] = \int_{-\infty}^{+\infty} g(x)\, f(x)\, dx. \tag{64}$$

We state Property 3 without proof.

Property 4. If a and b are constants, then

$$E[a\,\eta(\omega)+b] = a\cdot E\eta + b \tag{65}$$

for any random variable $\eta(\omega)$.

In other words, (65) says: "The expectation of a linear function of a single random variable is the linear function obtained by replacing the random variable by its expectation".

Corollary 1. (i) $E[a\eta] = a\cdot E\eta$ for any constant a,

(ii) if $\eta(\omega)=b$, then $Eb=b$, for any constant b, that is, the expectation of a constant is that constant.

Proof: Corollary 1 follows from (65) [(i) for $b=0$ and (ii) for $a=0$].

Remark 11. The expected value of a random variable $\eta(\omega)$, $E[\eta(\omega)]$ is also referred to as the mean or the first moment of $\eta(\omega)$.

One simple type of function whose expectation might be of interest is the power function $g(\eta)=\eta^k,\ k=1,2,3,\ldots.$

Definition 14. The quantity $E\eta^n$, $n\geq 1$ is called the nth moment of $\eta(\omega)$. By Properties 2 and 3 we note that

$$E\eta^n = \begin{cases} \sum\limits_i x_i^n\,p(x_i) & \text{if } \eta \text{ is a discrete random variable} \\ \int_{-\infty}^{+\infty} x^n\,f_\eta(x)\,dx & \text{if } \eta \text{ is an absolutely continuous random variable} \end{cases} \tag{66}$$

Property 5. If $\eta(\omega)\geq 0$ then $E\eta\geq 0$.

The proof is obvious.

For two–dimensional analog of Properties 2 and 3 which give the computational formulas for the expected value of a function of a random variable, suppose $\eta_1(\omega)$ and $\eta_2(\omega)$ are random variables and $g(x,y)$ is a real–valued continuous function of two variables. Then we have the following result (the proof is omitted).

Theorem 9. *If $\eta_1(\omega)$ and $\eta_2(\omega)$ have either a joint probability mass function $p(x,y)$ or a joint density function $f(x,y)$. Then*

$$E\left[g(\eta_1(\omega),\eta_2(\omega)\right] = \begin{cases} \sum_x \sum_y g(x,y)\, p(x,y) & \text{if} \quad (\eta_1(\omega),\eta_2(\omega)) \text{ is a discrete vector} \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y)\, f(x,y)\, dx\, dy & \text{if} \quad (\eta_1(\omega),\eta_2(\omega)) \text{ is absolutely continuous} \end{cases} \tag{67}$$

Remark 12. Properties 2 and 3 follows from Theorem 9 if we take $g(x,y) = g(x)$. Indeed, we can use properties $\sum_y p(x,y) = p_{\eta_1}(x)$ and (51).

Property 6. If $E\eta_1(\omega)$ and $E\eta_2(\omega)$ are finite, then

$$E\left(\eta_1(\omega) + \eta_2(\omega)\right) = E\eta_1(\omega) + E\eta_2(\omega). \tag{68}$$

Using Property 6, a simple induction proof shows that if $E\eta_i(\omega)$ is finite for all $i = 1, ..., n$, then

$$E\left(\eta_1(\omega) + ... + \eta_n(\omega)\right) = E\eta_1(\omega) + ... + E\eta_n(\omega). \tag{69}$$

Corollary 2. For any random variable $\eta(\omega)$ we have

$$E\left(\eta(\omega) - E\eta\right) = 0.$$

The proof follows from additivity of expectation and Corollary 1, (ii).

Property 7. If $\eta_1(\omega) \geq \eta_2(\omega)$ (that is for any outcome of the experiment, the value of the random variable $\eta_1(\omega)$ is greater than or equal the value of the random variable $\eta_2(\omega)$), then

$$E\eta_1 \geq E\eta_2.$$

Proof: $\eta_1(\omega) \geq \eta_2(\omega)$ implies $\eta_1(\omega) - \eta_2(\omega) \geq 0$ and therefore by Property 5 $E(\eta_1 - \eta_2) \geq 0$. By additivity of expectation

$$E(\eta_1 - \eta_2) = E\eta_1 - E\eta_2 \geq 0.$$

The proof is complete.

When one is dealing with an infinite collection of random variables $\eta_i(\omega)$, $i \geq 1$, each having a finite expectation, it is not necessarily true that

$$E\left[\sum_{i=1}^{\infty} \eta_i(\omega)\right] = \sum_{i=1}^{\infty} E\eta_i. \tag{70}$$

To determine when (70) is valid, we note that

$$\sum_{i=1}^{\infty} \eta_i(\omega) = \lim_{n\to\infty} \sum_{i=1}^{n} \eta_i(\omega)$$

and thus

$$\begin{aligned} E\left[\sum_{i=1}^{\infty} \eta_i(\omega)\right] &= E\left[\lim_{n\to\infty} \sum_{i=1}^{n} \eta_i(\omega)\right] \overset{?}{=} \\ \overset{?}{=} \lim_{n\to\infty} E\left[\sum_{i=1}^{n} \eta_i(\omega)\right] &= \lim_{n\to\infty} \sum_{i=1}^{n} E\eta_i = \sum_{i=1}^{\infty} E\eta_i. \end{aligned} \tag{71}$$

Hence Equation (70) is valid whenever we are justified in interchanging the expectation and limit operations in (71). Although, in general, this interchange is not justified, it can be shown to be valid in two important special cases:

1. $\eta_i(\omega)$ are all nonnegative random variables (that is, $P(\omega\colon\ \eta_i(\omega) \geq 0) = 1$ for all i).

2. $\sum\limits_{i=1}^{\infty} E|\eta_i| < \infty$.

Property 8. If η_1 and η_2 are independent, then

$$E\left[\eta_1 \cdot \eta_2\right] = E\eta_1 \cdot E\eta_2. \tag{72}$$

Corollary 3. If η_1 and η_2 are independent, then for any continuous functions $\varphi_1(x)$ and $\varphi_2(x)$

$$E\left[\varphi_1(\eta_1) \cdot \varphi_2(\eta_2)\right] = E\varphi_1(\eta_1) \cdot E\varphi_2(\eta_2). \tag{73}$$

Proof: The proof follows from Property 8 and Theorem 9.

APPENDIX I:

Proofs of Properties 4, 6, 8

Proof of Property 4: **(i) if $\eta(\omega)$ is a discrete, then by Property 2 we get**

$$E[a\,\eta(\omega)+b]=\sum_i(ax_i+b)\cdot p(x_i)=a\sum_i x_i\cdot p(x_i)+b\sum_i p(x_i)=a\cdot E\eta+b.$$

Above we used (38) and the definition of expectation for discrete random variable.
(ii) if $\eta(\omega)$ is an absolutely continuous, then by Property 3 we obtain

$$E(a\,\eta(\omega)+b)=\int_{-\infty}^{+\infty}(ax+b)\,f(x)\,dx=a\int_{-\infty}^{+\infty}x\,f(x)\,dx+b=a\cdot E\eta+b.$$

Proof of Property 6: **Using Theorem 9 for $g(x,y)=x+y$ we obtain:**
(i) in absolutely continuous case

$$E(\eta_1(\omega)+\eta_2(\omega))=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x+y)\,f(x,y)\,dx\,dy=$$
$$=\int_{-\infty}^{+\infty}x\,dx\int_{-\infty}^{+\infty}f(x,y)\,dy+\int_{-\infty}^{+\infty}y\,dy\int_{-\infty}^{+\infty}f(x,y)\,dx=$$
$$=\int_{-\infty}^{+\infty}x\,f_{\eta_1}(x)\,dx+\int_{-\infty}^{+\infty}y\,f_{\eta_2}(y)\,dy=E\eta_1(\omega)+E\eta_2(\omega);$$

and
(ii) in discrete case

$$E(\eta_1(\omega)+\eta_2(\omega))=\sum_{x_i}\sum_{y_j}(x_i+y_j)\,p(x_i,y_j)=$$
$$=\sum_{x_i}x_i\sum_{y_j}p(x_i,y_j)+\sum_{y_j}y_j\sum_{x_i}p(x_i,y_j)=\sum_{x_i}x_i\,p_{\eta_1}(x_i)+\sum_{y_j}y_j\,p_{\eta_2}(y_j)=E\eta_1(\omega)+E\eta_2(\omega).$$

The proof is complete.

Proof of Property 8: **Using Theorem 9 for $g(x,y)=x\cdot y$ we obtain:**
(i) in absolutely continuous case

$$E(\eta_1(\omega)\cdot\eta_2(\omega))=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x\cdot y\,f(x,y)\,dx\,dy=$$
$$(\textbf{by independence of }\eta_1\textbf{ and }\eta_2)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x\cdot y\,f_{\eta_1}(x(\,f_{\eta_2}(y)\,dx\,dy=$$
$$=\int_{-\infty}^{+\infty}x\,f_{\eta_1}(x)\,dx\cdot\int_{-\infty}^{+\infty}y\,f_{\eta_2}(y)\,dy=E(\eta_1)\cdot E(\eta_2);$$

and
(ii) in discrete case

$$E(\eta_1(\omega)\cdot\eta_2(\omega))=\sum_{x_i}\sum_{y_j}x_i\cdot y_j\,p(x_i,y_j)=$$
$$=\sum_{x_i}x_i\,p_{\eta_1}(x_i)\cdot\sum_{y_j}y_j\,p_{\eta_2}(y_j)=E\eta_1(\omega)\cdot E\eta_2(\omega).$$

The proof is complete.

Example 58. **Let us give an example which states that from (72) does not follow that η_1 and η_2 are independent, i. e. we construct an example of two dependent random variables η_1 and η_2 for which $E[\eta_1 \cdot \eta_2] = E\eta_1 \cdot E\eta_2$.**
Let $\eta_1(\omega)$ and $\zeta(\omega)$ be independent random variables and $E\eta_1 = E\zeta = 0$. Denote by η_2 the following random variable:

$$\eta_2(\omega) = \eta_1(\omega) \cdot \zeta(\omega).$$

It is obvious that η_1 and η_2 are dependent (if at outcome ω_0 random variable $\eta_1 = 0$ implies that $\eta_2(\omega_0) = 0$) and

$$E[\eta_1 \cdot \eta_2] = E\left[\eta_1^2 \cdot \zeta\right] = (\textbf{by Theorem 8 and Property 8}) = E\eta_1^2 \cdot E\zeta = E\eta_1^2 \cdot 0 = 0.$$

LECTURE 13

§25. VARIANCE

Given a random variable $\eta(\omega)$ along with its distribution function $F(x)$, it would be extremely useful if we were able to summarize the essential properties of $F(x)$ by certain suitable defined measures. One such measure would be $E\eta$, the expected value of η. However, although $E\eta$ yields the weighted average of the possible values of $\eta(\omega)$, it does not tell us anything about the variation, or spread, of these values. For instance, although discrete random variables $\eta_1(\omega)$, $\eta_2(\omega)$ and $\eta_3(\omega)$ determined by the following formulae

$$\eta_1(\omega) \equiv 0,$$

$$\eta_2(\omega) = \begin{cases} 1 & \text{if} \quad \omega \in A \\ -1 & \text{if} \quad \omega \notin A \end{cases},$$

$$\eta_3(\omega) = \begin{cases} +1000 & \text{if} \quad \omega \in A \\ -1000 & \text{if} \quad \omega \notin A \end{cases}$$

$\left(\text{where we take } P(A) = P(\overline{A}) = \dfrac{1}{2}\right)$ all have the same expectation — namely, 0 — there is much greater spread in the possible value of η_2 than in those of $\eta_1(\omega)$ (which is a constant) and in the possible value of $\eta_3(\omega)$ than in those of $\eta_2(\omega)$.

As we expect $\eta(\omega)$ to take on values around its mean $E\eta$, it would appear that a reasonable way of measuring the possible variation of η would be to look at how far apart η would be from its mean on the average. One possible way to measure this would be to consider the quantity $E|\eta - E\eta|$. However, it turns out to be mathematically inconvenient to deal with this quantity, and so a more tractable quantity is usually considered — namely, the expectation of the square of the difference between η and its mean. We thus have the following definition.

Definition 15. If $\eta(\omega)$ is a random variable, then the variance of η, denoted by $\mathrm{Var}\,(\eta)$, is defined by

$$\mathrm{Var}\,(\eta) = E\,[\eta - E\eta]^2\,. \tag{74}$$

An alternative formula for $\mathrm{Var}\,(\eta)$ is derived as follows:

$$\mathrm{Var}\,(\eta) = E\,[\eta - E\eta]^2 = E\left[\eta^2 - 2\eta\,E\eta + (E\eta)^2\right] = E\left[\eta^2\right] - 2(E\eta)^2 + (E\eta)^2. = E\left[\eta^2\right] - (E\eta)^2.$$

That is

$$\operatorname{Var}(\eta) = E\left[\eta^2\right] - [E\eta]^2 . \tag{75}$$

In words, the variance of $\eta(\omega)$ is equal to the expected value of η^2 minus the square of its expected value. This is, in practice, often the easiest way to compute $\operatorname{Var}(\eta)$.

Remark 13. The square root of the $\operatorname{Var}(\eta)$ is called the *standard deviation* of η, and we denote it by $SD(\eta)$. That is,

$$SD(\eta) = \sqrt{\operatorname{Var}(\eta)}$$

By Properties 2 and 3 of Expectation into which we substitute $g(x) = (x - E\eta)^2$ we obtain the computational formulae for $\operatorname{Var}(\eta)$

$$\operatorname{Var}(\eta) = \begin{cases} \sum_i (x_i - E\eta)^2\, p(x_i) & \text{if } \eta \text{ is a discrete random variable} \\ \int_{-\infty}^{+\infty} (x - E\eta)^2\, f_\eta(x)\, dx & \text{if } \eta \text{ is absolutely continuous} \end{cases} \tag{76}$$

or

$$\operatorname{Var}(\eta) = \begin{cases} \sum_i x_i^2\, p(x_i) - \left(\sum_i x_i\, p(x_i)\right)^2 & \text{if } \eta \text{ is discrete} \\ \int_{-\infty}^{+\infty} x^2\, f_\eta(x)\, dx - \left(\int_{-\infty}^{+\infty} x\, f_\eta(x)\, dx\right)^2 & \text{if } \eta \text{ is absolutely continuous} \end{cases} \tag{77}$$

Properties of the Variance of Random Variables

Property 1. For any random variable η

$$\operatorname{Var}(\eta) \geq 0.$$

Proof: As the random variable $\zeta(\omega) = (\eta(\omega) - E\eta)^2$ is nonnegative, then by Property 5 of Expectation we conclude

$$\operatorname{Var}(\eta) = E\left[\eta - E\eta\right]^2 \geq 0.$$

The proof is complete.

Corollary 4. For any random variable $\eta(\omega)$ we have

$$E\left[\eta^2\right] \geq (E\eta)^2 .$$

The proof immediately follows from Property 1 and (75).

Markov's and Chebyshev's Inequalities

Theorem 10. (Markov's Inequality). *If $\eta(\omega)$ is a random variable that takes only nonnegative values, then for any value $\varepsilon > 0$*

$$P\{\omega:\ \eta(\omega) \geq \varepsilon\} \leq \frac{E\eta}{\varepsilon}. \tag{78}$$

The proof of Theorem 10 can be found in the Appendix.

Corollary 5. If $\eta(\omega)$ is a random variable, then for any value $\varepsilon > 0$

$$P\{\omega:\ |\eta(\omega)| \geq \varepsilon\} \leq \frac{E|\eta|}{\varepsilon}. \tag{79}$$

The proof is obvious.

Corollary 6. If $\eta(\omega)$ is a random variable, then for any value $\varepsilon > 0$

$$P\{\omega:\ |\eta(\omega)| \geq \varepsilon\} \leq \frac{E|\eta|^r}{\varepsilon^r}, \qquad r = 1, 2, ... \tag{80}$$

Proof: As for any natural r we have

$$\eta(\omega) \geq \varepsilon \Longrightarrow \eta^r(\omega) \geq \varepsilon^r,$$

hence we obtain the inequality

$$P\{\omega:\ |\eta(\omega)| \geq \varepsilon\} \leq P\{\omega:\ |\eta^r(\omega)| \geq \varepsilon^r\}.$$

Now the proof follows from Corollary 5.

Corollary 7. (Chebyshev's inequality) If $\eta(\omega)$ is a random variable, then for any value $\varepsilon > 0$

$$P\{\omega:\ |\eta(\omega) - E\eta| \geq \varepsilon\} \leq \frac{\mathrm{Var}\,(\eta)}{\varepsilon^2}, \tag{81}$$

The proof follows from Corollary 6 with $r = 2$ and for random variable $\eta(\omega) - E\eta$.

Lemma 6. *If $\eta(\omega)$ is a random variable that takes only nonnegative values and $E\eta = 0$, then*

$$P\{\omega\colon\ \eta(\omega) = 0\} = 1.$$

The proof of Lemma 6 can be found in the Appendix.

Property 2. If $\mathrm{Var}\,(\eta) = 0$, then

$$P\{\omega\colon\ \eta(\omega) = E\eta\} = 1.$$

In other words, the only random variables having variance equal to 0 are those that are constant with probability 1.

Proof: As $\zeta = (\eta - E\eta)^2$ is nonnegative random variable and $E\zeta = \mathrm{Var}\,(\eta) = 0$, therefore by Lemma 6 we get $\zeta = (\eta - E\eta)^2 = 0$, that is Property 2.

Corollary 8. $\mathrm{Var}\,(\eta) = 0$ if and only if η is a constant.

The proof is obvious.

Property 3. For any constants a and b we have the following formula

$$\mathrm{Var}\,(a\eta + b) = a^2 \cdot \mathrm{Var}\,(\eta). \tag{82}$$

Proof: We have

$$\mathrm{Var}\,(a\eta + b) = E\,[a\eta + b - aE\eta - b]^2 = E\left[a^2(\eta - E\eta)^2\right] = a^2\,E\left[(\eta - E\eta)^2\right] = a^2 \cdot \mathrm{Var}\,(\eta).$$

Property 4. For any two random variables $\eta_1(\omega)$ and $\eta_2(\omega)$, we have

$$\mathrm{Var}\,(\eta_1 + \eta_2) = \mathrm{Var}\,(\eta_1) + \mathrm{Var}(\eta_2) + 2\,[E(\eta_1 \cdot \eta_2) - E\eta_1 \cdot E\eta_2]\,. \tag{83}$$

Proof: Using equation (75), we get

$$\mathrm{Var}\,(\eta_1 + \eta_2) = E\left[(\eta_1 + \eta_2)^2\right] - [E(\eta_1 + \eta_2)]^2 = \text{(by additivity of expectation)}$$

$$= E\left[\eta_1^2 + \eta_2^2 + 2\eta_1 \cdot \eta_2\right] - [E\eta_1 + E\eta_2]^2 = \text{(by additivity of expectation)}$$

$$= E\left[\eta_1^2\right] + E\left[\eta_2^2\right] + 2E\,[\eta_1 \cdot \eta_2] - [E\eta_1]^2 - E\,[\eta_2]^2 - 2E\eta_1 \cdot E\eta_2 =$$

$$= \mathrm{Var}\,(\eta_1) + \mathrm{Var}(\eta_2) + 2\,[E(\eta_1 \cdot \eta_2) - E\eta_1 \cdot E\eta_2]\,.$$

The proof is complete.

Corollary 9. If η_1 and η_2 are independent, then

$$\mathrm{Var}\,(\eta_1+\eta_2)=\mathrm{Var}\,(\eta_1)+\mathrm{Var}(\eta_2). \tag{84}$$

The proof follows from Property 4 and (72).

However, the converse is not true. A simple example of two dependent random variables η_1 and η_2 for which we have (84), we have constructed in Example 58.

§26. COVARIANCE

Just as the expected value and the variance of a single random variable give us information about this random variable, so does the covariance between two random variables gives us information about the relationship between the random variables.

Definition 16. The covariance between $\eta_1(\omega)$ and $\eta_2(\omega)$, denoted by $\mathrm{Cov}\,(\eta_1,\eta_2)$, is defined by

$$\mathrm{Cov}\,(\eta_1,\eta_2)=E\left[(\eta_1-E\eta_1)\cdot(\eta_2-E\eta_2)\right]. \tag{85}$$

Now, the right–hand side of (85) can be written as

$$\mathrm{Cov}\,(\eta_1,\eta_2)=E\left[\eta_1\eta_2-E\eta_1\cdot\eta_2-\eta_1\cdot E\eta_2+E\eta_1\cdot E\eta_2\right]=\text{(by additivity of expectation)}$$

$$=E\left[\eta_1\cdot\eta_2\right]-E\eta_1\cdot E\eta_2-E\eta_1\cdot E\eta_2+E\eta_1\cdot E\eta_2=E\left[\eta_1\cdot\eta_2\right]-E\eta_1\cdot E\eta_2.$$

Therefore

$$\mathrm{Cov}\,(\eta_1,\eta_2)=E\left[\eta_1\cdot\eta_2\right]-E\eta_1\cdot E\eta_2. \tag{86}$$

Note that if η_1 and η_2 are independent then, by (72), it follows that $\mathrm{Cov}\,(\eta_1,\eta_2)=0$. However, the converse is not true. A simple example of two dependent random variables η_1 and η_2 having zero covariance is obtained by letting η_1 be a random variable such that

-1	0	1
$1/3$	$1/3$	$1/3$

and define

$$\eta_2(\omega)=\begin{cases}0 & \text{if}\quad \eta_1\neq 0\\ 1 & \text{if}\quad \eta_1=0\end{cases}$$

Now, $\eta_1\cdot\eta_2=0$, so $E\left[\eta_1\cdot\eta_2\right]=0$. Also, $E\eta_1=0$ and thus

$$\mathrm{Cov}\,(\eta_1,\eta_2)=E\left[\eta_1\cdot\eta_2\right]-E\eta_1\cdot E\eta_2=0.$$

However, η_1 and η_2 are clearly not independent.

APPENDIX:

Proofs of Lemma 6 and Theorem 10

Proof of Lemma 6: **The following equation**

$$\{\omega\colon\ \eta(\omega)>0\}=\bigcup_{n=1}^{\infty}\left\{\omega\colon \eta(\omega)>\frac{1}{n}\right\}$$

holds. As $\left\{\omega\colon \eta(\omega)>\frac{1}{n}\right\}$ **is a nondecreasing sequence of events by Property 12 of probability we obtain**

$$P\{\omega\colon\ \eta(\omega)>0\}=\lim_{n\to+\infty}P\left\{\omega\colon \eta(\omega)>\frac{1}{n}\right\}\leq (\textbf{by Markov's inequality})\leq$$
$$\leq\lim_{n\to+\infty}[n\cdot E\eta]=\lim_{n\to+\infty}[n\cdot 0]=0.$$

The proof is complete.

Proof of Theorem 10: **It is not difficult to verify that for any** $\varepsilon>0$ **the inequality**

$$\varepsilon\cdot I_{\{\omega\colon\ \eta(\omega)\geq\varepsilon\}}\leq\eta(\omega)$$

holds.
Therefore, by Property 7 of Expectation we obtain

$$\varepsilon\, EI_{\{\omega\colon\ \eta(\omega)\geq\varepsilon\}}\leq E\eta.$$

Applying Property 1 of Expectation we obtain (78). The proof is complete.

L E C T U R E 14

TYPICAL PROBLEMS OF MID–TERM EXAMINATION-1

PROBLEM 1. An urn contains 10 balls, bearing numbers 0 to 9. 3 balls are randomly drawn without replacement. By placing the numbers in a row in the order in which they are drawn, an integer 0 to 999 is formed. What is the probability that the number thus formed is divisible by 51? (*Note:* regard 0 as being divisible by 51).

PROBLEM 2. A club has 12 members (6 married couples). Four members are randomly selected to form the club executive. Find the probability that the executive contains no married couple.

PROBLEM 3. A town has 2 fire engines operating independently. The probability that a specific engine is available when needed is 0.96. What is the probability that a fire engine is available when needed?

PROBLEM 4. Let a number be chosen from the integers 11 to 111111 in such a way that each of these numbers are equally likely to be chosen. What is the probability that the number chosen will be divisible either on 7 or on 11?

PROBLEM 5. Eight red blocks and four blue blocks are arranged at random in a row. What is the probability that all four blue blocks occur together in a row with no red blocks in between?

PROBLEM 6. Let A and B be events with probabilities $P(A) = 5/6$ and $P(B) = 1/4$. What is the maximum and minimum values of $P(A \cap B)$. Find corresponding bounds for $P(A \cup B)$.

ANSWERS

1. $\frac{1}{45} \approx 0.0(2)$.

2. $\frac{16}{33} = 0.(48)$.

3. 0.9984.

4. $\frac{24530}{111101} = 0.22079$.

5. $\frac{1}{55} = 0.0(18)$.

6. $\frac{1}{12} \leq P(A \cap B) \leq \frac{1}{4}$; $\frac{5}{6} \leq PA \cup B) \leq 1$.

LECTURE 15

Properties of Covariance of Random Variables

Property 1. For any two random variables η_1 and η_2

$$\operatorname{Cov}(\eta_1, \eta_2) = \operatorname{Cov}(\eta_2, \eta_1).$$

The proof is obvious.

Property 2.

$$\operatorname{Cov}(\eta, \eta) = \operatorname{Var}(\eta).$$

Proof:

$$\operatorname{Cov}(\eta, \eta) = E\left[(\eta - E\eta)\cdot(\eta - E\eta)\right] = E\left[(\eta - E\eta)^2\right] = \operatorname{Var}(\eta).$$

The proof is complete.

Property 3. For any two random variables η_1 and η_2 and for any constant a we have

$$\operatorname{Cov}(a\cdot\eta_1, \eta_2) = a\cdot\operatorname{Cov}(\eta_1, \eta_2).$$

Proof: We have

$$\operatorname{Cov}(a\cdot\eta_1, \eta_2) = E\left[(a\,\eta_1 - E(a\cdot\eta_1))\cdot(\eta_2 - E\eta_2)\right] = E\left[a\,(\eta_1 - E\eta_1)\cdot(\eta_2 - E\eta_2)\right] = a\cdot\operatorname{Cov}(\eta_1, \eta_2).$$

Property 4. For any random variables $\eta_1, ..., \eta_n$ and $\zeta_1, ..., \zeta_m$ we have

$$\operatorname{Cov}\left(\sum_{i=1}^{n}\eta_i, \sum_{j=1}^{m}\zeta_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m}\operatorname{Cov}(\eta_i, \zeta_j), \tag{87}$$

that is, the covariance is finitely additive operation.

Proof: Indeed,

$$\operatorname{Cov}\left(\sum_{i=1}^{n}\eta_i, \sum_{j=1}^{m}\zeta_j\right) = E\left[\left(\sum_{i=1}^{n}\eta_i - \sum_{i=1}^{n}E\eta_i\right)\left(\sum_{j=1}^{m}\zeta_j - \sum_{j=1}^{m}E\zeta_j\right)\right] =$$

$$= E\left[\sum_{i=1}^{n}\sum_{j=1}^{m}(\eta_i - E\eta_i)\cdot(\zeta_j - E\zeta_j)\right] = \sum_{i=1}^{n}\sum_{j=1}^{m}\operatorname{Cov}(\eta_i, \zeta_j),$$

where the last equality follows because the expected value of a sum of random variables is equal to the sum of the expected values.

Corollary 10. For any sequence $\eta_1, ..., \eta_n$ we have

$$\text{Var}\left(\sum_{i=1}^{n} \eta_i\right) = \sum_{i=1}^{n} \text{Var}\ (\eta_i) + 2 \sum_{i<j} \text{Cov}(\eta_i, \eta_j) \tag{88}$$

Proof: It follows from Properties 2 and 3, upon taking $\zeta_j = \eta_j$, $j = 1, ..., n$, that

$$\text{Var}\left(\sum_{i=1}^{n} \eta_i\right) = \text{Cov}\left(\sum_{i=1}^{n} \eta_i, \sum_{j=1}^{n} \eta_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \text{Cov}\,(\eta_i, \eta_j) =$$
$$= \sum_{i=1}^{n} \text{Var}\,(\eta_i) + \sum_{i \neq j} \text{Cov}\,(\eta_i, \eta_j).$$

Since each pair of indices, $i, j, i \neq j$, appear twice, we obtain (88). The proof is complete.

If $\eta_1, ..., \eta_n$ are pairwise independent, in that η_i and η_j are independent for $i \neq j$, then $\text{Cov}\,(\eta_i, \eta_j) = 0$ and (88) reduces to

$$\text{Var}\left(\sum_{i=1}^{n} \eta_i\right) = \sum_{i=1}^{n} \text{Var}\ (\eta_i)\,. \tag{89}$$

(compare with (83) and (84)).

Definition 17. The correlation of two random variables η_1 and η_2, denoted by $\rho(\eta_1, \eta_2)$, is defined, as long as η_1 and η_2 are not constant, by the following formula

$$\rho(\eta_1, \eta_2) = \frac{\text{Cov}\,(\eta_1, \eta_2)}{\sqrt{\text{Var}\,(\eta_1) \cdot \text{Var}\,(\eta_2)}}. \tag{90}$$

Properties of Correlation of Random Variables

The correlation has the following three properties:

Property 1. For any two random variables η_1 and η_2 its correlation satisfies the inequality

$$|\rho(\eta_1, \eta_2)| \leq 1.$$

Property 2. If η_1 and η_2 are independent, then

$$\rho(\eta_1, \eta_2) = 0.$$

Property 3. $\eta_2 = a \cdot \eta_1 + b$ for some constants a and b if and only if

$$\rho(\eta_1, \eta_2) = \pm 1,$$

moreover, the sign of a coincides with the sign of ρ.

Property 2 is obvious. To prove Properties 1 and 3 we are going to get the following additional relationship:

$$\mathrm{Var}\left(\frac{\eta_1}{\sqrt{\mathrm{Var}\,(\eta_1)}} \pm \frac{\eta_2}{\sqrt{\mathrm{Var}\,(\eta_2)}}\right) = 2[1 \pm \rho(\eta_1, \eta_2)]. \tag{91}$$

Indeed, by (88) we have

$$\mathrm{Var}\left(\frac{\eta_1}{\sqrt{\mathrm{Var}\,(\eta_1)}} \pm \frac{\eta_2}{\sqrt{\mathrm{Var}\,(\eta_2)}}\right) = \frac{\mathrm{Var}\,(\eta_1)}{\mathrm{Var}\,(\eta_1)} + \frac{\mathrm{Var}\,(\eta_2)}{\mathrm{Var}\,(\eta_2)} \pm 2\frac{\mathrm{Cov}\,(\eta_1, \eta_2)}{\sqrt{\mathrm{Var}(\eta_1) \cdot \mathrm{Var}\,(\eta_2)}} = 2[1 \pm \rho(\eta_1, \eta_2)].$$

Therefore, by Property 1 of Variance, we get

$$2[1 \pm \rho(\eta_1, \eta_2)] \geq 0.$$

The proof of Property 1 is complete. Let us prove Property 3.
As $\mathrm{Var}\,(\zeta) = 0$ implies that $\zeta(\omega)$ is constant with Probability 1 (see Corollary 8), we see from (91), that $\rho(\eta_1, \eta_2) = +1$ implies that

$$\frac{\eta_1}{\sqrt{\mathrm{Var}\,(\eta_1)}} - \frac{\eta_2}{\sqrt{\mathrm{Var}\,(\eta_2)}} = \text{const}$$

and therefore $\eta_2 = a\eta_1 + b$, where $a > 0$ and similarly, $\rho = -1$ implies that $\eta_2 = a\eta_1 + b$, where $a < 0$. We leave it as an exercise for the reader to show that the reverse is also true: that if $\eta_2 = a\eta_1 + b$, then $\rho(\eta_1, \eta_2)$ is either $+1$ or -1, depending on the sign of a. The proof of Property 3 is complete.
The correlation is a measure of the degree of linearity between η_1 and η_2. A value of $\rho(\eta_1, \eta_2)$ near $+1$ or -1 indicates a high degree of linearity between η_1 and η_2, whereas a value near 0 indicates a lack of such linearity. A positive value of $\rho(\eta_1, \eta_2)$ indicates that $\eta_2(\omega)$ tends to increase when η_1 does, whereas a negative value indicates that η_2 tends to decrease when η_1 increases.

Definition 18. If $\rho(\eta_1, \eta_2) = 0$, then η_1 and η_2 are said to be *uncorrelated*.

§27. BACK TO THE CLASSICAL DISTRIBUTIONS

I. Normal Random Variable. We shall now restrict our discussion to the famous Normal (or Gaussian) distribution, of great importance in statistics and probability theory. We say that $\eta(\omega)$ is a normal random variable, or simply that η is normally distributed, with parameters a and $\sigma > 0$ if the density function $f(x)$ of η is given by (42) (or the distribution function has the form (31)):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right).$$

This density function is a bell–shaped curve that is symmetric with respect to the vertical line $x = a$ (since $f(x) = f(2a - x)$). Furthermore, simple differentiation yields:

$$f'(x) = -\frac{x-a}{\sigma^2} f(x), \qquad f''(x) = \frac{1}{\sigma^4}\left[(x-a)^2 - \sigma^2\right] f(x).$$

It follows that the curve $f(x)$ has maximum at $x = a$, equal to $f(a) = \dfrac{1}{\sigma\sqrt{2\pi}}$, and two inflection points at $a + \sigma$ and $a - \sigma$.

Let us calculate $E\eta$ and $\mathrm{Var}\,(\eta)$. By definition (see (62)) we have

$$E\eta = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right) dx.$$

Let us make a change of variable:

$$y = \frac{x-a}{\sigma}, \qquad \sigma\, dy = dx.$$

Therefore

$$E\eta = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma\cdot y + a) \exp\left(-\frac{y^2}{2}\right) dy = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y \exp\left(-\frac{y^2}{2}\right) dy + \frac{a}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{y^2}{2}\right) dy.$$

By symmetry, the first integral must be 0, so

$$E\eta = \frac{a}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{y^2}{2}\right) dy = a.$$

Since $E\eta = a$, by (76), we have that

$$\mathrm{Var}\,(\eta) = \frac{1}{\sigma\,\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-a)^2 \,\exp\left(-\frac{(x-a)^2}{2\sigma^2}\right) dx.$$

Let us make a change of variable:

$$y = \frac{x-a}{\sigma}, \qquad \sigma\, dy = dx.$$

Therefore

$$\mathrm{Var}\,(\eta) = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^2 \,\exp\left(-\frac{y^2}{2}\right) dy = (\text{using integration by parts})$$
$$\frac{\sigma^2}{\sqrt{2\pi}} \left[-y\,\exp\left(-\frac{y^2}{2}\right)\Bigg|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} \exp\left(-\frac{y^2}{2}\right) dy\right] = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{y^2}{2}\right) dy = \sigma^2.$$

Therefore, we have obtained that

$$E\eta = a \qquad \mathrm{Var}\,(\eta) = \sigma^2.$$

If a random variable $\eta(\omega)$ is known to be normally distributed with mean a and variance σ^2, then for brevity one often writes η is $\mathcal{N}(a, \sigma^2)$.

L E C T U R E 16

An important fact about normal random variables is the following Lemma.

Lemma 7. *If $\eta(\omega)$ is normally distributed $\mathcal{N}(a, \sigma^2)$, then $\zeta(\omega) = \alpha \cdot \eta(\omega) + \beta$, $\alpha \neq 0$ is normally distributed $\mathcal{N}(\alpha a + \beta, \alpha^2 \sigma^2)$.*

Proof: This follows because $F_\zeta(x)$, the distribution function of the random variable ζ, is given by the formula

$$F_\zeta(x) = P\{\omega\colon\ \zeta(\omega) < x\} = P\{\omega\colon\ \alpha \cdot \eta(\omega) + \beta < x\} = P\{\omega\colon\ \alpha \cdot \eta(\omega) < x - \beta\} =$$
$$= \begin{cases} F_\eta\left(\dfrac{x-\beta}{\alpha}\right) & \text{if} \quad \alpha > 0 \\ P\left\{\omega\colon\ \eta(\omega) \geq \dfrac{x-\beta}{\alpha}\right\} = 1 - F_\eta\left(\dfrac{x-\beta}{\alpha}\right) & \text{if} \quad \alpha < 0 \end{cases}.$$

Therefore, density function of ζ is given by

$$f_\zeta(x) = \begin{cases} \dfrac{1}{\alpha} \cdot f_\eta\left(\dfrac{x-\beta}{\alpha}\right) & \text{if } \alpha > 0 \\ -\dfrac{1}{\alpha} \cdot f_\eta\left(\dfrac{x-\beta}{\alpha}\right) & \text{if } \alpha < 0 \end{cases} = \frac{1}{|\alpha|} \cdot f_\eta\left(\frac{x-\beta}{\alpha}\right) = \frac{1}{|\alpha| \cdot \sigma\sqrt{2\pi}} \exp\left(-\frac{\left(\frac{x-\beta}{\alpha} - a\right)^2}{2\sigma^2}\right)$$

The proof is complete.

An important implication of the preceding result is that if $\eta(\omega)$ is normally distributed with parameters a and σ^2, then

$$\zeta(\omega) = \frac{\eta - a}{\sigma}$$

is normally distributed with parameters 0 and 1. Such a random variable ζ is said to have the *standard normal distribution*.

Its density function has the form

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

It is traditional to denote the distribution function of the standard normal random variable by $\Phi(x)$. That is,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{y^2}{2}\right) dy. \tag{99}$$

The values of $\Phi(x)$ for nonnegative x are given in the Table 5.1 of the Textbook of Sheldon Ross (page 214) or Table 3 of the Textbook of Richard Johnson (page 586). For negative values of x, $\Phi(x)$ can be obtained from the following property of $\Phi(x)$:

$$\Phi(-x) = 1 - \Phi(x), \qquad \text{for any } x. \tag{100}$$

Proof: The proof of equation (100) follows from the following chain of implications:

$$\Phi(-x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-x} \exp\left(-\frac{y^2}{2}\right) dy = (\text{by making a change of variable } y = -z) =$$
$$= \frac{1}{\sqrt{2\pi}} \int_{x}^{+\infty} \exp\left(-\frac{z^2}{2}\right) dz = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{z^2}{2}\right) dz = 1 - \Phi(x).$$

Formula (100) states that if ζ is the standard normal random variable, then

$$P\{\omega:\ \zeta(\omega) < -x\} = P\{\omega:\ \zeta(\omega) \geq x\}, \qquad \text{for any real } x.$$

Since $\zeta(\omega) = \dfrac{\eta - a}{\sigma}$ is the standard normal random variable whenever $\eta(\omega)$ is normally distributed with parameters a and σ^2, it follows that the distribution function of $\eta(\omega)$ can be expressed as

$$F_\eta(x) = P\{\omega:\ \eta(\omega) < x\} = P\left\{\omega:\ \zeta(\omega) < \frac{x-a}{\sigma}\right\} = \Phi\left(\frac{x-a}{\sigma}\right).$$

Thus, to find the probability that random variable $\eta(\omega)$ will take on a value less than x, we look up $\Phi\left(\dfrac{x-a}{\sigma}\right)$ in the Table for $\Phi(x)$. Also, if we want to find the probability that a random variable having the normal distribution with mean a and the variance σ^2 will take on a value between x_1 and x_2, we have only calculate the probability that a random variable having the standard normal distribution will take on a value between $\dfrac{x_1-a}{\sigma}$ and $\dfrac{x_2-a}{\sigma}$. Therefore,

$$P\{\omega:\ x_1 < \eta(\omega) < x_2\} = \Phi\left(\frac{x_2-a}{\sigma}\right) - \Phi\left(\frac{x_1-a}{\sigma}\right). \tag{101}$$

Example 59. If $\eta(\omega)$ is the standard normal random variable, find

(a) $P\{\omega:\ 0.87 < \eta(\omega) < 1.28\}$,

(b) $P\{\omega:\ -0.34 < \eta(\omega) < 0.62\}$,

(c) $P\{\omega:\ \eta(\omega) > 0.85\}$,

(d) $P\{\omega:\ \eta(\omega) > -0.65\}$.

Solution: Looking up necessary values in the Table for standard normal distribution, we get

(a) $P\{\omega:\ 0.87 < \eta(\omega) < 1.28\} = \Phi(1.28) - \Phi(0.87) \approx 0.8997 - 0.8078 = 0.0919;$

(b) $P\{\omega:\ -0.34 < \eta(\omega) < 0.62\} =$

$$= \Phi(0.62) - \Phi(-0.34) = \Phi(0.62) - (1 - \Phi(0.34)) \approx 0.7324 - (1 - 0.6331) = 0.3655;$$

(c) $P\{\omega:\ \eta(\omega) > 0.85\} = 1 - \Phi(0.85) \approx 1 - 0.8023 = 0.1977;$

(d) $P\{\omega:\ \eta(\omega) > -0.65\} = 1 - \Phi(-0.65) = 1 - (1 - \Phi(0.65)) = \Phi(0.65) \approx 0.7422.$

Example 60. A random variable $\eta(\omega)$ is $\mathcal{N}(1,000, 2,500)$, that is normally distributed with mean $a = 1,000$ and variance $\sigma^2 = 2,500$. We have to find the probability that $\eta(\omega)$ is between 900 and $1,050$.

Solution: It follows from (101) that

$$P\{\omega:\ 900 \le \eta(\omega) \le 1,050\} = \Phi(1) - \Phi(-2).$$

Using (100), we conclude from the Table for $\Phi(x)$ that

$$P\{\omega:\ 900 \le \eta(\omega) \le 1,050\} = \Phi(1) + \Phi(2) - 1 \approx 0.819.$$

Example 61. Let $\eta(\omega)$ is $\mathcal{N}(2,9)$, that is normally distributed with mean $a = 2$ and variance $\sigma^2 = 9$. We have to find the probability that $\eta(\omega)$ is between 2 and 5.

Solution: Using (101)

$$P\{\omega:\ 2 \le \eta(\omega) \le 5\} = \Phi(1) - \Phi(0) = \Phi(1) - 0.5 \approx 0.841 - 0.5 = 0.341.$$

Here we used (100) for $x = 0$, that is $\Phi(0) = 1 - \Phi(0)$, or $2\,\Phi(0) = 1$ and the Table for Standard normal distribution function $\Phi(x)$.

Example 62. If $\eta(\omega)$ is a normal random variable with parameters $a = 3$ and $\sigma^2 = 9$, find

(a) $P\{\omega:\ 2 < \eta(\omega) < 5\},$

(b) $P\{\omega:\ \eta(\omega) > 0\},$

(c) $P\{\omega:\ |\eta(\omega) - 3| > 6\}.$

Solution: (a) $P\{\omega\colon\ 2<\eta(\omega)<5\}=$

$$=P\left\{\omega\colon\ \frac{2-3}{3}<\frac{\eta(\omega)-3}{3}<\frac{5-3}{3}\right\}=P\left\{\omega\colon\ -\frac{1}{3}<\zeta(\omega)<\frac{2}{3}\right\}=\Phi\left(\frac{2}{3}\right)-\Phi\left(-\frac{1}{3}\right)=$$

$$=\Phi\left(\frac{2}{3}\right)-\left[1-\Phi\left(\frac{1}{3}\right)\right]\approx 0.3779;$$

(b) $P\{\omega\colon\ \eta(\omega)>0\}=$

$$=P\left\{\omega\colon\ \frac{\eta(\omega)-3}{3}>\frac{0-3}{3}\right\}=P\{\omega\colon\ \zeta(\omega)>-1\}=1-\Phi(-1)=\Phi(1)\approx 0.8413;$$

(c) $P\{\omega\colon\ |\eta(\omega)-3|>6\}=P\{\omega\colon\ \eta(\omega)>9\}+P\{\omega\colon\ \eta(\omega)<-3\}=$

$$=P\left\{\omega\colon\ \frac{\eta(\omega)-3}{3}>\frac{9-3}{3}\right\}+P\left\{\omega\colon\ \frac{\eta(\omega)-3}{3}<\frac{-3-3}{3}\right\}=$$

$$=P\{\omega\colon\ \zeta(\omega)>2\}+P\{\omega\colon\ \zeta(\omega)<-2\}=1-\Phi(2)+\Phi(-2)=2[1-\Phi(2)]\approx 0.0456.$$

Example 63. A random variable $\eta(\omega)$ is normally distributed with mean zero and variance 1. Find a number ε such that

$$P\{|\eta(\omega)|>\varepsilon\}=\frac{1}{2}.$$

Solution:

$$P\{|\eta(\omega)|>\varepsilon\}=1-P\{|\eta(\omega)|\le\varepsilon\}=1-P\{-\varepsilon\le\eta(\omega)\le\varepsilon\}=$$

$$1-\Phi(\varepsilon)+\Phi(-\varepsilon)=1-\Phi(\varepsilon)+1-\Phi(\varepsilon)=2-2\,\Phi(\varepsilon).$$

Therefore, we obtain the equation $2-2\,\Phi(\varepsilon)=0.5$, that is $\Phi(\varepsilon)=\frac{3}{4}$. From the Table we have $\varepsilon=0.675$.

II. The Uniform Random Variable. We say that $\eta(\omega)$ is a uniform random variable on the interval (a,b) if its density function is given by (44) (or the distribution function has the form (32)).

Let us calculate $E\eta$ and Var (η). By definition (see (62)) we have

$$E\eta=\int_{-\infty}^{+\infty}x\,f_\eta(x)\,dx=\frac{1}{b-a}\int_a^b x\,dx=\frac{b^2-a^2}{2(b-a)}=\frac{a+b}{2}.$$

In words, the expected value of a random variable uniformly distributed over some interval is equal to the midpoint of that interval.

Let us calculate $\mathrm{Var}(\eta)$. Since $E\eta = \dfrac{a+b}{2}$, by (76), we have that

$$\mathrm{Var}(\eta) = \int_{-\infty}^{+\infty} \left(x - \frac{a+b}{2}\right)^2 f_\eta(x)\,dx = \frac{1}{b-a}\int_a^b \left(x - \frac{a+b}{2}\right)^2 dx = \frac{(b-a)^2}{12}.$$

Therefore, we have obtained that

$$E\eta = \frac{a+b}{2}, \qquad \mathrm{Var}(\eta) = \frac{(b-a)^2}{12}.$$

III. The Exponential Random Variables. We say that $\eta(\omega)$ is an exponential random variable if its density function is given by (44) (or the distribution function has the form (33)).

Let us calculate $E\eta$ and $\mathrm{Var}(\eta)$. By definition (see (62)) we have

$$E\eta = \int_{-\infty}^{+\infty} x\, f_\eta(x)\,dx = \int_0^{+\infty} x\,\lambda\, e^{-\lambda x}\,dx.$$

Integrating by parts $\left(\lambda e^{-\lambda x}\,dx = dv, u = x\right)$ yields

$$E\eta = -\left. x e^{-\lambda x}\right|_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x}\,dx = 0 - \left.\frac{1}{\lambda} e^{-\lambda x}\right|_0^{+\infty} = \frac{1}{\lambda}.$$

To obtain the variance of η, we first find $E\left[\eta^2\right]$.

$$E\left[\eta^2\right] = \int_0^{+\infty} x^2\,\lambda\, e^{-\lambda x}\,dx.$$

Integrating by parts $\left(\lambda e^{-\lambda x}\,dx = dv, u = x^2\right)$ gives

$$E\left[\eta^2\right] = -\left. x^2\, e^{-\lambda x}\right|_0^{+\infty} + 2\int_0^{+\infty} x\, e^{-\lambda x}\,dx = 0 + \frac{2}{\lambda}\,E\eta = \frac{2}{\lambda^2}.$$

Hence

$$\mathrm{Var}(\eta) = E\left[\eta^2\right] - \left[E\eta\right]^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.$$

Thus the mean of the exponential is the reciprocal of its parameter λ and the variance is the mean squared.

IV. The Poisson Random Variable. A discrete random variable $\eta(\omega)$, taking on one of the values $0, 1, 2, ...$ is said to be a Poisson random variable with parameter λ if for some $\lambda > 0$,

$$p(n) = P\{\omega\colon\ \eta(\omega) = n\} = \frac{\lambda^n}{n!}\, e^{-\lambda}, \qquad i = 0, 1, 2, ... \tag{102}$$

(compare with (27)).

Equation (102) defines a probability mass function, since

$$\sum_{n=0}^{\infty} p(n) = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = e^{-\lambda}\, e^{\lambda} = 1.$$

Let us calculate $E\eta$ and $\mathrm{Var}\,(\eta)$. By definition (see (62)) we have

$$E\eta = e^{-\lambda} \sum_{n=0}^{\infty} n\,\frac{\lambda^n}{n!} = e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} = \lambda\, e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} = \lambda\, e^{-\lambda}\, e^{\lambda} = \lambda.$$

To obtain the variance of η, we first find $E\left[\eta^2\right]$.

$$E\left[\eta^2\right] = e^{-\lambda} \sum_{n=0}^{\infty} n^2\,\frac{\lambda^n}{n!} = e^{-\lambda} \sum_{n=1}^{\infty} n\,\frac{\lambda^n}{(n-1)!} = e^{-\lambda} \sum_{n=1}^{\infty} (n-1+1)\,\frac{\lambda^n}{(n-1)!} =$$
$$= e^{-\lambda} \sum_{n=1}^{\infty} (n-1)\,\frac{\lambda^n}{(n-1)!} + e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} = \lambda^2 + \lambda.$$

Hence

$$\mathrm{Var}\,(\eta) = E\left[\eta^2\right] - \left[E\eta\right]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Therefore, the expected value and variance of a Poisson random variable are both equal to its parameter λ.

V. The Binomial Random Variables. Suppose now that n independent trials, each of which results in a "success" with probability p and in a "failure" with probability $1-p$, are to be performed. If $\eta(\omega)$ represents the number of successes that occur in the n trials, then $\eta(\omega)$ is said to be a *binomial* random variable with parameters n and p. $\eta(\omega)$ is a discrete random variable (see Example 53).

Denote by $\eta_i(\omega)$ a random variable which is the number of successes in ith trial. Therefore, η_i for any i is also discrete and take on only two values 0 and 1. $\eta_i = 1$ when the outcome is a success and $\eta_i = 0$ when it is a failure and thus the probability mass function of η_i is given by

$$p(0) = P\{\omega: \ \eta_i = 0\} = 1 - p, \qquad p(1) = P\{\omega: \ \eta_i = 1\} = p.$$

It is not difficult to verify that $\eta_1, ...\eta_n$ are independent identically distributed random variables and

$$\eta(\omega) = \sum_{i=1}^{n} \eta_i(\omega)$$

Let us compute its expected value and variance.

$$E\eta_i = 0 \cdot (1-p) + 1 \cdot p = p \qquad \text{for any} \quad i = 1, 2, ..., n.$$

Since $\eta_i(\omega) = \eta_i^2(\omega)$ we obtain

$$E\left[\eta_i^2\right] = E\eta_i = p.$$

Therefore

$$\text{Var}(\eta_i) = E\left[\eta_i^2\right] - \left[E\eta_i\right]^2 = p - p^2 = p \cdot (1-p).$$

By additivity of Expectation and additivity of $\text{Var}(\eta)$ for independent random variables (see (69) and (84)) we have

$$E\eta = n \cdot p \qquad \text{Var}(\eta) = n \cdot p \cdot (1-p).$$

VI. The Gamma Distributions.

The *gamma distribution* is important because it includes a wide class of specific distributions, some of which underlie fundamental statistical procedures. In addition to serving as a utility distribution, the gamma provides probabilities for yet another random variable associated with Poisson processes (the exponential distribution itself is a member of the gamma distributions).

The Gamma Function.

In order to describe the gamma distribution in detail, we must first consider a useful function, Gamma Function:

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} \, dx, \qquad \alpha > 0.$$

The symbol Γ (Greek uppercase *gamma*) is reserved for this function. The integration by parts of $\Gamma(\alpha)$ yields that

$$\Gamma(\alpha+1) = \alpha \, \Gamma(\alpha).$$

Note, that for any nonnegative integer k we have

$$\Gamma(k+1) = k!,$$

In particular, $\Gamma(1) = 1$.

An important class involves values with halves. We have

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

and for any positive integer k

$$\Gamma\left(k + \frac{1}{2}\right) = \frac{(2k-1)!!}{2^k} \sqrt{\pi},$$

where $(2k-1)!! = 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2k-1)$.

The Density Function of Gamma Random Variable

The following expression gives the density function for a gamma distribution.

$$f(x) = \begin{cases} 0 & \text{if} \quad x \le 0, \\ \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \text{if} \quad x > 0. \end{cases}$$

The two parameters λ and α may be any positive values ($\lambda > 0$ and $\alpha > 0$).
A special case of this function occurs when $\alpha = 1$. We have

$$f(x) = \begin{cases} 0 & \text{if} \quad x \le 0, \\ \lambda\, e^{-\lambda\, x} & \text{if} \quad x > 0 \end{cases}$$

which is the density function for the exponential distribution.
The expectation and Variance of gamma distribution have the forms:

$$E\eta = \frac{\alpha}{\lambda} \qquad \text{and} \qquad \mathrm{Var}\,(\eta) = \frac{\alpha}{\lambda^2}.$$

When α is natural, say $\alpha = n$, the gamma distribution with parameters (n, λ) often arises in practice:

$$f(x) = \begin{cases} 0 & \text{if} \quad x \le 0, \\ \dfrac{\lambda^n}{(n-1)!}\, x^{n-1}\, e^{-\lambda\, x} & \text{if} \quad x > 0. \end{cases}$$

This distribution is often referred to in the literature as the n-Erlang distribution.
The gamma distribution with $\lambda = 1/2$ and $\alpha = n/2$ (n is natural) is called χ_n^2 (read "chi-squared") distribution with n degrees of freedom:

$$f(x) = \begin{cases} 0 & \text{if} \quad x \le 0, \\ \dfrac{1}{2^{n/2}\,\Gamma(n/2)}\, x^{n/2-1}\, e^{-x/2} & \text{if} \quad x > 0. \end{cases}$$

We have

$$E\chi_n^2 = n \qquad \text{and} \qquad \mathrm{Var}\chi_n^2 = 2n.$$

VII. Beta Distributions.

A random variable is said to have a beta distribution if its density is given by

$$f(x) = \begin{cases} \dfrac{1}{B(a,b)}\, x^{a-1}\,(1-x)^{b-1} & \text{if} \quad 0 < x < 1, \\ 0 & \text{otherwise,} \end{cases}$$

where

$$B(a,b) = \int_0^1 x^{a-1}\,(1-x)^{b-1}\,dx.$$

Note that when $a = b = 1$ the beta density is the uniform density. When a and b are greater than 1 the density is bell-shaped, but when they are less than 1 it is U-shaped.

When $a = b$, the beta density is symmetric about $\frac{1}{2}$. When $b > a$, the density is skewed to the left (in the sense that smaller values become more likely), and it is skewed to the right when $a > b$. The following relationship exists between beta and gamma functions:

$$B(a,b) = \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}.$$

The expectation and Variance of beta distribution have the forms:

$$E\eta = \frac{a}{a+b} \qquad \text{and} \qquad \mathrm{Var}\,(\eta) = \frac{a\,b}{(a+b)^2\,(a+b+1)}.$$

VIII. The Weibull Distribution.

A random variable is said to have a Weibull distribution with parameters α and β, if its density is given by

$$f(x) = \begin{cases} \alpha\,\beta\,x^{\beta-1}\,\exp\left(-\alpha\,x^{\beta}\right) & \text{if} \quad x > 0, \\ 0 & \text{elsewhere,} \end{cases}$$

where $\alpha > 0$ and $\beta > 0$.

If we let $\beta = 1$, the Weibull distribution reduces to the exponential distribution.

The expectation and Variance of the Weibull distribution have the forms:

$$E\eta = \alpha^{-1/\beta}\,\Gamma\left(1 + \frac{1}{\beta}\right) \qquad \text{and} \qquad \mathrm{Var}\,(\eta) = \alpha^{-2/\beta}\left\{\Gamma\left(1+\frac{2}{\beta}\right) - \left[\Gamma\left(1+\frac{1}{\beta}\right)\right]^2\right\}.$$

LECTURE 18

Example 64. Expectation and Variance for the Number of Matches. A group of N people throw their hats into center of a room. The hats are mixed up, and each person randomly selects one. Find the expectation and the variance of the number of people that select their own hats.

Solution: Let $\eta(\omega)$ denote the number of matches, we can most easily compute $E[\eta(\omega)]$ by writing

$$\eta(\omega) = \eta_1(\omega) + \eta_2(\omega) + ... + \eta_N(\omega),$$

where

$$\eta_i(\omega) = \begin{cases} 1 & \text{if the } i\text{th person selects his own hat} \\ 0 & \text{otherwise} \end{cases}$$

Since, for each i, the ith person is equally likely to select any of the N hats,

$$E[\eta_i(\omega)] = P\{\eta_i(\omega) = 1\} = \frac{1}{N}$$

we see that

$$E[\eta(\omega)] = E[\eta_1(\omega)] + E[\eta_2(\omega)] + ... + E[\eta_N(\omega)] = N \cdot \frac{1}{N} = 1.$$

Hence, on the average, exactly one person selects his own hat.

Let us compute the variance of $\eta(\omega)$. Using (88) we have

$$\text{Var}\left(\sum_{i=1}^{N} \eta_i\right) = \sum_{i=1}^{N} \text{Var }(\eta_i) + 2\sum_{i<j} \text{Cov}(\eta_i, \eta_j). \tag{92}$$

Since $P\{\eta_i(\omega) = 1\} = \frac{1}{N}$, we see from (75) that

$$\text{Var}(\eta_i) = \frac{1}{N} - \frac{1}{N^2} = \frac{N-1}{N^2}.$$

Since

$$\eta_i(\omega)\,\eta_j(\omega) = \begin{cases} 1 & \text{if the } i\text{th and } j\text{th persons both select their own hats} \\ 0 & \text{otherwise} \end{cases}$$

we obtain

$$E[\eta_i(\omega)\,\eta_j(\omega)] = P\{\eta_i(\omega) = 1 \cap \eta_j(\omega) = 1\} = \frac{1}{N}\,\frac{1}{N-1}.$$

Hence by (86) we have

$$\text{Cov}(\eta_i(\omega), \eta_j(\omega)) = \frac{1}{N\,(N-1)} - \frac{1}{N^2} = \frac{1}{N^2\,(N-1)}.$$

Therefore from (92),

$$\text{Var}(\eta(\omega)) = \frac{N-1}{N} + 2\binom{N}{2}\frac{1}{N^2\,(N-1)} = 1.$$

Thus both the mean and variance of the number of matches are equal to 1.

Example 65. Let η be any random variable with $E\eta = a$ and $\text{Var}(\eta) = \sigma^2$. Then, if $\varepsilon = k\sigma$, $(k > 0)$, Chebyshev's inequality states that

$$P(|\eta - a| \geq k \cdot \sigma) \leq \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}.$$

Thus, for any random variable, the probability of a deviation from the mean of more than k standard deviations is less than or equal to $1/k^2$. If, for example, $k = 5$, $1/k^2 = 0.04$.

Example 66. Chebyshev's inequality is the best possible inequality in the sense that, for any $\varepsilon > 0$, it is possible to give an example of a random variable for which Chebyshev's inequality is in fact an equality. To see this, given $\varepsilon > 0$, choose a discrete random variable η which take on two equally likely values $-\varepsilon$ and ε, that is

$-\varepsilon$	$+\varepsilon$
$1/2$	$1/2$

Then $E\eta = 0$, $\text{Var}(\eta) = \varepsilon^2$, and

$$P(|\eta - E\eta| \geq \varepsilon) \leq \frac{\text{Var}(\eta)}{\varepsilon^2} = 1.$$

The importance of Markov's and Chebyshev's inequalities is that they enable us to derive bounds on probabilities when only the mean, or both the mean and the variance, of the probability distribution are known.

Example 67. Suppose that it is known that the number of items produced in a factory during a week is a random variable with mean 50.

a) What can be said about the probability that this week's production will exceed 75?

b) If the variance of a week's production is known to equal 25, then what can be said about the probability that this week's production will be between 40 and 60?

SOLUTION: Let $\eta(\omega)$ be the number of items that will be produced in a week:

a) By Markov's inequality

$$P(\eta > 75) \leq \frac{E\eta}{75} = \frac{50}{75} = \frac{2}{3}.$$

b) By Chebyshev's inequality

$$P(|\eta - 50| \geq 10) \leq \frac{\mathrm{Var}(\eta)}{100} = \frac{1}{4}.$$

Hence

$$P(|\eta - 50| < 10) \geq 1 - \frac{1}{4} = \frac{3}{4}$$

and so the probability that this week's production will be between 40 and 60 is at least 0.75.

§28. LIMIT THEOREMS

The most important theoretical results in probability theory are limit theorems. Of these, the most important are those that are classified either under the heading of "laws of large numbers" or under the heading of "central limit theorems". Usually, theorems are considered with stating conditions under which the average of a sequence of random variables converges (in some sense) to the expected average. On the other hand, central limit theorems are concerned with determining conditions under which the sum of a large number of random variables has a probability distribution that is approximately normal.

I. The Weak Law of Large Numbers

Theorem 11 (Markov's theorem). *Let $\eta_1(\omega)$, ..., $\eta_n(\omega)$ be a sequence of random variables, each having finite mean ($E\eta_k = a_k$) and variance. If*

$$\frac{1}{n^2} \mathit{Var}\left(\sum_{k=1}^{n} \eta_k(\omega)\right) \to 0 \qquad \text{as} \quad n \to \infty, \tag{103}$$

then for any $\varepsilon > 0$,

$$P\left\{\omega\colon \quad \left|\frac{1}{n}\sum_{k=1}^{n} (\eta_k - a_k)\right| \geq \varepsilon\right\} \to 0 \qquad \text{as} \quad n \to \infty. \tag{104}$$

Condition (103) we will call Markov's condition.

Proof: Denote by

$$\zeta_n(\omega) = \frac{1}{n}\sum_{k=1}^{n} \eta_k(\omega).$$

Therefore, by additivity of expectation we get

$$E\,(\zeta_n) = \frac{1}{n}\sum_{k=1}^{n} a_k. \tag{105}$$

By Chebyshev's inequality (see (81)) we have

$$P\left\{\omega:\ \left|\frac{1}{n}\sum_{k=1}^{n}(\eta_k - a_k)\right| \geq \varepsilon\right\} = P\left\{\omega:\ |\zeta_n(\omega) - E\,(\zeta_n)\,| \geq \varepsilon\right\} \leq \frac{\mathrm{Var}\,(\zeta_n)}{\varepsilon^2} = \frac{\mathrm{Var}\left(\sum\limits_{k=1}^{n}\eta_k(\omega)\right)}{n^2\,\varepsilon^2}$$

Above we used (82). As (103) we have (104). The proof is complete.

LECTURE 19

Corollary 11. Let $\eta_1(\omega)$, ..., $\eta_n(\omega)$ be a sequence of independent random variables, each having finite mean ($E\eta_k = a_k$) and variance ($\mathrm{Var}\,(\eta_k) = \sigma_k^2$). If there exists a constant C such that for any k $\sigma_k^2 \le C$ (i. e. a sequence of variances is bounded), then for any $\varepsilon > 0$,

$$P\left\{\omega: \left|\frac{1}{n}\sum_{k=1}^{n}(\eta_k - a_k)\right| \ge \varepsilon\right\} \to 0 \qquad \text{as } n \to \infty.$$

Proof: We have to prove that all conditions of Markov's theorem are satisfied. Hence we must to verify condition (103). Indeed,

$$\frac{1}{n^2}\mathrm{Var}\left(\sum_{k=1}^{n}\eta_k(\omega)\right) = \text{(by the additivity of variance for independent random variables)} =$$

$$= \frac{1}{n^2}\sum_{k=1}^{n}\mathrm{Var}\,(\eta_k) \le \frac{1}{n^2}\sum_{k=1}^{n}C \le \frac{C\cdot n}{n^2} = \frac{C}{n} \to 0 \qquad \text{as} \quad n \to \infty.$$

The proof is complete.

Corollary 12. Let $\eta_1(\omega)$, ..., $\eta_n(\omega)$ be a sequence of independent and identically distributed random variables, each having finite mean ($E\eta_k = a$) and variance. Then for any $\varepsilon > 0$,

$$P\left\{\omega: \left|\frac{\sum\limits_{k=1}^{n}\eta_k}{n} - a\right| \ge \varepsilon\right\} \to 0 \qquad \text{as} \quad n \to \infty.$$

Proof: The proof is obvious and follows from Corollary 11.

The general form of the weak law of large numbers is presented in the following theorem which we state without proof.

Theorem 12 (Khintchine). *Let $\eta_1(\omega)$, ..., $\eta_n(\omega)$ be a sequence of independent and identically distributed random variables, each having finite mean ($E\eta_k = a$). Then for any $\varepsilon > 0$,*

$$P\left\{\omega: \left|\frac{1}{n}\sum_{k=1}^{n}\eta_k(\omega) - a\right| \ge \varepsilon\right\} \to 0 \qquad \textit{as} \quad n \to \infty.$$

Thus, we have proved Corollary 12 under the additional assumption that the random variables have a finite variance.

Corollary 13 (Bernoulli). Suppose now that n independent trials, each of which results in a "success" with probability p and in a "failure" with probability $1-p$, are to be performed. Let $\eta(\omega)$ be the number of successes that occur in the n trials. Then for any $\varepsilon > 0$,

$$P\left\{\omega: \ \left|\frac{\eta(\omega)}{n} - p\right| \geq \varepsilon\right\} \to 0 \qquad \text{as} \quad n \to \infty.$$

Proof: Denote by $\eta_i(\omega)$ a random variable which is the number of successes in ith trial. Therefore, η_i for any i is discrete and take on only two values 0 and 1. $\eta_i = 1$ when the outcome is a success and $\eta_i = 0$ when it is a failure. Moreover, η_1, ..., η_n are independent identically distributed random variables and

$$\eta(\omega) = \sum_{i=1}^{n} \eta_i(\omega), \quad E\eta_i = p, \quad \mathrm{Var}(\eta_i) = p \cdot (1-p) \qquad \text{for any} \quad i = 1, 2, ..., n.$$

Now we can use Corollary 12 for random variables η_i. The proof is complete.

II. The Central Limit Theorem. The central limit theorem is one of the most remarkable result in probability theory. Loosely put, it states that the sum of a large number of independent random variables has a distribution that is approximately normal. Hence it not only provides a simple method for computing approximate probabilities for sums of independent random variables, but it also helps explain the remarkable fact that empirical frequencies of so many natural populations exhibit normal curves.

In its simplest form (for identically distributed random variables) the central limit theorem is as follows.

Theorem 13. *Let $\eta_1(\omega)$, ..., $\eta_n(\omega)$ be a sequence of independent identically distributed random variables, each having finite mean ($E\eta_k = a$) and variance σ^2. Then the distribution of*

$$\frac{\sum\limits_{k=1}^{n} \eta_k(\omega) - na}{\sigma \cdot \sqrt{n}}$$

tends to the standard normal distribution as $n \to \infty$. *That is,*

$$P\left\{\omega: \quad \frac{\sum\limits_{k=1}^{n} \eta_k(\omega) - na}{\sigma \cdot \sqrt{n}} < x\right\} \longrightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{y^2}{2}\right) dy \qquad \textit{as} \quad n \to \infty.$$

Remark 14. Although Theorem 13 only states that, for each x,

$$P\left\{\omega: \quad \frac{\sum\limits_{k=1}^{n} \eta_k(\omega) - na}{\sigma \cdot \sqrt{n}} < x\right\} \longrightarrow \Phi(x) \qquad \text{as} \quad n \to \infty \tag{106}$$

it can, in fact, be shown that the convergence is uniform in x (We say that $\varphi_n(x) \to \varphi(x)$ uniformly in x, if for each $\varepsilon > 0$, there exists an N such that $|\varphi_n(x) - \varphi(x)| < \varepsilon$ for all x whenever $n \geq N$).

The following theorem is known as the DeMoivre–Laplace limit theorem. It was originally proved for the special case $p = 1/2$ by DeMoivre in 1733 and then extended to general p by Laplace in 1812. It states that if we "standardize" a binomial random variable with parameters n and p by first subtracting its mean $n \cdot p$ and then dividing the result by its standard deviation $\sqrt{n \cdot p \cdot (1-p)}$, then this standardized random variable (which has mean 0 and variance 1) will, when n is large, approximately have a standard normal distribution.

Theorem 14. *If* $\eta(\omega)$ *denotes the number of successes that occur when* n *independent trials, each resulting in a success with probability* p, *are performed then*

$$P\left\{\omega: \quad \frac{\eta(\omega) - np}{\sqrt{n \cdot p \cdot (1-p)}} < x\right\} \longrightarrow \Phi(x) \qquad \textit{as} \quad n \to \infty \tag{107}$$

It should be noted that we now have two possible approximations to binomial probabilities: the Poisson approximation, which yields a good approximation when n is large and $n \cdot p$ moderate, and the normal approximation, which can be shown to be quite good when $n \cdot p \cdot (1-p)$ is large. The normal approximation will, in general, be quite good for values of n satisfying $n \cdot p \cdot (1-p) \geq 10$.

Example 68. Let $\eta(\omega)$ be the number of times that a fair coin, flipped 40 times, lands heads. Find the probability that $\eta(\omega) = 20$. Use the normal approximation and then compare it to the exact solution.

SOLUTION: Since the binomial is a discrete random variable and the normal continuous random variable, the best approximation is obtained by writing the desired probability as

$$P\{\omega\colon\ \eta(\omega) = 20\} = P\{\omega\colon\ 19.5 < \eta(\omega) < 20.5\} =$$
$$= P\left\{\omega\colon\ \frac{19.5-20}{\sqrt{10}} < \frac{\eta(\omega)-20}{\sqrt{10}} < \frac{20.5-20}{\sqrt{10}}\right\} \approx P\left\{\omega\colon\ -0.16 < \frac{\eta(\omega)-20}{\sqrt{10}} < 0.16\right\} \approx$$
$$\approx \Phi(0.16) - \Phi(-0.16) = 2\,\Phi(0.16) - 1 \approx 0.1272.$$

The exact result

$$P\{\omega\colon\ \eta(\omega) = 20\} = \binom{40}{20}\left(\frac{1}{2}\right)^{40}$$

which can be shown to approximately equal 0.1254.

Example 69. Let $\eta_i(\omega)$, $i = 1, ..., 10$ be independent random variables, each uniformly distributed over $(0,1)$. Calculate an approximation to

$$P\left\{\omega\colon\ \sum_{i=1}^{10}\eta_i(\omega) > 6\right\}.$$

SOLUTION: Since

$$E\eta_i = \frac{1}{2}, \qquad \mathrm{Var}\,(\eta_i) = \frac{1}{12}$$

we have by the central limit theorem

$$P\left\{\omega\colon\ \sum_{i=1}^{10}\eta_i(\omega) > 6\right\} = P\left\{\omega\colon\ \frac{\sum\limits_{i=1}^{10}\eta_i(\omega) - 5}{\sqrt{10\cdot\dfrac{1}{12}}} > \frac{6-5}{\sqrt{10\cdot\dfrac{1}{12}}}\right\} \approx 1 - \Phi(\sqrt{1.2}) \approx 0.16.$$

Hence only 16 percent of the time will $\sum\limits_{i=1}^{10}\eta_i(\omega)$ be greater than 6.

Example 70. Let probability of event A in each trial be 0.05. How many independent trials we have to perform, that probability of event A appears at least 5 times is equal to 0.8?

SOLUTION: We have

$$P\{\omega\colon\ \eta_n(\omega)\ge 5\}=P\left\{\omega\colon\ \frac{\eta_n(\omega)-n\,0.05}{\sqrt{n\,0.05\,0.95}}\ge\frac{5-n\,0.05}{\sqrt{n\,0.05\,0.95}}\right\}=$$

$$=1-P\left\{\frac{\eta_n(\omega)-n\,0.05}{\sqrt{n\,0.05\,0.95}}\le\frac{5-n\,0.05}{\sqrt{n\,0.05\,0.95}}\right\}\approx 1-\Phi\left(\frac{5-n\,0.05}{\sqrt{n\,0.05\,0.95}}\right)=\Phi\left(\frac{n\,0.05-5}{\sqrt{n\,0.05\,0.95}}\right).$$

Therefore, we obtain the equation

$$\Phi\left(\frac{n\,0.05-5}{\sqrt{n\,0.05\,0.95}}\right)=0.8$$

or

$$\frac{0.05\,n-5}{\sqrt{n\,0.05\,0.95}}=0.8416.$$

It follows that $n=144$.

LECTURE 20

Example 71. Consider n rolls of a die. Let η_j be the outcome of the jth roll. Then $S_n = \eta_1 + ... + \eta_n$ is the sum of the first n rolls. This is an independent trials process with

$$E\eta_j = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}.$$

Thus, by the Law of large numbers, for any $\varepsilon > 0$

$$P\left(\left|\frac{S_n}{n} - \frac{7}{2}\right| \geq \varepsilon\right) \to 0 \qquad \text{as} \qquad n \to \infty.$$

An equivalently way to state this is that, for any $\varepsilon > 0$

$$P\left(\left|\frac{S_n}{n} - \frac{7}{2}\right| < \varepsilon\right) \to 1 \qquad \text{as} \qquad n \to \infty.$$

Example 72. Let η_1, ..., η_n be a Bernoulli trials process with probability 0.3 for success and 0.7 for failure. Let $\eta_j = 1$ if the jth outcome is a success and 0 otherwise. Then, $E\eta_j = 0.3$ and $\mathrm{Var}(\eta_j) = 0.3 \cdot 0.7 = 0.21$. If

$$A_n = \frac{S_n}{n} = \frac{\sum_{i=1}^{n} \eta_i}{n}$$

is the average of the η_j, then $E(A_n) = 0.3$ and $\mathrm{Var}(A_n) = \frac{0.21}{n}$. Chebyshev's inequality states that if, for example, $\varepsilon = 0.1$

$$P(|A_n - 0.3| \geq 0.1) \leq \frac{0.21}{n \cdot (0.1)^2} = \frac{21}{n}.$$

Thus, if $n = 100$

$$P(|A_{100} - 0.3| \geq 0.1) \leq 0.21,$$

or if $n = 1,000$

$$P(|A_{1,000} - 0.3| \geq 0.1) \leq 0.021,$$

These can be written as

$$P(0.2 \leq A_{100} \leq 0.4) \geq 0.79,$$

$$P(0.2 \le A_{1,000} \le 0.4) \ge 0.979.$$

Example 73. Measurement Error. Suppose that η_1, ..., η_n are repeated, independent unbiased measurements of a quantity a, and that $\mathrm{Var}\,(\eta_k) = \sigma^2$, $(k = 1, 2, ..., n)$. The average of the measurements,

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} \eta_k$$

is used as an estimate of a. The law of large numbers tells as that $\overline{X} = \dfrac{S_n}{n} = \dfrac{\sum_{k=1}^{n} \eta_k}{n}$ converges to a, so we can hope that $\overline{X}$ is close to a if n is large. Chebyshev's inequality allows us to bound the probability of an error of a given size, but the central limit theorem gives a much sharper approximation to the actual error. Suppose that we wish to find

$$P(|\overline{X} - a| \le \varepsilon)$$

for some constant ε. To use the central limit theorem to approximate this probability, we first standardize, using $E(\overline{X}) = a$ and $\mathrm{Var}\,(\overline{X}) = \dfrac{\sigma^2}{n}$:

$$P(|\overline{X} - a| \le \varepsilon) = P(-\varepsilon \le \overline{X} - a \le \varepsilon) = P(\frac{-\varepsilon}{\sigma/\sqrt{n}} \le \frac{\overline{X} - a}{\sigma/\sqrt{n}} \le \frac{\varepsilon}{\sigma/\sqrt{n}}) \approx \Phi\left(\frac{\varepsilon}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{-\varepsilon}{\sigma/\sqrt{n}}\right).$$

For example, suppose that 16 measurements are taken with $\sigma = 1$. The probability that the average deviates from a by less than or equal to 0.5 is approximately

$$P(|\overline{X} - a| \le 0.5) \approx \Phi(2) - \Phi(-2) \approx 0.954.$$

§29. BIVARIATE NORMAL DISTRIBUTION

Up to this point all of the random variables have been of one dimensions. A very important two–dimensional probability law which is a generalization of the one–dimensional normal probability law is called *Bivariate normal distribution*.

The random variables $\eta_1(\omega)$ and $\eta_2(\omega)$ are said to have a Bivariate normal distribution with parameters $(a_1, a_2, \sigma_1, \sigma_2, \rho)$ if their joint density function is given by

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times$$
$$\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1 - a_1}{\sigma_1}\right)^2 - 2\rho \cdot \frac{(x_1 - a_1)(x_2 - a_2)}{\sigma_1\sigma_2} + \left(\frac{x_2 - a_2}{\sigma_2}\right)^2\right]\right\}. \tag{108}$$

We see that Bivariate normal distribution is determined by 5 parameters. These are a_1, a_2, σ_1, σ_2 and ρ such that $a_1 \in (-\infty, +\infty)$, $a_2 \in (-\infty, +\infty)$, $\sigma_1 > 0$, $\sigma_2 > 0$ and $\rho \in (-1, +1)$. By (51) the density function of η_1 is defined by the following formula

$$f_{\eta_1}(x_1) = \int_{-\infty}^{+\infty} f(x_1, x_2)\, dx_2 = \frac{1}{\sigma_1 \sqrt{2\pi}} \exp\left(-\frac{(x_1 - a_1)^2}{2\sigma_1^2} \right).$$

Similarly, for random variable η_2, we obtain (see (52)):

$$f_{\eta_2}(x_2) = \int_{-\infty}^{+\infty} f(x_1, x_2)\, dx_1 = \frac{1}{\sigma_2 \sqrt{2\pi}} \exp\left(-\frac{(x_2 - a_2)^2}{2\sigma_2^2} \right).$$

Therefore, $\eta_1(\omega)$ and $\eta_2(\omega)$ are both normal random variables with respective parameters $\mathcal{N}(a_1, \sigma_1)$ and $\mathcal{N}(a_2, \sigma_2)$.

Thus, the marginal distributions of $\eta_1(\omega)$ and $\eta_2(\omega)$ are both normal, even though the joint distributions for $\rho = 0$, $\rho = 0.3$, $\rho = 0.6$ and $\rho = 0.9$ are quite different from each other. It is not difficult to verify that

$$\operatorname{Cov}(\eta_1, \eta_2) = E\left[(\eta_1 - a_1) \cdot (\eta_2 - a_2)\right] = \rho\, \sigma_1 \sigma_2,$$

i. e. covariance between $\eta_1(\omega)$ and $\eta_2(\omega)$ is equal to $\rho\, \sigma_1\, \sigma_2$ and therefore

$$\rho(\eta_1, \eta_2) = \rho.$$

Thus, we have proved that

η_1 and η_2 are independent if and only if the correlation is equal to 0.

Indeed, substituting $\rho = 0$ in (108) we obtain

$$f(x_1, x_2) = f_{\eta_1}(x_1) \cdot f_{\eta_2}(x_2).$$

Definition 19. If $\eta_1(\omega)$ and $\eta_2(\omega)$ have a joint density function $f(x_1, x_2)$, then the conditional density function of $\eta_1(\omega)$, given that $\eta_2(\omega) = x_2$ is defined for all values of x_2 such that $f_{\eta_2}(x_2) \neq 0$, by

$$f_{\eta_1/\eta_2}(x_1 | x_2) = \frac{f(x_1, x_2)}{f_{\eta_2}(x_2)}.$$

To motivate this definition, multiply the left–hand side by dx_1 and the right–hand side by $(dx_1 \cdot dx_2)/(dx_2)$ to obtain

$$f_{\eta_1/\eta_2}(x_1|x_2)dx_1 = \frac{f(x_1, x_2)\, dx_1\, dx_2}{f_{\eta_2}(x_2)\, dx_2} \approx \frac{P\{\omega\colon\ x_1 \le \eta_1(\omega) \le x_1 + dx_1, x_2 \le \eta_2(\omega) \le x_2 + dx_2\}}{P\{\omega\colon\ x_2 \le \eta_2(\omega) \le x_2 + dx_2\}} =$$
$$= P\{\omega\colon\ x_1 \le \eta_1(\omega) \le x_1 + dx_1 | x_2 \le \eta_2(\omega) \le x_2 + dx_2\}.$$

In other words, for small values of dx_1 and dx_2, $f_{\eta_1/\eta_2}(x_1|x_2)$ represents the conditional probability that $\eta_1(\omega)$ is between x_1 and $x_1 + dx_1$, given that $\eta_2(\omega)$ is between x_2 and $x_2 + dx_2$.

One can show that the conditional density of η_1, given that $\eta_2 = x_2$, is the normal density with parameters

$$\mathcal{N}\left(a_1 + \rho\frac{\sigma_1}{\sigma_2}\,(x_2 - a_2), \sigma_1^2\,(1 - \rho^2)\right).$$

Similarly, the conditional density of η_2, given that $\eta_1 = x_1$, is the normal density with parameters

$$\mathcal{N}\left(a_2 + \rho\frac{\sigma_2}{\sigma_1}\,(x_1 - a_1), \sigma_2^2\,(1 - \rho^2)\right).$$

LECTURE 21

§30. SUMS OF INDEPENDENT RANDOM VARIABLES

It is often important to be able to calculate the distribution of $\eta_1+\eta_2$ from the distributions of η_1 and η_2 when η_1 and η_2 are independent. Suppose that η_1 and η_2 are independent, continuous random variables having probability density functions $f_1(x)$ and $f_2(x)$. The distribution function $F_{\eta_1+\eta_2}(x)$ of $\eta_1+\eta_2$ is obtained as follows:

$$F_{\eta_1+\eta_2}(x)=P\{\eta_1+\eta_2\le x\}=\iint_{y+z\le x} f_1(y)\,f_2(z)\,dy\,dz=\int_{-\infty}^{+\infty} f_1(y)\,dy\int_{-\infty}^{x-y} f_2(z)\,dz=$$

$$=\int_{-\infty}^{+\infty} F_2(x-y)\,f_1(y)\,dy,$$

where $F_2(x)$ is the distribution function of η_2.

The distribution function $F_{\eta_1+\eta_2}(x)$ is called the convolution of the distributions $F_1(x)$ and $F_2(x)$ (the distribution functions of η_1 and η_2, respectively).

By symmetry we have the following two formulae:

$$F_{\eta_1+\eta_2}(x)=\int_{-\infty}^{+\infty} F_2(x-y)\,f_1(y)\,dy=\int_{-\infty}^{+\infty} F_1(x-y)\,f_2(y)\,dy.$$

By differentiating we obtain that density function $f_{\eta_1+\eta_2}(x)$ of $\eta_1+\eta_2$ is given by

$$f_{\eta_1+\eta_2}(x)=\frac{d}{dx}\int_{-\infty}^{+\infty} F_2(x-y)\,f_1(y)\,dy=\int_{-\infty}^{+\infty}\frac{d}{dx}F_2(x-y)\,f_1(y)\,dy=\int_{-\infty}^{+\infty} f_2(x-y)\,f_1(y)\,dy.$$

By symmetry, we have two formulae for density function of $\eta_1+\eta_2$:

$$f_{\eta_1+\eta_2}(x)=\int_{-\infty}^{+\infty} f_1(x-y)\,f_2(y)\,dy=\int_{-\infty}^{+\infty} f_2(x-y)\,f_1(y)\,dy. \tag{109}$$

Example 74. (Sum of two independent uniform random variables). If η_1 and η_2 are independent random variables, both uniformly distributed over $(0,1)$, calculate density function of $\eta_1+\eta_2$.

SOLUTION: Since density functions of both η_1 and η_2 have the form

$$f(x)=\begin{cases}0 & \text{if} \quad x\notin(0,1)\\ 1 & \text{if} \quad 0\le x\le 1\end{cases},$$

from (109) we obtain

$$f_{\eta_1+\eta_2}(x) = \int_0^1 f_1(x-y)\,dy.$$

For $0 \le x \le 1$, this yields

$$f_{\eta_1+\eta_2}(x) = \int_0^x dy = x,$$

and for $1 < x < 2$, we get

$$f_{\eta_1+\eta_2}(x) = \int_{x-1}^1 dy = 2 - x.$$

Hence

$$f_{\eta_1+\eta_2}(x) = \begin{cases} x & \text{if} \quad 0 \le x \le 1 \\ 2-x & \text{if} \quad 1 < x < 2 \\ 0 & \text{otherwise} \end{cases}.$$

Because of the shape of its density function, the random variable $\eta_1 + \eta_2$ is said to have a triangular distribution.

An important property of gamma distribution is that for a fixed value of λ, it is closed under convolutions.

Theorem 15. If η_1 and η_2 are independent gamma random variables with respective parameters (α_1, λ) and (α_2, λ), then $\eta_1+\eta_2$ is a gamma random variable with parameters $(\alpha_1 + \alpha_2, \lambda)$.

Proof: Recall that gamma random variable has a density function of the form:

$$f(x) = \begin{cases} 0 & \text{if} \quad x \le 0, \\ \dfrac{\lambda^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-\lambda\, x} & \text{if} \quad x > 0. \end{cases}$$

The two parameters λ and α may be any positive values ($\lambda > 0$ and $\alpha > 0$).

Using (109), we obtain

$$f_{\eta_1+\eta_2}(x) = \begin{cases} 0 & \text{if} \quad x \le 0, \\ \dfrac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)}\, x^{\alpha_1+\alpha_2-1}\, e^{-\lambda\, x} & \text{if} \quad x > 0. \end{cases}$$

Hence the result is proved.

It is now a simpler matter to establish, by using Theorem 15 and the method of mathematical induction, that if η_i, $i = 1, 2, ..., n$ are independent gamma random variables with respective parameters (α_i, λ), $i = 1, 2, ..., n$, then

$$\sum_{i=1}^{n} \eta_i$$

is gamma random variable with parameters $\left(\sum_{i=1}^{n} \alpha_i, \lambda\right)$.

Example 75. Let η_1, η_2,..., η_n be n independent exponential random variables each having parameter λ. Then, as an exponential random variable with parameter λ is the same as a gamma random variable with parameters $(1, \lambda)$, we see from Theorem 15 that $\eta_1 + \eta_2 + ... + \eta_n$ is a gamma random variable with parameters (n, λ). Therefore, has the form:

$$f(x) = \begin{cases} 0 & \text{if} \quad x \leq 0, \\ \dfrac{\lambda^n}{(n-1)!}\, x^{n-1}\, e^{-\lambda\, x} & \text{if} \quad x > 0, \end{cases}$$

that is n-Erlang distribution.

If η_1, η_2, ..., η_n are independent standard normal random variables, then

$$\xi = \sum_{i=1}^{n} \eta_i^2$$

is said to have the ***chi-squared*** (sometimes seen as χ^2) distribution with n degrees of freedom. Let us compute its density function. When $n = 1$, $\xi = \eta_1^2$, and its density function is given by

$$f_{\eta_1^2}(x) = \begin{cases} 0 & \text{if} \quad x \leq 0, \\ \dfrac{1}{2\sqrt{x}}\,[f_{\eta_1}(\sqrt{x}) + f_{\eta_1}(-\sqrt{x})] & \text{if} \quad x > 0, \end{cases} =$$

$$= \begin{cases} 0 & \text{if} \quad x \leq 0, \\ \dfrac{1}{2\sqrt{x}}\,\dfrac{2}{\sqrt{2\,\pi}}\,exp(-x/2) & \text{if} \quad x > 0, \end{cases} =$$

$$= \begin{cases} 0 & \text{if} \quad x \leq 0, \\ \dfrac{1}{2\sqrt{\pi}}\,\left(\dfrac{x}{2}\right)^{-1/2}\, exp(-x/2) & \text{if} \quad x > 0. \end{cases}$$

But we recognize the above as the gamma distribution with parameters $(1/2, 1/2)$ (A by-product of this analysis is that $\Gamma(1/2) = \sqrt{\pi}$). But as each η_i^2 is gamma distribution with parameters $(1/2, 1/2)$, we obtain from Theorem 15 that the χ^2 distribution with n degrees of freedom is just the gamma distribution with parameters $(n/2, 1/2)$ and has a probability density function given by

$$f_{\chi^2}(x) = \begin{cases} 0 & \text{if} \quad x \leq 0, \\ \dfrac{x^{n/2-1}\, exp\Big(-\dfrac{x}{2}\Big)}{2^{n/2}\,\Gamma\left(\dfrac{n}{2}\right)} & \text{if} \quad x > 0. \end{cases}$$

When n is even, $\Gamma(n/2) = (n/2-1)!$, whereas when n is odd, $\Gamma(n/2)$ can be obtained from iterating the relationship $\Gamma(\lambda) = (\lambda-1)\,\Gamma(\lambda-1)$ and then using the previously obtained result that $\Gamma(1/2) = \sqrt{\pi}$. For example,

$$\Gamma\left(\frac{5}{2}\right) = \frac{3}{2}\,\Gamma\left(\frac{3}{2}\right) = \frac{3}{4}\sqrt{\pi}.$$

The chi-squared distribution often arises in practice as being the distribution of the square of the error involved when one attempts to hit a target in n-dimensional space when the coordinate errors are taken to be independent until normal random variables. It is also important in statistical analysis.

Theorem 16. If η_i, $i = 1, \ldots, n$ are independent random variables that are normally distributed with respective parameters μ_i, σ_i^2, $i = 1, \ldots, n$, then $\sum\limits_{i=1}^{n} \eta_i$ is normally distributed with parameters $\sum\limits_{i=1}^{n} \mu_i$ and $\sum\limits_{i=1}^{n} \sigma_i^2$.

Proof of Theorem 16: Let η_1 and η_2 be independent normal random variables, with η_1 having mean 0 and variance σ^2, and η_2 having mean 0 and variance 1. We have to determine the density function of $\eta_1 + \eta_2$ by utilizing formula (109). We have

$$f_{\eta_1}(x-y)\, f_{\eta_2}(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-y)^2}{2\sigma^2}\right\} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{y^2}{2}\right\} =$$

$$= \frac{1}{2\pi\,\sigma} \exp\left\{-\frac{x^2}{2\sigma^2}\right\} \exp\left\{-c\left(y^2 - 2y\,\frac{x}{1+\sigma^2}\right)\right\},$$

where

$$c = \frac{1}{\sigma^2} + \frac{1}{2} = \frac{1+\sigma^2}{2\sigma^2}$$

Hence, from formula (109),

$$f_{\eta_1+\eta_2}(x) = \frac{1}{2\pi\,\sigma} \exp\left\{-\frac{x^2}{2\sigma^2}\right\} \exp\left\{\frac{x^2}{2\sigma^2(1+\sigma^2)}\right\} \cdot \int_{-\infty}^{\infty} \exp\left\{-c\left(y - \frac{x}{1+\sigma^2}\right)^2\right\} dy =$$

$$= \frac{1}{2\pi\,\sigma} \exp\left\{-\frac{x^2}{2(1+\sigma^2)}\right\} \int_{-\infty}^{\infty} \exp\{-c\,x^2\}\,dx = \frac{1}{\sqrt{2\,\pi\,(1+\sigma^2)}} \exp\left\{-\frac{a^2}{2(1+\sigma^2)}\right\},$$

because

$$\int_{-\infty}^{\infty} \exp\{-c\,x^2\}\,dx = \sqrt{\frac{\pi}{c}}.$$

But this implies that $\eta_1 + \eta_2$ is normal with mean 0 and variance $1 + \sigma^2$.

Now, suppose that η_1 and η_2 are independent normal random variables, with η_i having mean μ_i and variance σ_i^2, $i = 1, 2$. Then

$$\eta_1 + \eta_2 = \sigma_2\left(\frac{\eta_1 - \mu_1}{\sigma_2} + \frac{\eta_2 - \mu_2}{\sigma_2}\right) + \mu_1 + \mu_2.$$

But since $(\eta_1 - \mu_1)/\sigma_2$ is normal with mean 0 and variance σ_1^2/σ_2^2, and $(\eta_2 - \mu_2)/\sigma_2$ is normal with mean 0 and variance 1, it follows from our previous result, that

$$\frac{\eta_1 - \mu_1}{\sigma_2} + \frac{\eta_2 - \mu_2}{\sigma_2}$$

is normal with mean 0 and variance $1 + \sigma_1^2/\sigma_2^2$, implying that $\eta_1 + \eta_2$ is normal with mean $\mu_1 + \mu_2$ and variance $\sigma_2^2(1 + \sigma_1^2/\sigma_2^2) = \sigma_1^2 + \sigma_2^2$.

Thus, Theorem 16 is established when $n = 2$. The general case now follows by induction. That is, assume that it is true when there are $n - 1$ random variables. Now consider the case of n, and write

$$\sum_{i=1}^{n} \eta_i = \sum_{i=1}^{n-1} \eta_i + \eta_n.$$

By the induction hypothesis, $\sum\limits_{i=1}^{n-1} \eta_i$ is normal with mean $\sum\limits_{i=1}^{n-1} \mu_i$ and variance $\sum\limits_{i=1}^{n-1} \sigma_i^2$. Therefore, by the result for $n = 2$, we can conclude that $\sum\limits_{i=1}^{n} \eta_i$ is normal with mean $\sum\limits_{i=1}^{n} \mu_i$ and variance $\sum\limits_{i=1}^{n} \sigma_i^2$.

Example 76. A basketball team will play a 44-game season. Twenty-six of these games are against class A teams and 18 are against class B teams. Suppose that the team will win each game against a class A team with probability 0.4 and will win each game against a class B team with probability 0.7. Assume also that the results of the different games are independent. Approximate the probability that

(a) the team wins 25 games or more;

(b) the team wins more games against class A teams than it does against class B teams.

SOLUTION: (a) Let η_A and η_B denote, respectively, the number of games the team wins against class A and against class B teams. Note that η_A and η_B are independent binomial random variables, and

$$E\eta_A = 26 \cdot 0.4 = 10.4 \qquad Var(\eta_A) = 26 \cdot 0.4 \cdot 0.6 = 6.24$$

$$E\eta_B = 18 \cdot 0.7 = 12.6 \qquad Var(\eta_B) = 18 \cdot 0.7 \cdot 0.3 = 3.78$$

By the normal approximation to the binomial it follows that η_A and η_B will approximately have the same distribution as would independent normal random variable with expectations and variances as given in the preceding. Hence, by Theorem 16, $\eta_A+\eta_B$ will approximately have a normal distribution with mean 23 and variance 10.02. Therefore, letting Z denote a standard normal random variable, we have

$$P(\eta_A + \eta_B \geq 25) = P(\eta_A + \eta_B \geq 24.5) = P\left(\frac{\eta_A + \eta_B - 23}{\sqrt{10.02}} \geq \frac{24.5 - 23}{\sqrt{10.02}}\right) \approx$$

$$\approx P\left(Z \geq \frac{1.5}{\sqrt{10.02}}\right) \approx 1 - P(Z < 0.4739) \approx 0.3178.$$

(b) We note that $\eta_A - \eta_B$ will approximately have a normal distribution with mean -2.2 and variance 10.02. Hence

$$P(\eta_A - \eta_B \geq 1) = P(\eta_A - \eta_B \geq 0.5) = P\left(\frac{\eta_A - \eta_B + 2.2}{\sqrt{10.02}} \geq \frac{0.5 + 2.2}{\sqrt{10.02}}\right) \approx$$

$$\approx P\left(Z \geq \frac{2.7}{\sqrt{10.02}}\right) \approx 1 - P(Z < 0.8530) \approx 0.1968.$$

Therefore, there is approximately a 31.78 percent chance that the team will win at least 25 games, and approximately a 19.68 percent chance that it will win more games against class A teams than against class B teams.

LECTURE 22

The random variable ξ is said to be a *lognormal* random variable with parameters μ and σ if $\log(\xi)$ is a normal random variable with mean μ and variance σ^2. That is, ξ is lognormal if it can be expressed as

$$\xi = e^{\eta},$$

where η is a normal random variable.

Example 77. Starting at some fixed time, let $S(n)$ denote the price of a certain security at the end of n additional weeks, $n \geq 1$. A popular model for the evolution of these prices assumes that the price ratios $S(n)/S(n-1)$, $n \geq 1$, are independent and identically distributed lognormal random variables. Assuming this model, with parameters $\mu = 0.0165$, $\sigma = 0.0730$, what is the probability that

(a) the price of the security increases over each of the next two weeks;

(b) the price at the end of two weeks is higher than it is today?

Solution. Let Z be a standard normal random variable. To solve part (a), we use the fact that $\log(x)$ increases in x to conclude that $x > 1$ if and only if $\log(x) > \log(1) = 0$. As a result, we have

$$P\left(\frac{S(1)}{S(0)} > 1\right) = P\left(\log\left(\frac{S(1)}{S(0)}\right) > 0\right) = P\left(Z > \frac{-0.0165}{0.073}\right) = P(Z < 0.226) = 0.5894.$$

Therefore, the probability that the price is up after one week is 0.5894. Since the successive price ratios are independent, the probability that the price increases over each of the next two weeks is $(0.5894)^2 = 0.3474$.

To solve part (b), reason as follows:

$$P\left(\frac{S(2)}{S(0)} > 1\right) = P\left(\frac{S(2)}{S(1)}\frac{S(1)}{S(0)} > 1\right) = P\left(\log\left(\frac{S(2)}{S(1)}\right) + \log\left(\frac{S(1)}{S(0)}\right) > 0\right).$$

However,

$$\log\left(\frac{S(2)}{S(1)}\right) + \log\left(\frac{S(1)}{S(0)}\right),$$

being the sum of two independent normal random variables with a common mean 0.0165 and a common standard deviation 0.073, is normal random variable with mean 0.033 and variance $2(0.073)^2$. Consequently,

$$P\left(\frac{S(2)}{S(0)} > 1\right) = P\left(Z > \frac{-0.033}{0.073\sqrt{2}}\right) = P(Z < 0.31965) = 0.6254.$$

Example 78. Sums of independent Poisson random variables.

If η_1 and η_2 are independent Poisson random variables with respective parameters λ_1 and λ_2, compute the distribution of $\eta_1 + \eta_2$.

SOLUTION: Because the event $\{\eta_1 + \eta_2 = n\}$ may be written as the union of the disjoint events $\{\eta_1 = k\}$ and $\{\eta_2 = n - k\}$, where $0 \le k \le n$, we have

$$P(\eta_1 + \eta_2 = n) = \sum_{k=0}^{n} P(\eta_1 = k \cap \eta_2 = n-k) = \sum_{k=0}^{n} P(\eta_1 = k)\, P(\eta_2 = n-k) =$$

$$= \sum_{k=0}^{n} e^{-\lambda_1} \frac{\lambda_1^k}{k!} e^{-\lambda_2} \frac{\lambda_2^{n-k}}{(n-k)!} = e^{-(\lambda_1+\lambda_2)} \sum_{k=0}^{n} \frac{\lambda_1^k \lambda_2^{n-k}}{k!(n-k)!} =$$

$$= \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} \lambda_1^k \lambda_2^{n-k} = \frac{e^{-(\lambda_1+\lambda_2)}}{n!} (\lambda_1 + \lambda_2)^n.$$

In words, $\eta_1 + \eta_2$ has a Poisson distribution with parameter $\lambda_1 + \lambda_2$.

Example 79. Sums of independent binomial random variables.

Let η_1 and η_2 be independent binomial random variables with respective parameters (n, p) and (m, p). Calculate the distribution of $\eta_1 + \eta_2$.

SOLUTION: Without any computation at all we can immediately conclude, by recalling the interpretation of a binomial random variable, that $\eta_1+\eta_2$ is binomial with parameters $(n + m, p)$. This follows because η_1 represents the number of successes in n independent trials, each of which results in a success with probability p; similarly, η_2 represents the number of successes in m independent trials, each trial being a success with probability p. Hence, as η_1 and η_2 are assumed independent, it follows that $\eta_1+\eta_2$ represents the number of successes in $n + m$ independent trials when each trial has a probability p of being a success. Therefore, $\eta_1+\eta_2$ is a binomial random variable with parameters $(n+m,p)$. To check this result analytically, note that

$$P(\eta_1+\eta_2=k)=\sum_{i=0}^{k}P(\eta_1=i\cap\eta_2=k-i)=\sum_{i=0}^{k}P(\eta_1=i)\,P(\eta_2=k-i)=$$

$$=\sum_{i=1}^{k}\binom{n}{i}\,p^{i}\,(1-p)^{n-i}\,\binom{m}{k-i}\,p^{k-i}(1-p)^{m-k+i},$$

where $\binom{r}{j}=0$ when $j>r$. Hence

$$P(\eta_1+\eta_2=k)=p^{k}(1-p)^{n+m-k}\sum_{i=0}^{k}\binom{n}{i}\binom{m}{k-i}.$$

and the result follows upon application of the combinatorial identity

$$\binom{n+m}{k}=\sum_{i=0}^{k}\binom{n}{i}\binom{m}{k-i}.$$

Example 80. Consider $n+m$ trials having a common probability of success. Suppose, however, that this success probability is not fixed in advance but it chosen from a uniform (0, 1) population. What is the conditional distribution of the success probability given that the $n+m$ trials result in n successes?

SOLUTION: If we let η denote the trial success probability, then η is the uniform random variable over the interval $(0,1)$. Also, given that $\eta=x$, the $n+m$ trials are independent with common success probability x, and so ξ, the number of successes, is a binomial random variable with parameters $(n+m,x)$. Hence the conditional density of η given that $\xi=n$ is as follows:

$$f_{\eta/\xi}(x/n)=\frac{P(\xi=n/\eta=x)\,f_{\eta}(x)}{P(\xi=n)}=\frac{\binom{n+m}{n}\,x^{n}(1-x)^{m}}{P(\xi=n)}=cx^{n}(1-x)^{m},\quad 0<x<1,$$

where c does not depend on x.

Therefore

$$f_{\eta/\xi}(x/n)=\begin{cases}0 & \text{if}\quad x\notin(0,1)\\ cx^{n}(1-x)^{m}, & \text{if}\quad 0<x<1.\end{cases}$$

Hence the conditional density is that of beta random variable with parameters $n+1$, $m+1$.

The preceding result is quite interesting, for it states that if the original or prior (to the collection of data) distribution of a trial success probability is uniformly distributed over $(0,1)$ [or, equivalently, is beta with parameters (1, 1)] then the posterior (or conditional) distribution given a total of n successes in $n+m$ trials is beta with parameters $(1+n, 1+m)$. This is valuable, for it enhances our intuition as to what it means to assume that a random variable has a beta distribution.

Example 81. Let (η_1, η_2) denote a random point in the plane and assume that the rectangular coordinates η_1 and η_2 are independent standard normal random variables. We are interested in the joint distribution of R, Θ, the polar coordinate representation of this point.
As the joint density function of η_1 and η_2 is

$$f(x,y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$$

It is not difficult to calculate that the joint density function of $R = \sqrt{\eta_1^2 + \eta_2^2}$ and $\Theta = \arctan \dfrac{\eta_2}{\eta_1}$, is given by

$$f(r,\theta) = \begin{cases} \frac{1}{2\pi} r\, e^{-r^2/2}, & \text{if} \quad 0 < \theta < 2\pi, \quad 0 < r < \infty \\ 0, & \text{otherwise.} \end{cases}$$

As this joint density factors into the marginal densities for R and Θ, we obtain that R and Θ are independent random variables, with Θ being uniformly distributed over $(0, 2\pi)$ and R having the Rayleigh distribution with density

$$f(r) = \begin{cases} 0, & \text{if} \quad x \le 0 \\ r\, e^{-r^2/2}, & \text{if} \quad 0 < r < \infty. \end{cases}$$

(Thus, for instance, when one is aiming at a target in the plane, if the horizontal and vertical miss distances are independent standard normals, then the absolute value of the error has the above Rayleigh distribution.)

The preceding result is quite interesting, for it certainly is not evident a priori that a random vector whose coordinates are independent standard normal variables will have an angle of orientation that is not only uniformly distributed, but is also independent of the vector's distance from the origin.

If we wanted the joint distribution of R^2 and Θ, then, as the transformation $D = R^2 = \eta_1^2 + \eta_2^2$ and $\Theta = \arctan \dfrac{\eta_2}{\eta_1}$ give us the following joint density function:

$$f(d,\theta) = \begin{cases} \frac{1}{2}e^{-d/2}\frac{1}{2\pi}, & \text{if} \quad 0 < d < \infty, \quad 0 < \theta < 2\pi \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, $D = R^2$ and Θ are independent, with R^2 having an exponential distribution with parameter $1/2$. But as $R^2 = \eta_1^2 + \eta_2^2$, it follows, by definition, that R^2 has a chi-squared distribution with 2 degrees of freedom. Hence we have a verification of the result that the exponential distribution with parameter $1/2$ is the same as the chi-squared distribution with 2 degrees of freedom.

The preceding result can be used to simulate (or generate) normal random variables by making a suitable transformation on uniform random variables. Let U_1 and U_2 be independent random variables each uniformly distributed over $(0,1)$. We will transform U_1, U_2 into two independent unit normal random variables η_1 and η_2 by first considering the polar coordinate representation (R, Θ) of the random vector (η_1, η_2).

From the above, R^2 and Θ will be independent, and, in addition, $R^2 = \eta_1^2 + \eta_2^2$ will have an exponential distribution with parameter $\lambda = 1/2$. But $-2\log U_1$ has such a distribution since, for $x > 0$,

$$P(-2\log U_1 < x) = P(\log U_1 > -\frac{x}{2}) = P(U_1 > e^{-x/2}) = 1 - e^{-x/2}.$$

Also, as $2\pi U_2$ is a uniform $(0, 2\pi)$ random variable, we can use it to generate Θ. That is, if we let

$$R^2 = -2\log U_1, \quad \Theta = 2\pi U_2,$$

then R^2 can be taken to be a square of the distance from the origin and θ as the angle of orientation of (η_1, η_2).

§31. SIMULATION

As our starting point in simulation, we shall suppose that we can simulate from the uniform $(0,1)$ distribution and we shall use the term "random numbers" to mean independent random variables from this distribution.

A general method for simulating a random variable having a continuous distribution — called the *inverse transformation method* — is based on the following Lemma.

Lemma 8. *Let* $\eta(\omega)$ *be a random variable with continuous distribution function* $F(x)$. *Then the random variable*

$$\zeta(\omega) = F(\eta(\omega))$$

has uniform distribution over $(0, 1)$.

Proof: For any $x \leq 0$ the distribution function F_ζ is equal to 0 and for $x > 1$ F_ζ is equal to 1. Let us consider $x \in (0, 1)$.

$$F_\zeta(x) = P\{\omega\colon\ \zeta(\omega) < x\} = P\{\omega\colon\ F(\eta(\omega)) < x\} = P\{\omega\colon\ \eta(\omega) < F^{-1}(x)\} = F(F^{-1}(x)) = x$$

Above $F^{-1}(x)$ is defined to equal that value y for which $F(y) = x$.

Corollary 14. Let U be the uniform $(0, 1)$ random variable. For any continuous distribution function F if we define the random variable $\eta(\omega)$ by

$$\eta(\omega) = F^{-1}(U)$$

then the random variable $\eta(\omega)$ has distribution function $F(x)$.
The proof immediately follows from Lemma 8.

Example 82 (Simulating an Exponential Random Variable). If $F(x) = 1 - e^{-x}$, then $F^{-1}(u)$ is that value of x such that $1 - e^{-x} = u$ or $x = -\log(1 - u)$.
Hence, if U is a uniform $(0, 1)$ variable, then

$$F^{-1}(U) = -\log(1 - U)$$

is exponentially distributed with mean 1. Since $1 - U$ is also uniformly distributed on $(0, 1)$, it follows that $-\log U$ is exponential with mean 1. Since $c\eta$ is exponential with mean c when η is exponential with mean 1, it follows that $-c\log U$ is exponential with mean c.

Example 83 (Simulating Independent, Standard Normal Random Variables). For the important case of the normal distribution, it is not possible to give a close form expression for the distribution function $F(x)$, and so we cannot solve the equation $F(y) = x$. For this reason, special methods have been developed. One such method relies on the fact that if U and V are independent random variables with uniform densities on $(0, 1)$, then (see the previous section) random variables

$$\eta_1 = \sqrt{-2\log(U)}\,\cos(2\pi V)$$

and

$$\eta_2 = \sqrt{-2\log(U)}\,\sin(2\pi V)$$

are independent, and have normal distribution function with parameters $a = 0$ and $\sigma = 1$.

LECTURE 23

§32. MOMENT GENERATING FUNCTIONS

Studying distributions of random variables and their basic quantitative properties, such as expressions for moments, occupies a central role in both statistics and probability. It turns out that a function called the probability generating function is often a very useful mathematical tool in studying distributions of random variables. It is useful to derive formulas for moments and for the probability mass function of random variables that appear too complicated at first glance. in this section, we introduce the probability generating function, study some of its properties, a nd apply it to a selection of examples. The moment generating function, which is related to the probability generating function, is also extremely useful as a mathematical tool in numerous problems and is also introduced ni this section.

Definition 20. The probability generating function, also called simply the generating function, of nonnegative integer-valued random variable $\eta(\omega)$ is defined as

$$G(s) = G_\eta(s) = E(s^\eta) = \sum_{k=0}^{\infty} s^k\, P(\eta = k) = \sum_{k=0}^{\infty} s^k\, p_k,$$

provided the expectation is finite.

Note, that $G(s)$ is always finite for $|s| \le 1$, but it could be finite over a larger interval, depending on the specific random variable η.

Two basic properties of the generating function are the following.

Theorem 17. a) Suppose $G(s)$ is finite in some open interval containing the origin. Then, $G(s)$ is infinitely differentiable in that open interval, and

$$P(\eta = k) = \frac{G^{(k)}(0)}{k!}, \qquad k \ge 0,$$

where $G^{(0)}(0)$ means $G(0)$.

b) If $\lim_{s\uparrow 1} G^{(k)}(s)$ is finite, then $E[\eta(\eta-1)...(\eta-k+1)]$ exists and is finite, and

$$G^{(k)}(1) = \lim_{s\uparrow 1} G^{(k)}(s) = E[\eta(\eta-1)...(\eta-k+1)].$$

Proof: The infinite differentiability is a fact from the theory of power series that converge in some nonempty open interval. The power series can be differentiated infinitely many times by term by term in that open interval. That

$$P(\eta = k) = \frac{G^{(k)}(0)}{k!}, \qquad k \geq 0,$$

follows on differentiating $G(s)$ term by term k times and setting $s = 0$, while part b) follows on differentiating $G(s)$ k times and letting $s \to 1$.

Definition 21. $E[\eta(\eta-1)...(\eta-k+1)]$ is called the kth factorial moment of $\eta(\omega)$.

Remark. The kth factorial moment of $\eta(\omega)$ exists if and only if the kth moment $E(\eta^k)$ exists.

One of the most important properties of generating functions is the following.

Theorem 18. Let η_1, η_2,..., η_n be independent random variables with generating functions $G_1(s)$, $G_2(s)$,..., $G_n(s)$. Then the generating function of $\eta_1 + \eta_2 + ...\eta_n$ equals

$$G_{\eta_1+\eta_2+...\eta_n}(s) = \prod_{i=1}^{n} G_i(s).$$

Proof: By definition,

$$G_{\eta_1+\eta_2+...\eta_n}(s) = E\left[s^{\eta_1+\eta_2+...\eta_n}\right] = E\left[s^{\eta_1}s^{\eta_2}...s^{\eta_n}\right] = E\left[s^{\eta_1}\right]E\left[s^{\eta_2}\right]...E\left[s^{\eta_n}\right] = \prod_{i=1}^{n} G_i(s)$$

by virtue of the independence of $\eta_1(\omega)$, $\eta_2(\omega)$, ..., $\eta_n(\omega)$, which would imply the independence of s^{η_1}, s^{η_2} ... s^{η_n}.

One reason that the generating function is useful as a tool is its distribution determining property, in the following sense.

Theorem 19. Let $G(s)$ and $H(s)$ be the generating functions of two random variables η_1 and η_2. If $G(s) = H(s)$ in any nonempty open interval, then η_1 and η_2 have the same distribution.

Proof: Let $P(\eta_1(\omega) = n) = p_n$ and $P(\eta_2(\omega) = n) = q_n$, $n \geq 0$. Then

$$G(s) = \sum_{n=0}^{\infty} p_n s^n, \qquad H(s) = \sum_{n=0}^{\infty} q_n s^n.$$

If there is a nonempty open interval in which

$$\sum_{n=0}^{\infty} p_n s^n = \sum_{n=0}^{\infty} q_n s^n,$$

then from the theory of power series, $p_n = q_n$ for any $n \geq 0$, and therefore η_1 and η_2 have the same distribution.

Summarizing, then, one can find from the generating function of a nonnegative integer–valued random variable η_1 the probability mass function of η and every moment of η, including the moments that are infinite.

Example 83 (Discrete uniform distribution). Suppose η has the discrete uniform distribution on $\{1, 2, ..., n\}$. Then its generating function is

$$G(s) = E\left[s^{\eta}\right] = \sum_{k=1}^{n} s^k \, P(\eta = k) = \sum_{k=1}^{n} s^k \, \frac{1}{n} = \frac{1}{n} \sum_{k=1}^{n} s^k = \frac{s\left(s^n - 1\right)}{n\left(s - 1\right)}$$

by summing the geometric series $\sum\limits_{k=1}^{n} s^n$. As a check, if we differentiate $G(s)$ once, we get

$$G'(s) = \frac{1 + s^n\left[n(s-1) - 1\right]}{n\left(s-1\right)^2}.$$

On applying L,Hospital's rule, we get that $G'(1) = \frac{n+1}{2}$, which therefore is the mean of η.

Example 84. Let $G(s) = \frac{(1+s)^n}{2^n}$. Then, by just expanding $(1+s)^n$ using the binomial theorem, we have

$$G(s) = \frac{1}{2^n} \sum_{k=0}^{n} \binom{n}{k} s^k.$$

We now recognize that the coefficients

$$\frac{\binom{n}{k}}{2^n}, \qquad k = 0, 1, ..., n,$$

are all nonnegative and that they do add to one. Therefore, $G(s) = \frac{(1+s)^n}{2^n}$ is a generating function, and indeed it is the generating function of the random variable η with the probability mass function

$$P(\eta = k) = \frac{\binom{n}{k}}{2^n}, \qquad k = 0, 1, 2, ..., n,$$

which is the binomial random variable with parameters n and $1/2$.

Example 85 (The Poisson distribution). Consider a nonnegative integer-valued random variable η with the probability mass function $p_k = e^{-1}\frac{1}{k!}$, $k = 0, 1, 2, \dots$. This is indeed a valid probability mass function. First, it is clear that $p_k \geq 0$ for any k. Also $\sum_{k=0}^{\infty} p_k = 1$. We find the generating function of this distribution. The generating function is

$$G(s) = E\left[s^\eta\right] = \sum_{k=0}^{\infty} s^k e^{-1}\frac{1}{k!} = e^{-1}\sum_{k=0}^{\infty}\frac{s^k}{k!} = e^{-1}e^s = e^{s-1}.$$

The first derivative of $G(s)$ is $G'(s) = e^{s-1}$, and therefore $G'(1) = e^0 = 1$. From our theorem above, we conclude that $E(\eta) = 1$. Indeed, the probability mass function that we have in this example is the probability mass function of the so-called Poisson distribution with mean one. The probability mass function of the Poisson distribution with a general mean λ is

$$p_k = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \dots.$$

The Poisson distribution is an extremely important distribution in probability theory.

§33. CHARACTERISTIC FUNCTION

An important tool in the study of random variables and their probability mass functions or distribution functions is the characteristic function. For any random variable η with distribution function $F(x)$, this is defined to be the function φ on $(-\infty, +\infty)$ as follows:

$$\varphi(t) = E(e^{it\eta}) = \int_{-\infty}^{+\infty} e^{itx}\, dF(x), \qquad t \in \mathbb{R}$$

Therefore, by definition,

$$\varphi(t) = \begin{cases} \displaystyle\sum_{k:\ p(k)>0} e^{itk}\, p(k) & \text{if } \eta \text{ is a discrete random variable} \\ \displaystyle\int_{-\infty}^{+\infty} e^{itx}\, f_\eta(x)\, dx & \text{if } \eta \text{ is an absolutely continuous random variable,} \end{cases}$$

depending on whether η is specified its density function $f_\eta(x)$, or its probability mass function $p(x)$.

In calculus, the characteristic function is known as the Fourier transformation of F.

The characteristic function has the following simple properties:

1) For any $t \in \mathbb{R}$ we have $|\varphi(t)| \leq 1 = \varphi(0)$ and $\varphi(-t) = \overline{\varphi(t)}$, where $\bar{z}$ denotes the conjugate complex of z.

The property follows from the equality $|e^{iy}| = 1$ for any $y \in \mathbb{R}$, since we have the Euler formula:

$$e^{iy} = \cos y + i \sin y.$$

2) φ is uniformly continuous in $\mathbb{R}$.

To see this, we write for real t and h

$$|\varphi(t+h) - \varphi(t)| \leq \int_{-\infty}^{+\infty} |e^{itx}|\,|e^{ihx} - 1| dF(x) = \int_{-\infty}^{+\infty} |e^{ihx} - 1| dF(x).$$

The last integrand is bounded by 2 and tends to 0 as $h \to 0$, for each x. Hence, the integral converges to 0 by bounded convergence. Since it does not involve t, the convergence is surely uniform with respect to t.

3) If we write φ_η for the characteristic function of η, then for any real numbers a and b, we have

$$\varphi_{a\eta+b}(t) = \varphi_\eta(at)\, e^{itb},$$

and, in particular,

$$\varphi_{-\eta}(t) = \overline{\varphi_\eta(t)}.$$

Indeed,

$$\varphi_{a\eta+b}(t) = E(e^{it(a\eta+b)}) = E(e^{i(ta)\eta} \cdot e^{itb}) = e^{itb}\, E(e^{i(ta)\eta}) = e^{itb}\, \varphi_\eta(at).$$

4) If $\{\varphi_n, n \geq 1\}$ are characteristic functions, $\lambda_n \geq 0$, and $\sum\limits_{n=1}^{\infty} \lambda_n = 1$, then

$$\sum_{n=1}^{\infty} \lambda_n\, \varphi_n(t)$$

is a characteristic function.

5) If $\{\varphi_j(t), 1 \leq j \leq n\}$ are characteristic functions, then

$$\prod_{j=1}^{n} \varphi_j(t)$$

is a characteristic function.

Proof: There exist independent random variables $\{\eta_j, 1 \le j \le n\}$ with distribution functions $\{F_j(x), 1 \le j \le n\}$. Letting

$$S_n = \sum_{j=1}^{n} \eta_j,$$

we have

$$E\left[e^{itS_n}\right] = E\left(\prod_{j=1}^{n} e^{it\eta_j}\right) = \prod_{j=1}^{n} E\left(e^{it\eta_j}\right) = \prod_{j=1}^{n} \varphi_j(t),$$

or

$$\varphi_{S_n}(t) = \prod_{j=1}^{n} \varphi_j(t).$$

Definition 22. The convolution of two distribution functions $F_1(x)$ and $F_2(x)$ is defined to be the distribution function $F(x)$ such that

$$F(x) = \int_{-\infty}^{+\infty} F_1(x-y)\, dF_2(y),$$

and written as

$$F = F_1 \star F_2.$$

It is easy to verify that $F(x)$ is indeed a distribution function (for details see §30).

Theorem 20. Let η_1 and η_2 be independent random variables with distribution functions $F_1(x)$ and $F_2(x)$, respectively. Then $\eta_1+\eta_2$ has the distribution function $F_1 \star F_2$.

Corollary 15. The binary operation of convolution $\star$ is commutative and associative.

Definition 23. The convolution of two density functions $f_1(x)$ and $f_2(x)$ is defined to be the density function $f(x)$ such that

$$f(x) = \int_{-\infty}^{+\infty} f_1(x-y)\, f_2(y)\, dy,$$

and written as

$$f = f_1 \star f_2.$$

We leave to the reader to verify that $f(x)$ is indeed a density function (see (109) in §30).

Theorem 21. The convolution of two absolutely continuous distribution functions with densities $f_1(x)$ and $f_2(x)$ is absolutely continuous with density $f_1 \star f_2$.

Theorem 22. Addition of (a finite number of) independent random variables corresponds to convolution of their distribution functions and multiplication of their characteristic functions.

Corollary 16. If $\varphi(x)$ is a characteristic function, then so is $|\varphi(x)|^2$.

Proof: Consider two independent identically distributed random variables η_1 and η_2. Let η_i has the characteristic function $\varphi(t)$, $i = 1, 2$. The characteristic function of the difference $\eta_1 - \eta_2$ is

$$E\left(e^{it(\eta_1-\eta_2)}\right) = E\left(e^{it\eta_1}\right)\,E\left(e^{-it\eta_2}\right) = \varphi(t)\,\varphi(-t) = |\varphi(t)|^2.$$

The technique of considering $\eta_1 - \eta_2$ and $|\varphi(t)|^2$ instead η_1 and $\varphi(t)$ will be used below and referred to as "symmetrization". This is often expedient, since a real and particularly a positive-valued characteristic function such as $|\varphi(t)|^2$ is easier to handle than a general one.

Let us list a few well-known characteristic functions together with their distribution functions or density functions, the last being given in the interval outside of which they vanish.

2) If $\eta(\omega) = c$ is a constant, then characteristic function equals

$$e^{ict}.$$

2) If η is a discrete random variable with probability mass function

$$\begin{array}{lcc} \eta(\omega): & -1 & +1 \\ & \frac{1}{2} & \frac{1}{2}, \end{array}$$

then characteristic function is $\cos t$, because

$$\cos t = \frac{e^{it} + e^{-it}}{2}.$$

3) The characteristic function of the binomial distribution has the following form:

$$(1-p+p\,e^{it})^n.$$

4) The characteristic function of the Poisson distribution with mean λ has the form:

$$e^{\lambda\,(e^{it}-1)}.$$

5) The characteristic function of the exponential distribution with mean λ^{-1}:

$$\left(1-\frac{it}{\lambda}\right)^{-1}.$$

6) The characteristic function of the uniform distribution over the interval $[-a,a]$ has the following form:

$$\frac{\sin at}{at}.$$

7) For normal distribution with mean a and variance σ^2, characteristic function has the form:

$$\exp\left(iat-\frac{\sigma^2\,t^2}{2}\right).$$

Therefore for standard normal distribution the characteristic function is

$$e^{-t^2/2}.$$

The following question arises: given a characteristic function φ, how can we find the corresponding distribution function? The formula for doing this, called the inversion formula, is of theoretical importance, since it will establish a one-to -one correspondence between the class of distribution functions and the class of characteristic functions.

Theorem 23. If $x_1 < x_2$ are points of continuity of distribution function $F(x)$, then we have

$$F(x_2)-F(x_1)=\frac{1}{2\,\pi}\lim_{c\to\infty}\int_{-c}^{c}\frac{e^{-itx_1}-e^{-itx_2}}{it}\,\varphi(t)\,dt.$$

We give next an important particular case of Theorem 23.

Theorem 24. If there exists $f(x)$ the density function of $F(x)$ then the inversion formula has the following form:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx}\, \varphi(t)\, dt.$$

Theorem 25. $\eta(\omega)$ **is symmetric with respect to the origin if and only if its characteristic function is real-valued for all** t**.**

A complex-valued function φ defined on the real axis is called positive definite if for any finite set of real numbers t_j and complex numbers z_j (with conjugate complex $\bar{z}_j$), $1 \le j \le n$, we have

$$\sum_{j=1}^{n} \sum_{k=1}^{n} \varphi(t_j - t_k)\, z_j\, \overline{z}_k \ge 0.$$

Theorem 26. $\varphi(t)$ **is a characteristic function if and only if it is positive definite and continuous at** 0 **with** $\varphi(0) = 1$**.**

L E C T U R E 24

§34. MARKOV CHAINS

There is an increasing interest in the study of systems which vary in time in a random manner. Mathematical models of such systems are known as stochastic processes. We have tried to select topics that are conceptually interesting and that have found fruitful application in various branches of science and technology.

A stochastic process can be defined quite generally as any collection of random variables $\eta(t, \omega)$, $t \in T$, defined on a common probability space, where T is a subset of $(-\infty, +\infty)$ and is thought of as the time parameter set. The process is called a ***continuous parameter process*** if T is an interval having positive length and a ***discrete parameter process*** if T is a subset of the integers. If the random variables $\eta(t, \omega)$ all take on values from the fixed set $\mathcal{G}$, then $\mathcal{G}$ is called the ***state space*** of the process.

Many stochastic processes of theoretical and applied interest possesses the property that, given the present state of the process, the past history does not affect conditional probabilities of events defined in terms of the future. Such processes are called Markov processes. In sections 34— 36 we study Markov chains, which are discrete parameter Markov processes whose state space is finite or countable infinite.

Consider a system that can be in any one of a finite or countably infinite number of states. Let $\mathcal{G}$ denote this set of states. We can assume that $\mathcal{G}$ is a subset of the integers. Let the system be observed at the discrete moments of time $n = 0, 1, 2, ...$, and let $\eta_n(\omega) = \eta(n, \omega)$ denote the state of the system at time n. If $\eta_n(\omega) = i$, then the process is said to be in state i at time n.

Since we are interested in non–deterministic systems, we think of $\eta_n(\omega)$, $n \geq 0$, as random variables defined on a common probability space. Little can be said about such random variables unless some additional structure is imposed upon them.

The simplest possible structure is that of independent random variables. This would be a good model for such systems as repeated experiments in which future states of the system are independent of past and present states. In most systems that arise in practice, however, past and present states of the system influence the future states even if they do not uniquely determine them.

Many systems have the property that given the present state, the past states have no influence on the future. This property is called the *Markov property*, and systems having this property are called *Markov chains*. The Markov property is defined precisely by the requirement that

$$\begin{aligned} &P\left\{\eta_{n+1}(\omega)=j\,/\eta_0(\omega)=i_0,\eta_1(\omega)=i_1,...,\eta_{n-1}(\omega)=i_{n-1},\eta_n(\omega)=i\,\right\}= \\ &=P\left\{\eta_{n+1}(\omega)=j\,/\eta_n(\omega)=i\,\right\} \end{aligned} \tag{110}$$

for every choice of the nonnegative integer n and the numbers i_0, i_1, ..., i_{n-1}, i, j, each in $\mathcal{G}$. The conditional probabilities $P\left(\eta_{n+1}(\omega)=j\,/\eta_n(\omega)=i\,\right)$ are called the *one–step transition probabilities* of the chain. We study Markov chains having *stationary* transition probabilities, that is, those such that

$$P\left(\eta_{n+1}(\omega)=j\,/\eta_n(\omega)=i\,\right)=P_{i,j}$$

is independent of n. From now on, when we say that $\eta_n(\omega)$, $n\geq 0$ forms a Markov chain, we mean that these random variables satisfy the Markov property and have stationary transition probabilities.

Equation (110) may be interpreted as stating that, for a Markov chain, the conditional distribution of any future state $\eta_{n+1}(\omega)$ given the past states $\eta_0(\omega)$, $\eta_1(\omega)$, ..., $\eta_{n-1}(\omega)$ and the present state $\eta_n(\omega)$, is independent of the past states and depends only on the present state.

The value $P_{i,j}$ represents the probability that a system in state i will enter state j at the next transition. Since probabilities are nonnegative and since the process must make a transition into some state, we have that

$$P_{i,j}\geq 0 \quad \text{for any} \quad i,j\in\mathcal{G} \tag{111}$$

and

$$\sum_j P_{i,j}=1 \quad \text{for any} \quad i\in\mathcal{G}. \tag{112}$$

Let ¶ denote the matrix of one–step transition probabilities $P_{i,j}$, so that

$$¶ = \|P_{i,j}\|_{i,j \in \mathcal{G}}.$$

The function $\pi_0(i)$, $i \in \mathcal{G}$, defined by

$$\pi_0(i) = P\{\eta_0(\omega) = i\}, \qquad i \in \mathcal{G}$$

is called the initial distribution of the Markov chain. It is such that

$$\pi_0(i) \geq 0 \quad \text{for any} \quad i \in \mathcal{G} \tag{113}$$

and

$$\sum_i \pi_0(i) = 1. \tag{114}$$

The joint distribution of $\eta_0(\omega)$, ..., $\eta_n(\omega)$ can easily be expressed in terms of the transition function and the initial distribution. For example,

$$P\{\eta_0(\omega) = i_0, \eta_1(\omega) = i_1\} = P\{\eta_0(\omega) = i_0\} \cdot P\{\eta_1(\omega) = i_1 \, / \eta_0(\omega) = i_0\} = \pi_0(i_0) \cdot P_{i_0,i_1}.$$

By induction it is easily seen that

$$P\{\eta_0(\omega) = i_0, \eta_1(\omega) = i_1, ..., \eta_n(\omega) = i_n\} = \pi_0(i_0) \cdot P_{i_0,i_1} \cdot P_{i_1,i_2} \cdot ... \cdot P_{i_{n-1},i_n}. \tag{115}$$

It is usually more convenient, however, to reverse the order of our definitions. We say that $P_{i,j}$, $i \in \mathcal{G}$ and $j \in \mathcal{G}$, is a *transition function* if it satisfies (111) and (112), and we say that $\pi_0(i)$, $i \in \mathcal{G}$, is an *initial distribution* if it satisfies (113) and (114). It can be shown that given any transition function $P_{i,j}$ and any initial distribution $\pi_0(i)$, there is a probability space and random variables $\eta_n(\omega)$, $n \geq 0$, defined on that space satisfying (115). It is not difficult to show that, these random variables form Markov chain having transition function $P_{i,j}$ and initial distribution $\pi_0(i)$.

A state i of a Markov chain is called an absorbing state if $P_{i,i} = 1$ or, equivalently, if $P_{i,j} = 0$ for $j \neq i$.

Example 86 (A Gambling model). Consider a gambler who, at each play of the game, either wins 1 unit with probability p or loses 1 unit with probability $1-p$. If we suppose that the gambler will quit playing when his fortune hits either 0 or M, then the gambler's sequence of fortunes is a Markov chain having transition probabilities

$$P_{i,i+1} = p = 1 - P_{i,i-1}, \qquad i = 1, ..., M-1, \qquad P_{0,0} = P_{M,M} = 1.$$

In this case 0 and M are both absorbing states.

We can also define the two–stage transition probability, $P_{i,j}(2)$, that a system, presently in state i, will be in state j after two additional transitions. That is,

$$P_{i,j}(2) = P\left\{\eta_{m+2}(\omega) = j \,/\eta_m(\omega) = i\right\}.$$

In general, we define the n–stage transition probabilities, denoted as $P_{i,j}(n)$, that a process in state i will be in state j after n additional transitions:

$$P_{i,j}(n) = P\left\{\eta_{m+n}(\omega) = j \,/\eta_m(\omega) = i\right\}$$

and $P_{i,j}(1) = P_{i,j}$.

Theorem 27 (Chapman–Kolmogorov equations).

$$P_{i,j}(n) = \sum_{k\in\mathcal{G}} P_{i,k}(r)\cdot P_{k,j}(n-r) \tag{116}$$

for all $0 < r < n$.

The Chapman–Kolmogorov equations (116) provide a method for computing n–step transition probabilities. Equations (116) are most easily understood by noting that $P_{i,k}(r)\cdot P_{k,j}(n-r)$ represents the probability that starting in i the process will go to state j in n transitions through a path which takes it into state k at the rth transition. Hence, summing over all intermediate states k yields the probability that the process will be in state j after n transitions. Formally, we have

$$P_{i,j}(n) = P(\eta_n(\omega) = j \,/\eta_0(\omega) = i\,) = \sum_{k\in\mathcal{G}} P(\eta_n(\omega) = j, \eta_r(\omega) = k \,/\eta_0(\omega) = i\,) =$$
$$= \sum_{k\in\mathcal{G}} P(\eta_r(\omega) = k \,/\eta_0(\omega) = i\,)\cdot P(\eta_n(\omega) = j \,/\eta_r(\omega) = k\,) = \sum_{k\in\mathcal{G}} P_{i,k}(r)\cdot P_{k,j}(n-r).$$

If we let $\P(n)$ denote the matrix of n–step transition probabilities $P_{i,j}(n)$, then equation (116) asserts that

$$\P(n) = \P(r) \cdot \P(n-r),$$

where the dot represents matrix multiplication[2]. Hence, in particular $(n = 2,\ r = 1)$

$$\P(2) = \P \cdot \P = \P^2$$

and by induction

$$\P(n) = \P^{n-1} \cdot \P = \P^n.$$

That is, the n–step transition probability matrix may be obtained by multiplying the matrix $\P$ by itself n times.

Example 87. Let $\{\eta_n(\omega); n \in \mathbb{N}\}$ be a Markov chain with three states and transition matrix

$$\P = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{3}{5} & \frac{2}{5} & 0 \end{pmatrix}$$

Then

$$P(\eta_1 = 2 \cap \eta_2 = 3 \cap \eta_3 = 1 \cap \eta_4 = 3 \cap \eta_5 = 1 \cap \eta_6 = 3 \cap \eta_7 = 2 / \eta_0 = 3) =$$
$$= p_{3,2} \cdot p_{2,3} \cdot p_{3,1} \cdot p_{1,3} \cdot p_{3,1} \cdot p_{1,3} \cdot p_{3,2} = \frac{2}{5} \cdot \frac{1}{3} \cdot \frac{3}{5} \cdot \frac{1}{4} \cdot \frac{3}{5} \cdot \frac{1}{4} \cdot \frac{2}{5} = \frac{3}{2,500}.$$

The two–step transition probabilities are given by

$$\P(2) = \P^2 = \begin{pmatrix} \frac{17}{30} & \frac{9}{40} & \frac{5}{24} \\ \frac{8}{15} & \frac{3}{10} & \frac{1}{6} \\ \frac{17}{30} & \frac{3}{20} & \frac{17}{60} \end{pmatrix}$$

so that, for example $P_{2,3}(2) = \frac{1}{6}$.

[2]If $\mathbf{A}$ is an $n \times m$ matrix whose element in the ith row and jth column is a_{ij} and $\mathbf{B}$ is an $m \times k$ matrix whose element in the ith row and jth column is b_{ij}, then $\mathbf{A} \cdot \mathbf{B}$ is defined to be the $n \times k$ matrix whose element in the ith row and jth column is $\sum_{k=1}^{m} a_{ik} \cdot b_{kj}$

Example 88 (Number of successes in Bernoulli process). Let $\eta_n(\omega)$ denote the number of successes in n independent trials, where the probability of a success in any one trial is p. The sequence $\{\eta_n(\omega); n \in \mathbb{N}\}$ is a Markov chain. Here the state space is $\{0, 1, 2, ...\}$, the initial distribution is $\pi(0) = 1$ and $\pi(j) = 0$ for $j \geq 1$. The transition matrix is

$$\begin{pmatrix} 1-p & p & 0 & ... & 0 \\ 0 & 1-p & p & ... & 0 \\ 0 & 0 & 1-p & ... & 0 \\ 0 & ... & 0 & 1-p & p \end{pmatrix}$$

If the state space $\mathcal{G}$ of a Markov chain is finite, then computing $P_{i,j}(n)$ is relatively straightforward.

Example 89. We observe the state of a system at discrete points in time. We say that the system is in state 1 if it is operating properly. If the system is undergoing repair (following a breakdown), then the system state is denoted by state 2. If we assume that the system possesses the Markov property, then we have a two-state (homogeneous) Markov chain. We have

$$\P = \begin{pmatrix} 1-p_{1,2} & p_{1,2} \\ p_{2,1} & 1-p_{2,1} \end{pmatrix}$$

In this particular case we can compute $P_{i,j}(n)$. We will impose the condition

$$|1 - p_{1,2} - p_{2,1}| \neq 1 \quad \text{or} \quad < 1 \tag{117}$$

on the one–step transition probabilities.
If $|1 - p_{1,2} - p_{2,1}| < 1$, then n–step transition probability matrix $\P(n) = \P^n$ is given by:

$$\P(n) = \begin{pmatrix} \frac{p_{2,1} + p_{1,2}\,(1 - p_{1,2} - p_{2,1})^n}{p_{1,2} + p_{2,1}} & \frac{p_{1,2} - p_{1,2}\,(1 - p_{1,2} - p_{2,1})^n}{p_{1,2} + p_{2,1}} \\ \frac{p_{2,1} - p_{2,1}\,(1 - p_{1,2} - p_{2,1})^n}{p_{1,2} + p_{2,1}} & \frac{p_{1,2} + p_{2,1}\,(1 - p_{1,2} - p_{2,1})^n}{p_{1,2} + p_{2,1}} \end{pmatrix} \tag{118}$$

Since $p_{1,2}$ and $p_{2,1}$ are probabilities, condition (117) can be violated only if $p_{1,2} = p_{2,1} = 0$ or $p_{1,2} = p_{2,1} = 1$. These two cases is treated separately.

Example 90. Let $p_{1,2} = p_{2,1} = 0$. Clearly, $|1 - p_{1,2} - p_{2,1}| = 1$, and therefore (118) does not apply. The transition probability matrix ¶ is the identity matrix:

$$\P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

The two states do not communicate with each other. The matrix $\P(n) = \P^n$ is easily seen to be the identity matrix. In other words, the chain never changes state.

Example 91. Let $p_{1,2} = p_{2,1} = 1$. Clearly, $|1 - p_{1,2} - p_{2,1}| = 1$, and therefore (118) does not apply. The transition probability matrix ¶ is given by:

$$\P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

It can be verified by induction that:

$$\P(n) = \P^n = \begin{cases} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} & \text{if } n \text{ is even} \\ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} & \text{if } n \text{ is odd} \end{cases}$$

This Markov chain has an interesting behavior. Starting in state 1 (or 2), we return to state 1 (or 2) after an even number of steps. Therefore, the time between visits to a given state exhibits a periodic behavior. Such a chain is called a periodic Markov chain (with period 2). (Formal definition is given below).

L E C T U R E 25

TYPICAL PROBLEMS OF MID–TERM EXAMINATION-2

PROBLEM 1. A joint density function of random variables η_1 and η_2 is given by

$$f(x,y) = \begin{cases} k\,x & \text{if} \quad 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

where k is a constant. What is $\mathrm{Cov}(\eta_1, \eta_2)$?

PROBLEM 2. The loss due to a fire in a commercial building is modeled by a random variable η with density function

$$f(x) = \begin{cases} 0.005\,(20 - x) & \text{for} \quad 0 < x < 20 \\ 0 & \text{otherwise} \end{cases}$$

Given that a fire loss exceeds 8, what is the probability that it exceeds 16?

PROBLEM 3. The random variables η_1 and η_2 are normal with standard deviations are 2 and 3 respectively and correlation 0.4. Calculate the variance of

$$\frac{\eta_1}{2} - \eta_2.$$

PROBLEM 4. A random variable η has a normal distribution with mean 2 and variance 9 (that is $N(2,9)$). Find z, if the probability that η takes on the value greater than z is 0.5517 (that is $P(\eta > z) = 0.5517$).

PROBLEM 5. If η is uniformly distributed over the interval $(-1, 1)$, find probability that $|\eta|$ is greater than 1/5?

PROBLEM 6. A group insurance policy covers the medical claims of the employees of a small company. The value, V, of the claims made in one year is described by

$$V = 100,000\,\eta,$$

where η is a random variable with density function

$$f(y) = \begin{cases} k\,(1 - y)^4 & \text{if} \quad 0 < y < 1, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant. What is the conditional probability that V exceeds 40,000, given that V exceeds 10,000?

ANSWERS

1. 0

2. 1/9

3. 7.6

4. 1.61

5. 4/5

6. 0.132

LECTURE 26

§35. CLASSIFICATION OF STATES

State j is said to be ***accessible*** from state i if $P_{i,j}(n) > 0$ for some $n \geq 0$. Note that this implies that state j is accessible from state i if and only if, starting in i, it is possible that the process will ever enter state j. This is true since if j is not accessible from i, then

$$P(\text{ever enter } j\,/\text{start in } i) = P(\bigcup_{n=0}^{\infty}\{\eta_n(\omega) = j\,/\eta_0(\omega) = i\}) \leq$$
$$\leq \sum_{n=0}^{\infty} P(\eta_n(\omega) = j\,/\eta_0(\omega) = i) = \sum_{n=0}^{\infty} P_{i,j}(n) = 0.$$

Two states i and j that are accessible to each other are said to be ***communicate***. Note that any state communicates with itself since, by definition,

$$P_{i,i}(0) = P(\eta_0(\omega) = i\,/\eta_0(\omega) = i) = 1.$$

The relation of communication satisfies the following three properties:
1) State i communicates with state i for all $i \in \mathcal{G}$;
2) If state i communicates with state j, then state j communicates with state i;
3) If state i communicates with state j, and state j communicates with state k, then state i communicates with state k.

Properties 1) and 2) follow immediately from the definition of communication. To prove 3) suppose that i communicates with j, and j communicates with k. Thus, there exist integers n and m such that $P_{i,j}(n) > 0$ and $P_{j,k}(m) > 0$. Now by (116), we have that

$$P_{i,k}(n+m) = \sum_{l \in \mathcal{G}} P_{i,l}(n)P_{l,k}(m) \geq P_{i,j}(n)P_{j,k}(m).$$

Hence, state k is accessible from state i. Similarly, we can show that state i is accessible from state k. Hence, states i and k communicates.
Two states that communicate are said to be in the same ***class***. It is an easy consequence of 1), 2), and 3) that any two classes of states are either identical or disjoint. Therefore, the concept of communication is an equivalent relation. In other words, the concept of

communication divides the state space up into a number of separate classes. The Markov chain is said to be *irreducible* if there is only one class, that is, if all states communicate with each other.

Consider an arbitrary, but fixed state i. We define, for each integer $n \geq 1$,

$$f_i(n) = P\{\eta_n = i, \eta_k \neq i \quad \text{for any} \quad k = 1, 2, ..., n-1 \,/ \eta_0 = i\}.$$

In other words, $f_i(n)$ is the probability that, starting from state i, the first return to state i occurs at the nth transition. Clearly, $f_i(1) = P_{ii}$ and $f_i(n)$ may be calculated recursively according to

$$P_{ii}(n) = \sum_{k=0}^{n} f_i(k)\, P_{ii}(n-k), \qquad n \geq 1 \tag{119}$$

where we define $f_i(0) = 0$ for all i. Equation (119) is derived by decomposing the event from which $P_{ii}(n)$ is computed according to the time of the first returns to state i. Indeed, consider all the possible realizations of the process for which $\eta_0 = i$, $\eta_n = i$ and the first return to state i occurs at the kth transition. Call this event B_k. The events B_k, $k = 1, 2, ..., n$ are clearly mutually exclusive. The probability of the event that the first return is at the kth transition is by definition $f_i(k)$. In the remaining $n - k$ transitions, we are dealing only with those realizations for which $\eta_n = i$. Using the Markov property, we have

$$P(B_k) = P(\text{first return is at } k\text{th transition} \,/ \eta_0 = i)\, P(\eta_n = i \,/ \eta_k = i) = f_i(k)\, P_{ii}(n-k)$$

(recall that $P_{ii}(0) = 1$). Hence

$$P(\eta_n = i \,/ \eta_0 = i) = \sum_{k=1}^{n} P(B_k) = \sum_{k=1}^{n} f_i(k)\, P_{ii}(n-k) = \sum_{k=0}^{n} f_i(k)\, P_{ii}(n-k),$$

since by definition $f_i(0) = 0$.

When the process starts from state i, the probability that the process will ever reenter state i is

$$F_i = \sum_{n=0}^{\infty} f_i(n) = \lim_{N \to \infty} \sum_{n=0}^{N} f_i(n).$$

State i is said to be ***recurrent*** if $F_i = 1$, and ***transient***, if $F_i < 1$. Therefore, a state i is recurrent if and only if, after the process starts from state i, the probability of its returning to state i after some finite length of time is one.

Suppose that the process starts in state i and i is recurrent. Hence, with probability 1, the process will eventually reenter state i. However, by the definition of a Markov chain, it follows that the process will be starting over again when it reenters state i and, therefore, state i will eventually be visited again. Continual repetition of this argument leads to the conclusion that if state i is recurrent then, starting in state i, the process will reenter state i infinitely often.

On the other hand, suppose that state i is transient. Hence, each time the process enters state i there will be a positive probability, namely $1 - F_i$, that it will never again enter that state. Therefore, starting in state i, the probability that the process will be in state i for exactly n time period equals $F_i^{n-1} \cdot (1 - F_i)$, $n \geq 1$. In other words, if state i is transient then, starting in state i, the number of time periods that the process will be in state i has a geometric distribution with finite mean $1/(1 - F_i)$.

Consider a transient state i. Then the probability that a process starting from state i returns to state i at least once is $F_i < 1$. Because of the Markov property, the probability that the process returns to state i at least twice is $(F_i)^2$, and, repeating the argument, we see that the probability that the process returns to i at least k times is $(F_i)^k$ for $k = 1, 2, ...$

For a recurrent state i, $p_{i,i}(n) > 0$ for some $n \geq 1$. Define the period of state i, denoted by d_i, as the greatest common divisor of the set of positive integers n such that $p_{i,i}(n) > 0$.

A recurrent state i is said to be aperiodic if its period $d_i = 1$, and ***periodic***, if $d_i > 1$.

In Example 91 both states 0 and 1 are periodic with period 2. States of Example 90 are all aperiodic.

Example 92. Consider the Markov chain consisting of the three states $1, 2, 3$ and having transition probability matrix

$$\P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/4 & 1/4 \\ 0 & 1/3 & 2/3 \end{pmatrix}$$

It is easy to verify that this Markov chain is irreducible. It is possible to go from state 1 to state 3 since $P_{1,2} = 1/2$, $P_{2,3} = 1/4$. That is, one way of getting from state 1 to state 3 is to go from state 1 to state 2 (with probability $1/2$) and then go from state 2 to state 3 (with probability $1/4$).

Example 93. Consider a Markov chain consisting of the four states $1, 2, 3, 4$, and having a transition probability matrix

$$¶ = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The classes of this Markov chain are $\{1, 2\}$, $\{3\}$ and $\{4\}$. Note that while state 1 (or 2) is accessible from state 3, the reverse is not true. Since state 4 is an absorbing state, that is, $P_{4,4} = 1$, no other state is accessible from it.

Thus, for irreducible Markov chain every state can be reached from every other state in a finite number of steps. In other words, for all $i, j \in \mathcal{G}$, there is an integer $n \geq 1$ such that $p_{i,j}(n) > 0$.

It follows that state i is recurrent if and only if, starting in state i, the expected number of time periods that the process is in state i is infinite. Letting

$$A_n = \begin{cases} 1 & \text{if} \quad \eta_n(\omega) = i, \\ 0 & \text{if} \quad \eta_n(\omega) \neq i \end{cases}$$

we have that $\sum\limits_{n=0}^{\infty} A_n$ represents the number of periods that the process is in state i. Also

$$E\left[\sum_{n=0}^{\infty} A_n/\eta_0 = i\right] = \sum_{n=0}^{\infty} E\left[A_n/\eta_0 = i\right] =$$
$$= \sum_{n=0}^{\infty} P\{\eta_n = i/\eta_0 = i\} = \sum_{n=0}^{\infty} P_{ii}(n).$$

We have proved the following theorem.

Theorem 28. State i is recurrent if and only if

$$\sum_{n=0}^{\infty} P_{ii}(n) = +\infty.$$

Equivalently, state i is transient if and only if

$$\sum_{n=0}^{\infty} P_{ii}(n) < +\infty.$$

Theorem 29 (Solidarity). All states of an irreducible Markov chain are of the same type:

1) If one state of an irreducible Markov chain is periodic, then all states are periodic and have the same period.

2) If one state is recurrent, then so are all states.

Example 94. Consider a two–state Markov chain with $p_{1,2} = 0$ and $p_{2,1} = 1$, that is

$$\P = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$$

In this case the state 2 is transient and state 1 is absorbing. The chain is not irreducible, but the limiting state probabilities exist (since $\P(n) = \P$) and are given by $p_1 = 1$ and $p_2 = 0$. This says that eventually the chain will remain in state 1 (after at most one transition).

Example 95 (A random walk). Consider a Markov chain whose state space consists of the integers $\{0, \pm 1, \pm 2, ...\}$ and have transition probabilities given by

$$P_{i,i+1} = p = 1 - P_{i,i-1}, \qquad i = 0, \pm 1, \pm 2, ...,$$

where $0 < p < 1$.

In other words, at each transition the particle moves with probability p one unit to the right and with probability $1 - p$ one unit to the left. One colorful interpretation of this process is that it represents the wanderings of a drunken man as he walks along a straight line. Another is that it represents the winnings of a gambler who on each play of the game either wins or loses one dollar.

Since all states clearly communicate it follows from Theorem 18 that they are either all transient or all recurrent. So let us consider state 0 and attempt to determine if $\sum\limits_{n=0}^{\infty} P_{00}(n)$ is finite or infinite. Obviously, we have

$$P_{00}(2n+1) = 0, \qquad n = 1, 2, ...$$

On the other hand, we have

$$P_{00}(2n) = \binom{2n}{n} p^n \, (1-p)^n.$$

Using now Stirling's formula

$$n! = n^{n+1/2} e^{-n} \sqrt{2\,\pi}$$

we obtain

$$P_{00}(2n) \approx \frac{(4p(1-p))^n}{\sqrt{\pi\, n}}.$$

It is obvious that $4p(1-p) \le 1$ with equality holding if and only if $p = 1/2$. Hence

$$\sum_{n=0}^{\infty} P_{00}(n) = +\infty$$

if and only if $p = 1/2$. Therefore from Theorem 18, the one-dimensional random walk is recurrent if and only if $p = 1/2$. Intuitively, if $p \neq 1/2$ there is positive probability that a particle initially at the origin will drift to $+\infty$ if $p > 1/2$ (to $-\infty$ if $p < 1/2$) without ever returning to the origin.

LECTURE 27

§36. LIMITING DISTRIBUTIONS

For a Markov chain with a countably infinite state space $\mathcal{G}$, computation of $P_{i,j}(n)$ poses problems.

For a large number of Markov chains it turns out that $P_{i,j}(n)$ converges, as $n \to \infty$, to a value p_j that depends only on j. That is, for large values of n, the probability of being in state j after n transitions is approximately equal to p_j no matter what the initial state was. It can be shown that for a finite element state space $\mathcal{G}$, a sufficient condition for a Markov chain to possess this property is that for some $n > 0$,

$$P_{i,j}(n) > 0 \quad \text{for all} \quad i, j \in \mathcal{G}. \tag{120}$$

Markov chains that satisfy (120) are said to be ergodic. Since Theorem 17 yields

$$P_{i,j}(n+1) = \sum_{k \in \mathcal{G}} P_{i,k}(n) \cdot P_{k,j}$$

it follows, by letting $n \to \infty$, that for ergodic chains

$$p_j = \sum_{k \in \mathcal{G}} p_k \cdot P_{k,j}. \tag{121}$$

Furthermore, since $1 = \sum_{j \in \mathcal{G}} P_{i,j}(n)$, we also obtain, by letting $n \to \infty$,

$$\sum_{j \in \mathcal{G}} p_j = 1. \tag{122}$$

In fact, it can be shown that the p_j, $j \in \mathcal{G}$ are the unique nonnegative solutions of equations (121) and (122). All this is summed up in Theorem 30, which we state without proof.

Theorem 30. For an ergodic Markov chain $p_j = \lim_{n\to\infty} P_{i,j}(n)$ exists, and the p_j, $j \in \mathcal{G}$ are the unique nonnegative solutions of equations (121) and (122).

Returning to the Markov chain of Example 94, we see that Theorem 30 does not apply. Although the limits

$$\lim_{n\to\infty} p_{i,j}(n)$$

of $p_{i,j}(n)$ exist (in fact $¶(n) = ¶$), the limit is dependent upon the initial distribution. We consider the two–state Markov chain of Example 89 with the condition (117). This implies (118). From this we conclude that

$$p_1 = \frac{p_{2,1}}{p_{1,2} + p_{2,1}} \quad \text{and} \quad p_2 = \frac{p_{1,2}}{p_{1,2} + p_{2,1}}. \tag{123}$$

This result can also be derived using the system (121), (122). Using (121) we obtain

$$p_1 = (1 - p_{1,2})\, p_1 + p_{2,1}\, p_2$$

$$p_2 = p_{1,2}\, p_1 + (1 - p_{2,1})\, p_2$$

Note that these two equations are linearly dependent, and thus we need one more equation (supplied by the condition (122)):

$$p_1 + p_2 = 1.$$

Solving we get the same limiting probability distribution (123).

Example 96. Let for a Markov chain with three states and the one-step transition probabilities matrix:

$$¶ = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

It is not difficult to verify that this Markov chain is ergodic, since elements of the two-step transition probabilities matrix are all positive (condition (120)). Using (121) we obtain

$$2\, p_1 = p_2 + p_3,$$

$$2\, p_2 = p_1 + p_3,$$

and

$$2\, p_3 = p_2 + p_3.$$

Solving the system we obtain

$$p_1 = p_2 = p_3$$

Using (122) we get $p_1 = p_2 = p_3 = 1/3$.

Example 97. Consider Markov chain of Example 85:

$$\P = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{3}{5} & \frac{2}{5} & 0 \end{pmatrix}$$

This Markov chain is ergodic, since all elements of the two-step transition probabilities matrix are positive (see Example 85). Using (121) and (122) we obtain the following system of linear equations:

$$p_1 = 1/2\, p_1 + 2/3\, p_2 + 3/5\, p_3,$$

$$p_2 = 1/4\, p_1 + 2/5\, p_3,$$

$$p_3 = 1/4\, p_1 + 1/3\, p_2,$$

and

$$p_1 + p_2 + p_3 = 1.$$

Solving the system we obtain

$$p_1 = \frac{52}{93} \qquad p_2 = \frac{21}{93} \qquad p_3 = \frac{20}{93}.$$

§37. THE POISSON PROCESS

Before defining a Poisson process, recall that a function f is said to be $o(h)$ if

$$\lim_{h\to 0} \frac{f(h)}{h} = 0.$$

That is, f is $o(h)$ if, for small values of h, $f(h)$ is small even in relation to h. Suppose now that "events" are occuring at random time points and let $N(t)$ denote the number of events that occur in the time interval $[0, t]$. The collection of random variables $\{N(t),\, t \geq 0\}$ is said to be a Poisson process having rate λ, $\lambda > 0$ if

(i) $N(0) = 0$.

(ii) The number of events that occur in disjoint time intervals are independent.

(iii) The distribution of the number of events that occur in a given interval depends only on the length of that interval and not on its location.

(iv) $P(N(h) = 1) = \lambda\, h + o(h)$.

(v) $P(N(h) \geq 2) = o(h)$.

Thus condition (i) states that the process begins at time 0. Condition (ii), the independent increment assumption, states, for instance, that the number of events by time t (that is, $N(t)$) is independent of the number of events that occur between t and $t+s$ (that is, $N(t+s) - N(t)$). Condition (iii), the stationary increment assumption, states that the probability distribution of $N(t+s) - N(t)$ is the same for all values of t.

We have represented an argument, based on the Poisson distribution being a limiting version of the binomial distribution, that the foregoing conditions imply that $N(t)$ has a Poisson distribution with mean $\lambda\, t$. We will now obtain this result by a different method.

Lemma 9. For a Poisson process with rate λ

$$P(N(t) = 0) = e^{-\lambda\, t}.$$

PROOF: Let $P_0(t) = P(N(t) = 0)$. We derive a differential equation for $P_0(t)$ in the following manner:

$$P_0(t+h) = P(N(t+h) = 0) = P(N(t) = 0 \cap \; N(t+h) - N(t) = 0) =$$

$$= P(N(t) = 0)\, P(N(t+h) - N(t) = 0) = P_0(t)[1 - \lambda\, h + o(h)],$$

where the final two equations follow from condition (ii) plus the fact that conditions (iv) and (v) imply that $P(N(h) = 0) = 1 - \lambda\, h + o(h)$. Hence

$$\frac{P_0(t+h) - P_0(t)}{h} = \lambda\, P_0(t) + \frac{o(h)}{h}.$$

Now, letting $h \to 0$, we obtain

$$P_0'(t) = -\lambda P_0(t)$$

or, equivalently,

$$\frac{P_0'(t)}{P_0(t)} = -\lambda$$

which implies, by integration, that

$$\log P_0(t) = -\lambda\, t + c$$

or

$$P_0(t) = K\, e^{-\lambda\, t}.$$

Since $P_0(0) = P(N(0) = 0) = 1$, we arrive at

$$P_0(t) = e^{-\lambda\, t}.$$

For a Poisson process, let us denote by T_1 the time of the first event. Further, for $n > 1$, let T_n denote the elapsed time between the $(n-1)$st and the nth event. The sequence $\{T_n,\, n = 1, 2, \ldots\}$ is called the sequence of interarrival times. For instance, if $T_1 = 5$ and $T_2 = 10$, then the first event of the Poisson process would have occurred at time 5 and the second at time 15.

We shall now determine the distribution of the T_n. To do so, we first note that the event $\{T_1 > t\}$ takes place if and only if no events of the Poisson process occur in the interval $[0, t]$, and thus

$$P(T_1 > t) = P(N(t) = 0) = e^{-\lambda\, t}.$$

Hence T_1 has an exponential distribution with mean $1/\lambda$. Now

$$P(T_2 > t) = E[P(T_2 > t/T_1)].$$

However,

$$P(T_2 > t/T_1 = s) = P(0 \text{ events in } (s, s+t)/T_1 = s) = P(0 \text{ events in } (s, s+t)) = e^{-\lambda\, t},$$

where the last two equations followed from the assumptions about independent and stationary increments. Therefore, from the preceding we conclude that T_2 is also an exponential random variable with mean $1/\lambda$, and furthermore, that T_2 is independent of T_1. Repeating the same argument yields Theorem 31.

Theorem 31. T_1, $T_2, \ldots$ are independent exponential random variables each with mean $1/\lambda$.

Another quantity of interest is S_n, the arrival time of the nth event, also called the waiting time until the nth event. It is easily seen that

$$S_n = \sum_{i=1}^{n} T_i, \quad n \geq 1$$

and hence from Theorem 31 and the result that the sum of n independent exponentially distributed random variables has gamma distribution, it follows that S_n has a gamma distribution with parameters n and λ. That is, the probability density of S_n is given by

$$f_{S_n}(x) = \begin{cases} 0 & \text{if} \quad x \leq 0 \\ \lambda\, e^{-\lambda\, x}\, \frac{(\lambda\, x)^{n-1}}{(n-1)!}, & \text{if} \quad x \geq 0. \end{cases}$$

We are now ready to prove that $N(t)$ is a Poisson random variable with mean $\lambda\, t$.

Theorem 32. For a Poisson process with rate λ,

$$P(N(t) = n) = \frac{e^{-\lambda\, t}(\lambda\, t)^n}{n!}.$$

PROOF: Note that the nth event of the Poisson process will occur before or at time t if and only if the number of events that occur by t is at least n. That is,

$$N(t) \geq n \quad \text{if and only if} \quad S_n \geq t$$

so

$$P(N(t) = n) = P(N(t) \geq n) - P(N(t) \geq n+1) = P(S_n \leq t) - P(S_{n+1} \leq t) =$$
$$= \int_0^t \lambda\, e^{-\lambda\, x}\, \frac{(\lambda\, x)^{n-1}}{(n-1)!}\, dx - \int_0^t \lambda\, e^{-\lambda\, x}\, \frac{(\lambda\, x)^n}{n!}\, dx.$$

But the integration by parts formula $\int u\, dv = u\, v - \int v\, du$ yields, with $u = e^{-\lambda\, x}$, $dv = \lambda[(\lambda\, x)^{n-1}/(n-1)!]\, dx$,

$$\int_0^t \lambda\, e^{-\lambda\, x}\, \frac{(\lambda\, x)^{n-1}}{(n-1)!}\, dx = e^{-\lambda\, t}\, \frac{(\lambda\, t)^n}{n!} + \int_0^t \lambda\, e^{-\lambda\, x}\, \frac{(\lambda\, x)^n}{n!}\, dx$$

which completes the proof.

L E C T U R E 28

§38. SECOND ORDER PROCESSES

A stochastic process $\eta(\omega,t)$, $t\in T$ is called a second order process, if

$$E[\eta(\omega,t)]^2<\infty,$$

for each $t\in T$.

Let $\eta(\omega,t)$, $t\in T$, be a second order process. The mean function $\mu(t)$, $t\in T$, of the process is defined by

$$\mu(t)=E\eta(\omega,t).$$

The covariance function $r_\eta(s,t)$, $s\in T$ and $t\in T$, is defined by

$$r_\eta(s,t)=\text{Cov}\,(\eta_1,\eta_2)(\eta(\omega,s),\eta(\omega,t))=E[\eta(\omega,s)\cdot\eta(\omega,t)]-E[\eta(\omega,s)]\cdot E[\eta(\omega,t)].$$

Since

$$\text{Var}\,(\eta)[\eta(\omega,t)]=\text{Cov}\,(\eta_1,\eta_2)(\eta(\omega,t),\eta(\omega,t)),$$

the variance of $\eta(\omega,t)$ can be expressed in terms of the covariance function as

$$\text{Var}\,(\eta)[\eta(\omega,t)]=r_\eta(t,t),\qquad t\in T. \tag{124}$$

By a finite linear combination of the random variables $\eta(\omega,t)$, $t\in T$, we mean a random variable of the form

$$\sum_{j=1}^n b_j\,\eta(\omega,t_j),$$

where n is natural, t_1, ..., t_n are points in T, and b_1,..., b_n are real constants. The covariance between two such finite linear combinations is given by

$$\text{Cov}\,(\eta_1,\eta_2)\left(\sum_{i=1}^m a_i\,\eta(\omega,s_j),\sum_{j=1}^n b_j\,\eta(\omega,t_j)\right)=\sum_{i=1}^m\sum_{j=1}^n a_i\,b_j\,r(s_i,t_j).$$

In particular

$$\text{Var}\,(\eta)\left(\sum_{j=1}^n b_j\,\eta(\omega,t_j)\right)=\sum_{i=1}^n\sum_{j=1}^n b_i\,b_j\,r(t_i,t_j). \tag{125}$$

It follows immediately from the definition of the covariance function that it is symmetric in s and t, that is

$$r_\eta(s,t) = r_\eta(t,s), \qquad s,t \in T. \tag{126}$$

It is also nonnegative definite. That is, if n is natural, t_1,..., t_n are in T, and b_1,..., b_n are real numbers, then

$$\sum_{i=1}^{n}\sum_{j=1}^{n} b_i\, b_j\, r_\eta(t_i,t_j) \geq 0.$$

This is an immediate consequence of (125).

We say that $\eta(\omega,t)$, $-\infty < t < \infty$, is a second order stationary process, if for every number τ the second order process $\xi(\omega,t)$, $-\infty < t < \infty$, defined by

$$\xi(\omega,t) = \eta(\omega,t+\tau), \qquad -\infty < t < \infty,$$

has the same mean and covariance functions as $\eta(\omega,t)$ process. It is not difficult to show that this is the case if and only if $\mu(t)$ is independent of t and $r_\eta(s,t)$ depends only on the difference between s and t.

Let $\eta(\omega,t)$, $-\infty < t < \infty$, be a second order stationary process. Then

$$\mu(t) = \mu, \qquad -\infty < t < \infty,$$

where μ denotes the common mean of the random variables $\eta(\omega,t)$, $-\infty < t < \infty$.
Since $r_\eta(s,t)$ depends only on the difference between s and t, we have

$$r_\eta(s,t) = r_\eta(0,t-s), \qquad -\infty < s,t < \infty.$$

The function $r_\eta(t)$, $-\infty < s,t < \infty$, defined by

$$r_\eta(t) = r_\eta(0,t), \qquad -\infty < s,t < \infty, \tag{127}$$

is also called the covariance function of the stationary process. We see from (126) and (127) that

$$r_\eta(s,t) = r_\eta(t-s), \qquad -\infty < s,t < \infty.$$

It follows from (125) that $r_\eta(t)$ is symmetric about the origin, that is

$$r_\eta(-t) = r_\eta(t), \qquad -\infty < s, t < \infty.$$

The random variables $\eta(\omega, t)$, $-\infty < t < \infty$, have a common variance given by

$$\mathrm{Var}\,(\eta)\eta(\omega, t) = r_\eta(0), \qquad -\infty < t < \infty.$$

Recall Schwarz's inequality, which asserts that if η_1 and η_2 are random variables having finite second moment, then

$$[E\,(\eta_1 \cdot \eta_2)]^2 \le E\eta_1^2 \cdot E\eta_2^2.$$

Applying Schwarz's inequality to the random variables $\eta_1 - E\eta_1$ and $\eta_2 - E\eta_2$, we see that

$$\left[\mathrm{Cov}\,(\eta_1, \eta_2)\,(\eta_1, \eta_2)\right]^2 \le \mathrm{Var}\,(\eta)(\eta_1) \cdot \mathrm{Var}\,(\eta)(\eta_2).$$

It follows from this last inequality that

$$|\mathrm{Cov}\,(\eta_1, \eta_2)\,(\eta(\omega, 0), \eta(\omega, t))\,| \le \sqrt{\mathrm{Var}\,(\eta)(\eta(\omega, 0) \cdot \mathrm{Var}\,(\eta)(\eta(\omega, t))},$$

and hence that

$$|r_\eta(t)| \le r_\eta(0), \qquad -\infty < t < \infty.$$

If $r_\eta(0) > 0$, the correlation between $\eta(\omega, s)$ and $\eta(\omega, s + t)$ is given independently of s by

$$\frac{\mathrm{Cov}\,(\eta_1, \eta_2)\,(\eta(\omega, s), \eta(\omega, s + t))}{\sqrt{\mathrm{Var}\,(\eta)\eta(\omega, s)} \cdot \sqrt{\mathrm{Var}\,(\eta)\eta(\omega, t)}} = \frac{r_\eta(t)}{r_\eta(0)}, \quad -\infty < t < \infty.$$

Example 98. Let η_1 and η_2 be independent normally distributed random variables each having mean 0 and variance σ^2. Let λ be a real constant and set

$$\eta(\omega, t) = \eta_1 \cos(\lambda t) + \eta_2 \sin(\lambda t), \quad -\infty < t < \infty.$$

Find the mean and covariance functions of $\eta(\omega, t)$, $-\infty < t < \infty$, and show that it is a second order stationary process.

Proof: We observe first that

$$\mu(t) = E\eta_1 \cos(\lambda t) + E\eta_2 \sin(\lambda t) = 0, \quad -\infty < t < \infty.$$

Next,

$$r_\eta(s,t) = \text{Cov}\,(\eta_1,\eta_2)\,(\eta(\omega,s),\eta(\omega,t)) = E[\eta(\omega,s)\cdot\eta(\omega,t)] - E[\eta(\omega,s)]\cdot E[\eta(\omega,t)] =$$

$$= E[\eta(\omega,s)\cdot\eta(\omega,t)] = E\left[(\eta_1\cos(\lambda s) + \eta_2 \sin(\lambda s))\cdot(\eta_1\cos(\lambda t) + \eta_2 \sin(\lambda t))\right] =$$

$$= E(\eta_1^2)\cos(\lambda\, s)\cos(\lambda t) + E(\eta_2^2)\sin(\lambda\, s)\sin(\lambda t) =$$

$$= \sigma^2\cdot(\cos(\lambda\, s)\cos(\lambda t) + \sin(\lambda\, s)\sin(\lambda t)) = \sigma^2\cos\lambda(t-s).$$

This shows that $\eta(\omega,t)$, $-\infty < t < \infty$, is a second order stationary process having mean zero and covariance function

$$r_\eta(t) = \sigma^2\, \cos(\lambda t), \quad -\infty < t < \infty.$$

Example 99. Consider a Poisson process $\eta(\omega,t)$, $-\infty < t < \infty$, with parameter λ (see Section 37). This process satisfies the following properties:

(i) $\eta(\omega,0) = 0$.

(ii) $\eta(\omega,t) - \eta(\omega,s)$ has a Poisson distribution with mean $\lambda(t-s)$ for $s \le t$.

(iii) $\eta(\omega,t_2) - \eta(\omega,t_1)$, $\eta(\omega,t_3) - \eta(\omega,t_2)$,..., $\eta(\omega,t_n) - \eta(\omega,t_{n-1})$ are independent for any $t_1 \le t_2 \le ... \le t_n$.

We now find the mean and covariance function of a process $\eta(\omega,t)$, satisfying (i) – (iii). It follows from properties (i) and (ii) that $\eta(\omega,t)$ has a Poisson distribution with mean λt for $t \ge 0$ and $-\eta(\omega,t)$ has a Poisson distribution with mean $\lambda(-t)$ for $t < 0$. Thus

$$\mu(t) = \lambda t, \quad -\infty < t < \infty.$$

Since the variance of a Poisson distribution equals its mean, we see that $\eta(\omega,t)$ has a finite second moment and that

$$\text{Var}\,(\eta)\eta(\omega,t) = \lambda|t|.$$

Let $0 \leq s \leq t$. Then

$$\mathrm{Cov}\,(\eta_1,\eta_2)(\eta(\omega,s),\eta(\omega,s)) = \mathrm{Var}\,(\eta)\eta(\omega,s) = \lambda s.$$

It follows from properties (i) and (iii) that $\eta(\omega,s)$ and $\eta(\omega,t)-\eta(\omega,s)$ are independent, and hence

$$\mathrm{Cov}\,(\eta_1,\eta_2)(\eta(\omega,s),\eta(\omega,t)-\eta(\omega,s)) = 0.$$

Thus

$$\mathrm{Cov}\,(\eta_1,\eta_2)(\eta(\omega,s),\eta(\omega,t)) = \mathrm{Cov}\,(\eta_1,\eta_2)(\eta(\omega,s),\eta(\omega,s)+\eta(\omega,t)-\eta(\omega,s)) =$$

$$= \mathrm{Cov}\,(\eta_1,\eta_2)(\eta(\omega,s),\eta(\omega,s)) + \mathrm{Cov}\,(\eta_1,\eta_2)(\eta(\omega,s),\eta(\omega,t)-\eta(\omega,s)) = \lambda s.$$

If $s < 0$ and $t > 0$, then by properties (i) and (iii) the random variables $(\eta(\omega,s)$ and $\eta(\omega,t)$ are independent, and hence

$$\mathrm{Cov}\,(\eta_1,\eta_2)(\eta(\omega,s),\eta(\omega,t)) = 0.$$

The other cases can be handled similarly. We find in general that

$$r_\eta(s,t) = \begin{cases} \lambda\,\min(|s|,|t|), & \text{if} \quad s\,t \geq 0, \\ 0, & \text{if} \quad s\,t < 0, \end{cases}$$

Therefore, the process is second order, but not stationary.

In the next example we consider a closely related process which is a second order stationary process.

Example 100. Let $\eta(\omega,t)$, $-\infty < t < \infty$, be A Poisson process with parameter λ (see Section 37). Set

$$\xi(\omega,t) = \eta(\omega,t+1) - \eta(\omega,t), \qquad -\infty < t < \infty.$$

Find the mean and covariance function of the $\xi(\omega,t)$ process, and show that it is a second order stationary process.

Solution: Since $E\eta(\omega,t)=\lambda t$, it follows that

$$E\xi(\omega,t)=E\eta(\omega,t+1)-E\eta(\omega,t)=\lambda(t+1)-\lambda t=\lambda,$$

so the random variables $\xi(\omega,t)$ have common mean λ. To compute the covariance function of the $\xi(\omega,t)$ process, we observe that if $|t-s|\geq 1$, then the random variables $\eta(\omega,s+1)-\eta(\omega,s)$ and $\eta(\omega,t+1)-\eta(\omega,t)$ are independent by property (iii). Consequently,

$$r_\xi(s,t)=0, \quad \text{for} \quad |t-s|\geq 1.$$

Suppose $s\leq t<s+1$. Then

$$\text{Cov}\,(\eta_1,\eta_2)(\xi(\omega,s),\xi(\omega,t))=\text{Cov}\,(\eta_1,\eta_2)[(\eta(\omega,s+1)-\eta(\omega,s)),(\eta(\omega,t+1)-\eta(\omega,t))]=$$

$$=\text{Cov}\,(\eta_1,\eta_2)[(\eta(\omega,t)-\eta(\omega,s)+\eta(\omega,s+1)-\eta(\omega,t)),(\eta(\omega,s+1)-\eta(\omega,t)+\eta(\omega,t+1)-\eta(\omega,s+1))].$$

It follows from property (iii) and the assumptions on s and t that

$$\text{Cov}\,(\eta_1,\eta_2)[(\eta(\omega,t)-\eta(\omega,s),(\eta(\omega,s+1)-\eta(\omega,t))]=0,$$

$$\text{Cov}\,(\eta_1,\eta_2)[(\eta(\omega,t)-\eta(\omega,s),(\eta(\omega,t+1)-\eta(\omega,s+1))]=0,$$

and

$$\text{Cov}\,(\eta_1,\eta_2)[(\eta(\omega,s+1)-\eta(\omega,t),(\eta(\omega,t+1)-\eta(\omega,s+1))]=0,$$

By property (ii)

$$\text{Cov}\,(\eta_1,\eta_2)[(\eta(\omega,s+1)-\eta(\omega,t),(\eta(\omega,s+1)-\eta(\omega,t))]=\text{Var}\,(\eta)(\eta(\omega,s+1)-\eta(\omega,t))=\lambda(s+1-t).$$

Thus

$$\text{Cov}\,(\eta_1,\eta_2)(\xi(\omega,s),\xi(\omega,t))=\lambda(s+1-t).$$

By using symmetry we find in general that

$$r_\xi(s,t)=\begin{cases}\lambda\,(1-|t-s|), & \text{if} \quad |t-s|<1,\\ 0, & \text{if} \quad |t-s|\geq 1.\end{cases}$$

Thus $\xi(\omega,t)$, $-\infty<t<\infty$, is a second order stationary process having mean λ and covariance function

$$r_\xi(t)=\begin{cases}\lambda\,(1-|t|), & \text{if} \quad |t|<1,\\ 0, & \text{if} \quad |t|\geq 1.\end{cases}$$

LECTURE 29

§39. GAUSSIAN PROCESSES.

A stochastic process $\eta(\omega,t)$, $t\in T$, is called a Gaussian process if every finite linear combination of the random variables $\eta(\omega,t)$, is normally distributed. In this context constant random variables are regarded as normally distributed with zero variance. Gaussian processes are also called normal processes. If $\eta(\omega,t)$, $t\in T$, is a Gaussian process, then for each $t\in T$, $\eta(\omega,t)$ is normally distributed, and, in particular, $E\eta(\omega,t)^2<\infty$. Thus a Gaussian process is necessary a second order process. Gaussian processes have many nice theoretical properties that do not hold for second order processes in general. They are also widely used in applications, especially in engineering and financial mathematics. It is not difficult to prove that the process from Example 96 is a Gaussian process.

One of the most useful properties of Gaussian processes is that if two such processes have the same mean and covariance functions, then they also have the same joint distribution functions. We omit the proof of this result.

The mean and covariance functions can also be used to find the higher moments of a Gaussian process.

Example 99. Let $\eta(\omega,t), t\in T$, be a Gaussian process having zero means. Find $E\eta^4(\omega,t)$ in terms of the covariance function of the process.

We recall that if η is normally distributed with mean 0 and variance σ^2, then $E\eta^4(\omega,t)=3\sigma^4$. Since $\eta(\omega,t)$ is normally distributed with mean 0 and variance $r_\eta(t,t)$, we see that

$$E\eta^4(\omega,t)=3(r_\eta(t,t))^2.$$

Let n be a positive integer and let $\eta_1,\dots,\eta_n$ be random variables. They are said to have a joint normal (or Gaussian) distribution, if

$$a_1\eta_1+\cdots+a_n\eta_n$$

is normally distributed for every choice of the constants $a_,\dots,a_n$. A stochastic process $\eta(\omega,t)$, $t\in T$, is a Gaussian process if and only if for every positive integer n and every choice of $t_1,\dots,t_n$ all in T, the random variables $\eta(\omega,t_1),\dots,\eta(\omega,t_n)$ have a joint normal distribution.

§40. THE WIENER PROCESS

It has long been known from microscopic observations that particles suspended in a liquid are in a state of constant highly irregular motion. It gradually came to be realized that the cause of this motion is the bombardment of the particles by the smaller invisible molecules of the liquid. Such motion is called "Brownian motion," named after one of the first scientists to study it carefully.

Many mathematical models for this physical process have been proposed. We now describe by a Cartesian coordinate system whose origin is the location of the particle at time $t = 0$. Then the three coordinates of the position of the particle vary independently, each according to a stochastic process $W(\omega, t)$, $-\infty < t < \infty$, satisfying the following properties:

(i) $W(\omega, 0) = 0$.

(ii) $W(\omega, t) - W(\omega, s)$ has a normal distribution with mean 0 and variance $\sigma^2 (t - s)$ for $s \leq t$.

(iii) $W(\omega, t_2) - W(\omega, t_1)$, $W(\omega, t_3) - W(\omega, t_2)$,..., $W(\omega, t_n) - W(\omega, t_{n-1})$ are independent for $t_1 \leq t_2 \leq ... \leq t_n$.

Here σ^2 is some positive constant.

Property (i) follows from our choice of the coordinate system. Properties (ii) and (iii) are plausible if the motion is caused by an extremely large number of unrelated and individually negligible collisions which have no more tendency to move the particle in one direction than in the opposite direction. In particular, the central limit theorem makes it reasonable to suppose that the increments $W(\omega, t) - W(\omega, s)$ are normally distributed.

A stochastic process $W(\omega, t)$, $-\infty < t < \infty$, satisfying properties (i) – (iii) is called the Wiener process with parameter σ^2. It follows immediately from the properties of the Wiener process that the random variables $W(\omega, t)$ all have mean 0 and that

$$E[(W(\omega, t_2) - W(\omega, t_1)) \cdot (W(\omega, t_4) - W(\omega, t_3))] = 0, \quad t_1 \leq t_2 \leq t_3 \leq t_4.$$

The covariance function of the process is

$$r_\eta(s, t) = \begin{cases} \sigma^2 \min(|s|, |t|), & \text{if} \quad s\,t \geq 0, \\ 0, & \text{if} \quad s\,t < 0, \end{cases} \tag{128}$$

The proof of (128) is identical to that for the covariance function of the Poisson process. It is not difficult to prove that

$$E[(W(\omega, s) - W(\omega, a)) \cdot (W(\omega, t) - W(\omega, a))] = \sigma^2 \min(s - a, t - a), \quad s \geq a \quad \text{and} \quad t \geq a.$$

§41. GENERALIZED BLACK–SCHOLES FORMULA, A SEASONALITY FACTOR

The Black–Scholes model or Black–Scholes–Merton is a mathematical model of a financial market containing certain derivative investment instruments. From the model, one can deduce the Black–Scholes formula, which gives the price of European options. The formula led to a boom in options trading and legitimized scientifically the activities of the Chicago Board Options Exchange and other options markets around the world. lt is widely used by options market participants. Many empirical tests have shown the Black–Scholes price is fairly close to the observed prices, although there are well-known discrepancies such as the option smile.

The model was first articulated by Fischer Black and Myron Scholes in their 1973 paper, The Pricing of Options and Corporate Liabilities", published in the Journal of Political Economy. They derived a partial differential equation, now called the Black–Scholes equation, which governs the price of the option over time. The key idea behind the derivation was to hedge perfectly the option by buying and selling the underlying asset in just the right way and consequently "eliminate risk". This hedge is called delta hedging and is the basis of more complicated hedging strategies such as those engaged in by Wall Street investment banks. The hedge implies there is only one right price for the option and it is given by the Black–Scholes formula.

Robert C. Merton was the first to publish a paper expanding the mathematical understanding of the options pricing model and coined the term Black–Scholes options pricing model. Merton and Scholes received the 1997 Nobel Prize in Economics (The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel) for their work. Though ineligible for the prize because of his death in 1995, Black was mentioned as a contributor by the Swedish academy.

Let S be the price of the stock,

$V(S,t)$ be the price of a derivative as a function of time and stock price,

$C(S,t)$ be the price of a European call option and $P(S,t)$ the price of a European put option,

K be the strike of the option,

r be the annualized risk-free interest rate, continuously compounded.

μ be the drift rate of S, annualized,

σ be the volatility of the stock's returns; this is the square root of the quadratic variation of the stock's log price process.

t be a time in years; we generally use: now=0, expiry=T.

Π be the value of a portfolio.

The Black–Scholes equation simulated geometric Brownian motions with parameters from market data, the Black–Scholes equation is a partial differential equation, which describes the price of the option over time. The key idea behind the equation is that one can perfectly hedge the option by buying and selling the underlying asset in just the right way and consequently eliminate risk". This hedge, in turn, implies that there is only one right price for the option, as returned by the Black–Scholes formula given above. The Equation:

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + r\,S\frac{\partial V}{\partial S} - r\,V = 0.$$

Per the model assumptions above, the price of the underlying asset (typically a stock) follows a geometric Brownian motion. That is,

$$\frac{dS}{S} = \mu\,dt + \sigma\,dW,$$

where W is Brownian motion. Note that W, and consequently its infinitesimal increment dW, represents the only source of uncertainty in the price history of the stock. Intuitively, $W(t)$ is a process that "wiggles up and down" in such a random way that its expected change over any time interval is 0. (In addition, its variance over time T is equal to T; a good discrete analogue for W is a simple random walk. Thus the above equation states that the infinitesimal rate of return on the stock has an expected value of $\mu\,dt$ and a variance of $\sigma^2 dt$.

The Black–Scholes formula calculates the price of European put and call options. This price is consistent with the Black–Scholes equation; this follows since the formula can be obtained by solving the equation for the corresponding terminal and boundary conditions.

Classical Black–Scholes analysis consider the following expectation

$$\mathbf{E}[(e^Z - K)^+],$$

where $\mathbf{E}$ stands for expectation, $K > 0$.

We consider a generalized form

$$\mathbf{E}[(a\,e^{Z} - K)^{+}],$$

where the parameter a can be interpreted as seasonality factor

Theorem 32. Let Z have normal $\mathcal{N}(\gamma, \sigma^2)$ distribution. Then for any $a > 0$ and $K > 0$ we have

$$\mathbf{E}[(a\,e^{Z} - K)^{+}] = a\,e^{\gamma + \frac{1}{2}\sigma^2}\,\Phi\left(A + \sigma\right) - K\,\Phi(A), \tag{129}$$

where Φ is the standard normal $\mathcal{N}(0,1)$ distribution function, while

$$A \equiv A(a, K, \gamma, \sigma) = \frac{\gamma + \ln a - \ln K}{\sigma}$$

Proof: By definition we obtain

$$\mathbf{E}[(a\,e^{Z} - K)^{+}] = \frac{1}{\sigma\sqrt{2\,\pi}} \int_{\{z : a\,e^{z} > K\}} (a\,e^{z} - K)\, \exp\left(-\frac{(z-\gamma)^2}{2\,\sigma^2}\right) dz =$$

$$= \frac{1}{\sigma\sqrt{2\,\pi}} \left[a \int_{\ln(K/a)}^{\infty} e^{z}\, \exp\left(-\frac{(z-\gamma)^2}{2\,\sigma^2}\right) dz - K \int_{\ln(K/a)}^{\infty} \exp\left(-\frac{(z-\gamma)^2}{2\,\sigma^2}\right) dz \right]$$

Since

$$\frac{1}{\sigma\sqrt{2\,\pi}} \int_{\ln(K/a)}^{\infty} \exp\left(-\frac{(z-\gamma)^2}{2\,\sigma^2}\right) dz = 1 - \mathbf{P}\left(Z < \ln(K/a)\right) = 1 - \Phi\left(\frac{\ln(K/a) - \gamma}{\sigma}\right)$$

and using the property of standard normal distribution function, that for any real x:

$$1 - \Phi(x) = \Phi(-x)$$

we obtain

$$-K\, \frac{1}{\sigma\sqrt{2\,\pi}} \int_{\ln(K/a)}^{\infty} \exp\left(-\frac{(z-\gamma)^2}{2\,\sigma^2}\right) dz = -K\,\Phi\left(\frac{\gamma + \ln a - \ln K}{\sigma}\right).$$

Now we pass on the term

$$\frac{a}{\sigma\sqrt{2\,\pi}} \int_{\ln(K/a)}^{\infty} e^{z}\, \exp\left(-\frac{(z-\gamma)^2}{2\,\sigma^2}\right) dz$$

It is not difficult to verify an elementary identity

$$e^{z}\, \exp\left(-\frac{(z-\gamma)^2}{2\,\sigma^2}\right) = \exp\left(-\frac{(z-(\gamma+\sigma^2))^2}{2\,\sigma^2}\right)\, \exp\{\gamma + \frac{1}{2}\,\sigma^2\},$$

we obtain

$$\frac{a}{\sigma\sqrt{2\pi}}\int_{\ln(K/a)}^{\infty} e^{z}\,\exp\left(-\frac{(z-\gamma)^2}{2\sigma^2}\right)\,dz = a\,e^{\gamma+\frac{1}{2}\sigma^2}\,\mathbf{P}(Z_1 > \ln(K/a)),$$

where random variable Z_1 has normal $\mathcal{N}(\gamma+\sigma^2,\sigma^2)$ distribution.

Therefore,

$$\frac{a}{\sigma\sqrt{2\pi}}\int_{\ln(K/a)}^{\infty} e^{z}\,\exp\left(-\frac{(z-\gamma)^2}{2\sigma^2}\right)\,dz =$$

$$= a\,e^{\gamma+\frac{1}{2}\sigma^2}\left(1-\Phi\left(\frac{\ln(K/a)-\gamma-\sigma^2}{\sigma}\right)\right) = a\,e^{\gamma+\frac{1}{2}\sigma^2}\,\Phi\left(\frac{\ln(a/K)+\gamma}{\sigma}+\sigma\right) =$$

$$= a\,e^{\gamma+\frac{1}{2}\sigma^2}\,\Phi\left(\frac{\gamma+\ln a-\ln K}{\sigma}+\sigma\right)$$

The proof is complete.

Corollary. Let Z have normal $\mathcal{N}(\gamma,\sigma^2)$ distribution. Then for any $K>0$ we have

$$\mathbf{E}[(e^{Z}-K)^{+}] = e^{\gamma+\frac{1}{2}\sigma^2}\,\Phi\left(\frac{\gamma-\ln K}{\sigma}+\sigma\right) - K\,\Phi\left(\frac{\gamma-\ln K}{\sigma}\right), \tag{130}$$

where Φ is the standard normal $\mathcal{N}(0,1)$ distribution function.

Proof follows from the previous formula if we substitute $a=1$.

L E C T U R E 30

TYPICAL PROBLEMS OF FINAL EXAMINATION

PROBLEM 1. A coin is tossed 10 times and all possible outcomes of the 10 tosses are assumed equally likely. What is the probability that a random realization either begins with three tails or ends with two heads?

PROBLEM 2. The diameter of the disk is measured approximately. Considering that its value is uniformly distributed over interval $(0, 2)$, find the distribution of area of the disk.

PROBLEM 3. One thousand independent rolls of a fair die will be made. Compute an approximation to the probability that number 6 will appear between 170 and 250.

PROBLEM 4. Let η be uniformly distributed random variable over the interval $[a, b]$. Let $E\eta^2 = 1$ and $E\eta = -E\eta^3$. Find a and b.

PROBLEM 5. If 10 married couples are randomly seated at a round table, compute the expected value and the variance of the number of wives who are seated next to their husbands.

PROBLEM 6. Let η be uniformly distributed over the interval $[0, \frac{\pi}{2}]$. Find the density function of $\xi = \sin\eta$.

ANSWERS

1. $\frac{11}{32} = 0.34275.$

2. For distribution function

$$F_S(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \frac{\sqrt{x}}{\sqrt{\pi}} & \text{if} \quad x \in [0; \pi] \\ 1 & \text{if} \quad x > \pi \end{cases}$$

or for density function

$$f_S(x) = \begin{cases} \frac{1}{2\sqrt{\pi\, x}} & \text{if} \quad x \in [0; \pi] \\ 0 & \text{otherwise} \end{cases}$$

3. $\approx 0.3897.$

4. $a = -\sqrt{3},\ b = \sqrt{3}$

5. $\frac{20}{19}$; $\frac{360}{361}$.

6. For distribution function

$$F_\xi(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \dfrac{2\arcsin x}{\pi} & \text{if} \quad x \in [0;1] \\ 1 & \text{if} \quad x > 1 \end{cases}$$

or for density function

$$f_\xi(x) = \begin{cases} \dfrac{2}{\pi\sqrt{1-x^2}} & \text{if} \quad x \in [0;1] \\ 0 & \text{otherwise} \end{cases}$$