TABLE OF CONTENTS

PROBABILITY THEORY

STOCHASTIC PROCESSES

Note: These lecture notes are revised periodically with new materials and examples added from time to time. Lectures 1 $\rightarrow$ 11 are used at Polytechnic for a first level graduate course on "Probability theory and Random Variables". Parts of lectures 14 $\rightarrow$ 19 are used at Polytechnic for a "Stochastic Processes" course. These notes are intended for unlimited worldwide use. However the user must acknowledge the present website www.mhhe.com/papoulis as the source of information. Any feedback may be addressed to pillai@hora.poly.edu

S. UNNIKRISHNA PILLAI

PROBABILITY THEORY

1. Basics

Probability theory deals with the study of random phenomena, which under repeated experiments yield different outcomes that have certain underlying patterns about them. The notion of an experiment assumes a set of repeatable conditions that allow any number of identical repetitions. When an experiment is performed under these conditions, certain elementary events ξ_i occur in different but *completely uncertain* ways. We can assign nonnegative number $P(\xi_i)$, as the probability of the event ξ_i in various ways:

Laplace's Classical Definition: The Probability of an event A is defined a-priori without actual experimentation as

$$P(A) = \frac{\text{Number of outcomes favorable to } A}{\text{Total number of possible outcomes}}, \qquad (1\text{-}1)$$

provided all these outcomes are *equally likely*.

Consider a box with n white and m red balls. In this case, there are two elementary outcomes: white ball or red ball. Probability of "selecting a white ball" $= \dfrac{n}{n+m}.$
We can use above classical definition to determine the probability that a given number is divisible by a prime p.

If p is a prime number, then every p^{th} number (starting with p) is divisible by p. Thus among p consecutive integers there is one favorable outcome, and hence

$$P\{a \text{ given number is divisible by a prime } p\} = \frac{1}{p} \quad \text{(1-2)}$$

Relative Frequency Definition: The probability of an event A is defined as

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n} \quad \text{(1-3)}$$

where n_A is the number of occurrences of A and n is the total number of trials.
We can use the relative frequency definition to derive (1-2) as well. To do this we argue that among the integers $1, 2, 3, \cdots, n$, the numbers $p, 2p, \cdots$ are divisible by p.

Thus there are n/p such numbers between 1 and n. Hence

$$P\{a \text{ given number } N \text{ is divisible by a prime } p\}$$

$$= \lim_{n\to\infty} \frac{n/p}{n} = \frac{1}{p}. \tag{1-4}$$

In a similar manner, it follows that

$$P\{\, p^2 \text{ divides any given number } N\} = \frac{1}{p^2} \tag{1-5}$$

and

$$P\{\, pq \text{ divides any given number } N\} = \frac{1}{pq}. \tag{1-6}$$

The axiomatic approach to probability, due to Kolmogorov, developed through a set of axioms (below) is generally recognized as superior to the above definitions, (1-1) and (1-3), as it provides a solid foundation for complicated applications.

The totality of all ξ_i, *known a priori*, constitutes a set Ω, the set of all experimental outcomes.

$$\Omega = \{ \xi_1, \xi_2, \cdots, \xi_k, \cdots \} \qquad (1\text{-}7)$$

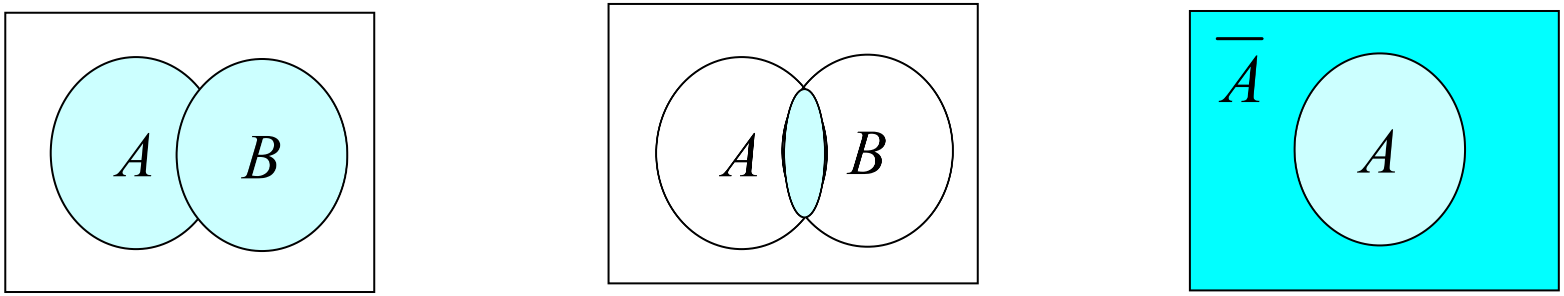
Ω has subsets $A, B, C, \cdots$. Recall that if A is a subset of Ω, then $\xi \in A$ implies $\xi \in \Omega$. From A and B, we can generate other related subsets $A \cup B$, $A \cap B$, $\overline{A}$, $\overline{B}$, etc.

$$A \cup B = \{ \xi \mid \xi \in A \ \text{ or } \ \xi \in B \}$$

$$A \cap B = \{ \xi \mid \xi \in A \ \text{ and } \ \xi \in B \}$$

and

$$\overline{A} = \{ \xi \mid \xi \notin A \} \qquad (1\text{-}8)$$

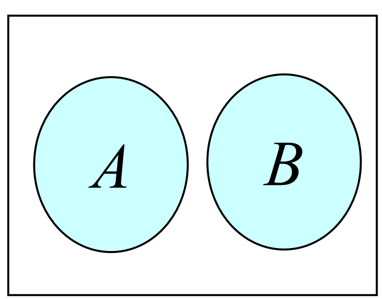


$A \cup B$ $\qquad$ $A \cap B$ $\qquad$ $\overline{A}$

Fig.1.1

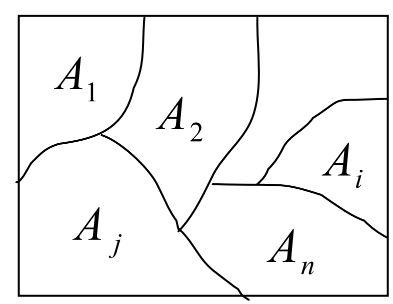
- If $A \cap B = \phi$, the empty set, then A and B are said to be mutually exclusive (M.E).
- A partition of Ω is a collection of mutually exclusive subsets of Ω such that their union is Ω.

$$A_i \cap A_j = \phi, \quad \text{and} \quad \bigcup_{i=1} A_i = \Omega. \tag{1-9}$$



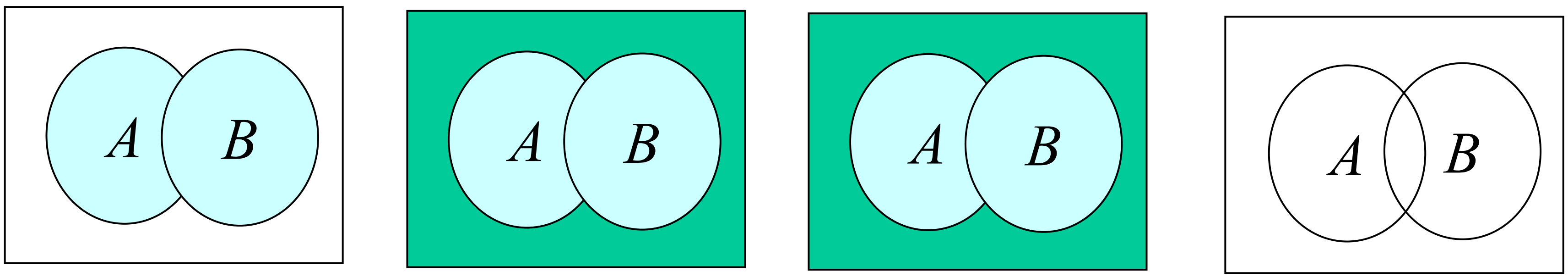
$A \cap B = \phi$

Fig. 1.2



De-Morgan's Laws:

$$\overline{A \cup B} = \overline{A} \cap \overline{B}; \qquad \overline{A \cap B} = \overline{A} \cup \overline{B} \qquad (1\text{-}10)$$



$A \cup B$ $\qquad \overline{A \cup B}$ $\qquad \overline{A} \cap \overline{B}$

Fig.1.3

- Often it is meaningful to talk about at least some of the subsets of Ω as events, for which we must have mechanism to compute their probabilities.

Example 1.1: Consider the experiment where two coins are simultaneously tossed. The various elementary events are

$$\xi_1 = (H,H),\ \ \xi_2 = (H,T),\ \ \xi_3 = (T,H),\ \ \xi_4 = (T,T)$$

and

$$\Omega = \{\xi_1, \xi_2, \xi_3, \xi_4\}.$$

The subset $A = \{\xi_1, \xi_2, \xi_3\}$ is the same as "Head has occurred at least once" and qualifies as an event.

Suppose two subsets A and B are both events, then consider

"Does an outcome belong to A or $B = A \cup B$"

"Does an outcome belong to A and $B = A \cap B$"

"Does an outcome fall outside A"?

Thus the sets $A \cup B$, $A \cap B$, $\overline{A}$, $\overline{B}$, etc., also qualify as events. We shall formalize this using the notion of a Field.

•**Field**: A collection of subsets of a nonempty set Ω forms a field F if

$$\begin{aligned} &\text{(i)} \quad \Omega \in F \\ &\text{(ii)} \quad \text{If } A \in F, \quad \text{then } \overline{A} \in F \\ &\text{(iii)} \quad \text{If } A \in F \quad \text{and } B \in F, \quad \text{then } A \cup B \in F. \end{aligned} \tag{1-11}$$

Using (i) - (iii), it is easy to show that $A \cap B$, $\overline{A} \cap B$, etc., also belong to F. For example, from (ii) we have $\overline{A} \in F$, $\overline{B} \in F$, and using (iii) this gives $\overline{A} \cup \overline{B} \in F$; applying (ii) again we get $\overline{\overline{A} \cup \overline{B}} = A \cap B \in F$, where we have used De Morgan's theorem in (1-10).

Thus if $A \in F$, $B \in F$, then

$$F = \left\{ \Omega, A, B, \overline{A}, \overline{B}, A \cup B, A \cap B, \overline{A} \cup B, \cdots \right\}. \quad (1\text{-}12)$$

From here on wards, we shall reserve the term 'event' only to members of F.

Assuming that the probability $p_i = P(\xi_i)$ of elementary outcomes ξ_i of Ω are apriori defined, how does one assign probabilities to more 'complicated' events such as A, B, AB, etc., above?

The three axioms of probability defined below can be used to achieve that goal.

Axioms of Probability

For any event A, we assign a number $P(A)$, called the probability of the event A. This number satisfies the following three conditions that act the axioms of probability.

$$
\begin{aligned}
&\text{(i)} \quad P(A) \geq 0 \quad \text{(Probability is a nonnegative number)} \\
&\text{(ii)} \quad P(\Omega) = 1 \quad \text{(Probability of the whole set is unity)} \qquad (1\text{-}13) \\
&\text{(iii)} \quad \text{If } A \cap B = \phi, \quad \text{then} \quad P(A \cup B) = P(A) + P(B).
\end{aligned}
$$

(Note that (iii) states that if A and B are mutually exclusive (M.E.) events, the probability of their union is the sum of their probabilities.)

The following conclusions follow from these axioms:

a. Since $A \cup \overline{A} = \Omega$, we have using (ii)

$$\mathrm{P}(A \cup \overline{A}) = P(\Omega) = 1.$$

But $A \cap \overline{A} \in \phi$, and using (iii),

$$\mathrm{P}(A \cup \overline{A}) = P(A) + \mathrm{P}(\overline{A}) = 1 \quad \text{or} \quad \mathrm{P}(\overline{A}) = 1 - P(A). \qquad \text{(1-14)}$$

b. Similarly, for any *A*, $A \cap \{\phi\} = \{\phi\}$.

Hence it follows that $P\left(A \cup \{\phi\}\right) = P(A) + P(\phi)$.

But $A \cup \{\phi\} = A$, and thus $P\{\phi\} = 0$. (1-15)

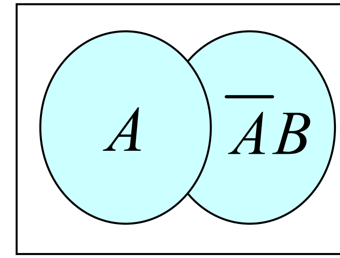
c. Suppose *A* and *B* are *not* mutually exclusive (M.E.)?

How does one compute $P(A \cup B) = ?$

To compute the above probability, we should re-express $A \cup B$ in terms of M.E. sets so that we can make use of the probability axioms. From Fig.1.4 we have

$$A \cup B = A \cup \overline{A}B, \qquad (1\text{-}16)$$

where A and $\overline{A}B$ are clearly M.E. events. Thus using axiom (1-13-iii)



$A \cup B$

Fig.1.4

$$P(A \cup B) = P(A \cup \overline{A}B) = P(A) + P(\overline{A}B). \qquad (1\text{-}17)$$

To compute $P(\overline{A}B)$, we can express B as

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

$$= (B \cap A) \cup (B \cap \overline{A}) = BA \cup B\overline{A} \qquad (1\text{-}18)$$

Thus

$$P(B) = P(BA) + P(B\overline{A}), \qquad (1\text{-}19)$$

since $BA = AB$ and $B\overline{A} = \overline{A}B$ are M.E. events.

From (1-19),

$$P(\overline{A}B) = P(B) - P(AB) \tag{1-20}$$

and using (1-20) in (1-17)

$$P(A \cup B) = P(A) + P(B) - P(AB). \tag{1-21}$$

- Question: Suppose every member of a denumerably infinite collection A_i of pair wise disjoint sets is an event, then what can we say about their union

$$A = \bigcup_{i=1}^{\infty} A_i \ ? \tag{1-22}$$

i.e., suppose all $A_i \in F$, what about A? Does it belong to F? (1-23)

Further, if A also belongs to F, what about $P(A)$? (1-24)

The above questions involving infinite sets can only be settled using our intuitive experience from plausible experiments. For example, in a coin tossing experiment, where the same coin is tossed indefinitely, define

$$A = \text{“head eventually appears”}. \tag{1-25}$$

Is A an event? Our intuitive experience surely tells us that A is an event. Let

$$\begin{aligned} A_n &= \{\text{head appears for the 1st time on the } n\text{th toss}\} \\ &= \{\underbrace{t, t, t, \cdots, t}_{n-1}, h\} \end{aligned} \tag{1-26}$$

Clearly $A_i \cap A_j = \phi$. Moreover the above A is

$$A = A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_i \cup \cdots. \tag{1-27}$$

We cannot use probability axiom (1-13-iii) to compute $P(A)$, since the axiom only deals with two (or a finite number) of M.E. events.

To settle both questions above (1-23)-(1-24), extension of these notions must be done based on our intuition as new axioms.

σ-Field (Definition):

A field F is a σ-field if in addition to the three conditions in (1-11), we have the following:

For every sequence $A_i, i = 1 \rightarrow \infty,$ of pair wise disjoint events belonging to F, their union also belongs to F, i.e.,

$$A = \bigcup_{i=1}^{\infty} A_i \in F. \tag{1-28}$$

In view of (1-28), we can add yet another axiom to the set of probability axioms in (1-13).

(iv) If A_i are pair wise mutually exclusive, then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n). \tag{1-29}$$

Returning back to the coin tossing experiment, from experience we know that if we keep tossing a coin, eventually, a head must show up, i.e.,

$$P(A) = 1. \tag{1-30}$$

But $A = \bigcup_{n=1}^{\infty} A_n,$ and using the fourth probability axiom in (1-29),

$$P(A) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n). \tag{1-31}$$

From (1-26), for a fair coin since only one in 2^n outcomes is in favor of A_n, we have

$$P(A_n) = \frac{1}{2^n} \quad \text{and} \quad \sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} \frac{1}{2^n} = 1, \tag{1-32}$$

which agrees with (1-30), thus justifying the

'reasonableness' of the fourth axiom in (1-29).

In summary, the triplet (Ω, *F*, *P*) composed of a nonempty set Ω of elementary events, a σ-field *F* of subsets of Ω, and a probability measure *P* on the sets in *F* subject the four axioms ((1-13) and (1-29)) form a probability model.

The probability of more complicated events must follow from this framework by deduction.

Conditional Probability and Independence

In N independent trials, suppose N_A, N_B, N_{AB} denote the number of times events A, B and AB occur respectively. According to the frequency interpretation of probability, for large N

$$P(A) \approx \frac{N_A}{N}, \quad P(B) \approx \frac{N_B}{N}, \quad P(AB) \approx \frac{N_{AB}}{N}. \tag{1-33}$$

Among the N_A occurrences of A, only N_{AB} of them are also found among the N_B occurrences of B. Thus the ratio

$$\frac{N_{AB}}{N_B} = \frac{N_{AB}/N}{N_B/N} = \frac{P(AB)}{P(B)} \tag{1-34}$$

is a measure of "the event A given that B has already occurred". We denote this conditional probability by

$P(A|B) =$ Probability of "the event A given that B has occurred".

We define

$$P(A \mid B) = \frac{P(AB)}{P(B)}, \tag{1-35}$$

provided $P(B) \neq 0.$ As we show below, the above definition satisfies all probability axioms discussed earlier.

We have

(i) $$P(A \mid B) = \frac{P(AB) \geq 0}{P(B) > 0} \geq 0, \tag{1-36}$$

(ii) $$P(\Omega \mid B) = \frac{P(\Omega B)}{P(B)} = \frac{P(B)}{P(B)} = 1, \text{ since } \Omega B = B. \tag{1-37}$$

(iii) Suppose $A \cap C = 0$. Then

$$P(A \cup C \mid B) = \frac{P((A \cup C) \cap B)}{P(B)} = \frac{P(AB \cup CB)}{P(B)}. \tag{1-38}$$

But $AB \cap AC = \phi$, hence $P(AB \cup CB) = P(AB) + P(CB)$.

$$P(A \cup C \mid B) = \frac{P(AB)}{P(B)} + \frac{P(CB)}{P(B)} = P(A \mid B) + P(C \mid B), \tag{1-39}$$

satisfying all probability axioms in (1-13). Thus (1-35) defines a legitimate probability measure.

Properties of Conditional Probability:

a. If $B \subset A$, $AB = B$, and

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{P(B)}{P(B)} = 1 \qquad \text{(1-40)}$$

since if $B \subset A$, then occurrence of B implies automatic occurrence of the event A. As an example, but

$$A = \{\text{outcome is even}\}, \; B = \{\text{outcome is 2}\},$$

in a dice tossing experiment. Then $B \subset A$, and $P(A \mid B) = 1.$

b. If $A \subset B$, $AB = A$, and

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{P(A)}{P(B)} > P(A). \qquad \text{(1-41)}$$

(In a dice experiment, $A = \{\text{outcome is 2}\}$, $B=\{\text{outcome is even}\}$, so that $A \subset B$. The statement that B has occurred (outcome is even) makes the odds for "outcome is 2" greater than without that information).

c. We can use the conditional probability to express the probability of a complicated event in terms of "simpler" related events.

Let $A_1, A_2, \cdots, A_n$ are pair wise disjoint and their union is Ω. Thus $A_i A_j = \phi$, and

$$\bigcup_{i=1}^{n} A_i = \Omega . \tag{1-42}$$

Thus

$$B = B(A_1 \cup A_2 \cup \cdots \cup A_n) = BA_1 \cup BA_2 \cup \cdots \cup BA_n. \tag{1-43}$$

But $A_i \cap A_j = \phi \Rightarrow BA_i \cap BA_j = \phi$, so that from (1-43)

$$P(B) = \sum_{i=1}^{n} P(BA_i) = \sum_{i=1}^{n} P(B \mid A_i) P(A_i). \qquad (1\text{-}44)$$

With the notion of conditional probability, next we introduce the notion of "independence" of events.

Independence: *A* and *B* are said to be independent events, if

$$P(AB) = P(A) \cdot P(B). \qquad (1\text{-}45)$$

Notice that the above definition is a probabilistic statement, *not* a set theoretic notion such as mutually exclusiveness.

Suppose A and B are independent, then

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A). \qquad (1\text{-}46)$$

Thus if A and B are independent, the event that B has occurred does not shed any more light into the event A. It makes no difference to A whether B has occurred or not. An example will clarify the situation:

Example 1.2: A box contains 6 white and 4 black balls. Remove two balls at random without replacement. What is the probability that the first one is white and the second one is black?

Let $W_1 =$ "first ball removed is white"

$B_2 =$ "second ball removed is black"

We need $P(W_1 \cap B_2) = ?$ We have $W_1 \cap B_2 = W_1 B_2 = B_2 W_1$. Using the conditional probability rule,

$$P(W_1 B_2) = P(B_2 W_1) = P(B_2 \mid W_1) P(W_1). \qquad (1\text{-}47)$$

But

$$P(W_1) = \frac{6}{6+4} = \frac{6}{10} = \frac{3}{5},$$

and

$$P(B_2 \mid W_1) = \frac{4}{5+4} = \frac{4}{9},$$

and hence

$$P(W_1 B_2) = \frac{5}{9} \cdot \frac{4}{9} = \frac{20}{81} \approx 0.25 .$$

Are the events W_1 and B_2 independent? Our common sense says No. To verify this we need to compute $P(B_2)$. Of course the fate of the second ball very much depends on that of the first ball. The first ball has two options: W_1 = "first ball is white" or B_1= "first ball is black". Note that $W_1 \cap B_1 = \phi$, and $W_1 \cup B_1 = \Omega$. Hence W_1 together with B_1 form a partition. Thus (see (1-42)-(1-44))

$$P(B_2) = P(B_2 \mid W_1)P(W_1) + P(B_2 \mid R_1)P(B_1)$$
$$= \frac{4}{5+4} \cdot \frac{3}{5} + \frac{3}{6+3} \cdot \frac{4}{10} = \frac{4}{9} \cdot \frac{3}{5} + \frac{1}{3} \cdot \frac{2}{5} = \frac{4+2}{15} = \frac{2}{5},$$

and

$$P(B_2)P(W_1) = \frac{2}{5} \cdot \frac{3}{5} \neq P(B_2 W_1) = \frac{20}{81}.$$

As expected, the events W_1 and B_2 are dependent.

From (1-35),

$$P(AB) = P(A \mid B) P(B). \tag{1-48}$$

Similarly, from (1-35)

$$P(B \mid A) = \frac{P(BA)}{P(A)} = \frac{P(AB)}{P(A)},$$

or

$$P(AB) = P(B \mid A) P(A). \tag{1-49}$$

From (1-48)-(1-49), we get

$$P(A \mid B) P(B) = P(B \mid A) P(A).$$

or

$$P(A \mid B) = \frac{P(B \mid A)}{P(B)} \cdot P(A) \tag{1-50}$$

Equation (1-50) is known as Bayes' theorem.

Although simple enough, Bayes' theorem has an interesting interpretation: $P(A)$ represents the a-priori probability of the event A. Suppose B has occurred, and assume that A and B are not independent. How can this new information be used to update our knowledge about A? Bayes' rule in (1-50) take into account the new information ("B has occurred") and gives out the a-posteriori probability of A given B.

We can also view the event B as new knowledge obtained from a fresh experiment. We know something about A as $P(A)$. The new information is available in terms of B. The new information should be used to improve our knowledge/understanding of A. Bayes' theorem gives the exact mechanism for incorporating such new information.

A more general version of Bayes' theorem involves partition of Ω. From (1-50)

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{P(B)} = \frac{P(B \mid A_i)P(A_i)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)}, \quad (1\text{-}51)$$

where we have made use of (1-44). In (1-51), $A_i, \; i = 1 \to n$, represent a set of mutually exclusive events with associated a-priori probabilities $P(A_i), \; i = 1 \to n.$ With the new information "B has occurred", the information about A_i can be updated by the n conditional probabilities $P(B \mid A_i), \; i = 1 \to n,$ using (1 - 47).

Example 1.3: Two boxes B_1 and B_2 contain 100 and 200 light bulbs respectively. The first box (B_1) has 15 defective bulbs and the second 5. Suppose a box is selected at random and one bulb is picked out.

(a) What is the probability that it is defective?

Solution: Note that box B_1 has 85 good and 15 defective bulbs. Similarly box B_2 has 195 good and 5 defective bulbs. Let D = "Defective bulb is picked out".

Then

$$P(D \mid B_1) = \frac{15}{100} = 0.15, \quad P(D \mid B_2) = \frac{5}{200} = 0.025 .$$

Since a box is selected at random, they are equally likely.

$$P(B_1) = P(B_2) = \frac{1}{2}.$$

Thus B_1 and B_2 form a partition as in (1-43), and using (1-44) we obtain

$$P(D) = P(D \mid B_1)P(B_1) + P(D \mid B_2)P(B_2)$$
$$= 0.15 \times \frac{1}{2} + 0.025 \times \frac{1}{2} = 0.0875 .$$

Thus, there is about 9% probability that a bulb picked at random is defective.

(b) Suppose we test the bulb and it is found to be defective. What is the probability that it came from box 1? $P(B_1 \mid D) = ?$

$$P(B_1 \mid D) = \frac{P(D \mid B_1)P(B_1)}{P(D)} = \frac{0.15 \times 1/2}{0.0875} = 0.8571 \ . \quad \text{(1-52)}$$

Notice that initially $P(B_1) = 0.5;$ then we picked out a box at random and tested a bulb that turned out to be defective. Can this information shed some light about the fact that we might have picked up box 1?

From (1-52), $P(B_1 \mid D) = 0.857 > 0.5,$ and indeed it is more likely at this point that we must have chosen box 1 in favor of box 2. (Recall box1 has six times more defective bulbs compared to box2).

2. Independence and Bernoulli Trials (Euler, Ramanujan and Bernoulli Numbers)

Independence: Events A and B are independent if

$$P(AB) = P(A)P(B). \tag{2-1}$$

- It is easy to show that A, B independent implies $\overline{A}, B$; $A, \overline{B}$; $\overline{A}, \overline{B}$ are all independent pairs. For example, $B = (A \cup \overline{A})B = AB \cup \overline{A}B$ and $AB \cap \overline{A}B = \phi$, so that

$$P(B) = P(AB \cup \overline{A}B) = P(AB) + P(\overline{A}B) = P(A)P(B) + P(\overline{A}B)$$

or

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

i.e., $\overline{A}$ and B are independent events.

As an application, let A_p and A_q represent the events

$$A_p = \text{"the prime } p \text{ divides the number } N\text{"}$$

and

$$A_q = \text{"the prime } q \text{ divides the number } N\text{"}.$$

Then from (1-4)

$$P\{A_p\} = \frac{1}{p}, \qquad P\{A_q\} = \frac{1}{q}$$

Also

$$P\{A_p \cap A_q\} = P\{\text{"}pq \text{ divides } N\text{"}\} = \frac{1}{pq} = P\{A_p\}\, P\{A_q\} \tag{2-2}$$

Hence it follows that A_p and A_q are independent events!

- If $P(A) = 0$, then since the event $AB \subset A$ always, we have

$$P(AB) \le P(A) = 0 \Rightarrow P(AB) = 0,$$

and (2-1) is always satisfied. Thus the event of zero probability is independent of every other event!

- Independent events obviously cannot be mutually exclusive, since $P(A) > 0$, $P(B) > 0$ and A, B independent implies $P(AB) > 0$. Thus if A and B are independent, the event AB cannot be the null set.

- More generally, a family of events $\{A_i\}$ are said to be independent, if for every finite sub collection $A_{i_1}, A_{i_2}, \cdots, A_{i_n}$, we have

$$P\left(\bigcap_{k=1}^{n} A_{i_k}\right) = \prod_{k=1}^{n} P(A_{i_k}). \tag{2-3}$$

- Let

$$A = A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n, \tag{2-4}$$

a union of n independent events. Then by De-Morgan's law

$$\overline{A} = \overline{A}_1 \overline{A}_2 \cdots \overline{A}_n$$

and using their independence

$$P(\overline{A}) = P(\overline{A}_1 \overline{A}_2 \cdots \overline{A}_n) = \prod_{i=1}^{n} P(\overline{A}_i) = \prod_{i=1}^{n} (1 - P(A_i)). \tag{2-5}$$

Thus for any A as in (2-4)

$$P(A) = 1 - P(\overline{A}) = 1 - \prod_{i=1}^{n} (1 - P(A_i)), \tag{2-6}$$

a useful result.

We can use these results to solve an interesting number theory problem.

Example 2.1 Two integers M and N are chosen at random. What is the probability that they are relatively prime to each other?

Solution: Since *M* and *N* are chosen at random, whether *p* divides *M* or not does not depend on the other number *N*. Thus we have

$$P\{\text{"}p \text{ divides both } M \text{ and } N\text{"}\}$$

$$= P\{\text{"}p \text{ divides } M\text{"}\}\, P\{\text{"}p \text{ divides } N\text{"}\} = \frac{1}{p^2}$$

where we have used (1-4). Also from (1-10)

$$P\{\text{"}p \text{ does } \textit{not} \text{ divede both } M \text{ and } N\text{"}\}$$

$$= 1 - P\{\text{"}p \text{ divides both } M \text{ and } N\text{"}\} = 1 - \frac{1}{p^2}$$

Observe that "*M* and *N* are relatively prime" if and only if there exists no prime p that divides both *M* and *N*.

Hence

$$"M \text{ and } N \text{ are relatively prime}" = \bar{X}_2 \cap \bar{X}_3 \cap \bar{X}_5 \cap \cdots$$

where X_p represents the event

$$X_p = "p \text{ divides both } M \text{ and } N".$$

Hence using (2-2) and (2-5)

$$P\{"M \text{ and N are relatively prime}"\} = \prod_{p \text{ prime}} P(\bar{X}_p)$$

$$= \prod_{p \text{ prime}} \left(1 - \tfrac{1}{p^2}\right) = \frac{1}{\sum_{k=1}^{\infty} 1/k^2} = \frac{1}{\pi^2/6} = \frac{6}{\pi^2} = 0.6079,$$

where we have used the Euler's identity[1]

---

[1]See Appendix for a proof of Euler's identity by Ramanujan.

$$\sum_{k=1}^{\infty} 1/k^s = \prod_{p \text{ prime}} (1-\tfrac{1}{p^s})^{-1}.$$

The same argument can be used to compute the probability that an integer chosen at random is "square free".
Since the event

$$\text{"An integer chosen at random is square free"}$$
$$= \bigcap_{p \text{ prime}} \{\text{"}p^2 \text{ does } \textit{not} \text{ divide } N\text{"}\},$$

using (2-5) we have

$$P\{\text{"An integer chosen at random is square free"}\}$$
$$= \prod_{p \text{ prime}} P\{p^2 \text{ does } \textit{not} \text{ divide } N\} = \prod_{p \text{ prime}} (1-\tfrac{1}{p^2})$$
$$= \frac{1}{\sum_{k=1}^{\infty} 1/k^2} = \frac{1}{\pi^2/6} = \frac{6}{\pi^2}.$$

Note: To add an interesting twist to the 'square free' number problem, Ramanujan has shown through elementary but clever arguments that the inverses of the n^{th} powers of all 'square free' numbers add to S_n / S_{2n}, where (see (2-E))

$$S_n = \sum_{k=1}^{\infty} 1/k^n.$$

Thus the sum of the inverses of the squares of 'square free' numbers is given by

$$\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{5^2}+\frac{1}{6^2}+\frac{1}{7^2}+\frac{1}{10^2}+\frac{1}{11^2}+\frac{1}{13^2}+\frac{1}{14^2}+\cdots=\frac{S_2}{S_4}$$

$$=\frac{\pi^2/6}{\pi^4/90}=\frac{15}{\pi^2}=1.51198.$$

Example 2.2: Three switches connected in parallel operate independently. Each switch remains closed with probability p. (a) Find the probability of receiving an input signal at the output. (b) Find the probability that switch S_1 is open given that an input signal is received at the output.

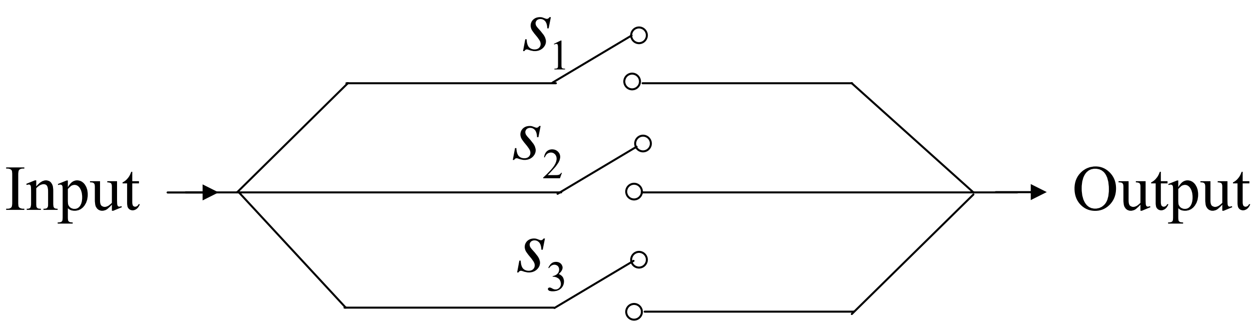


Fig.2.1

Solution: a. Let A_i = "Switch S_i is closed". Then $P(A_i) = p$, $i = 1 \to 3$. Since switches operate independently, we have

$$P(A_i A_j) = P(A_i)P(A_j);\ \ P(A_1 A_2 A_3) = P(A_1)P(A_2)P(A_3).$$

Let R = "input signal is received at the output". For the event R to occur either switch 1 or switch 2 or switch 3 must remain closed, i.e.,

$$R = A_1 \cup A_2 \cup A_3. \tag{2-7}$$

Using (2-3) - (2-6),

$$P(R) = P(A_1 \cup A_2 \cup A_3) = 1 - (1-p)^3 = 3p - 3p^2 + p^3. \tag{2-8}$$

We can also derive (2-8) in a different manner. Since any event and its compliment form a trivial partition, we can always write

$$P(R) = P(R \mid A_1)P(A_1) + P(R \mid \overline{A}_1)P(\overline{A}_1). \tag{2-9}$$

But $P(R \mid A_1) = 1,$ and $P(R \mid \overline{A}_1) = P(A_2 \cup A_3) = 2p - p^2$

and using these in (2-9) we obtain

$$P(R) = p + (2p - p^2)(1-p) = 3p - 3p^2 + p^3, \tag{2-10}$$

which agrees with (2-8).

Note that the events A_1, A_2, A_3 do not form a partition, since they are not mutually exclusive. Obviously any two or all three switches can be closed (or open) simultaneously. Moreover, $P(A_1) + P(A_2) + P(A_3) \neq 1.$

b. We need $P(\overline{A}_1 \mid R).$ From Bayes' theorem

$$P(\overline{A}_1 \mid R) = \frac{P(R \mid \overline{A}_1) P(\overline{A}_1)}{P(R)} = \frac{(2p - p^2)(1-p)}{3p - 3p^2 + p^3} = \frac{2 - 2p + p^2}{3p - 3p^2 + p^3}. \quad (2\text{-}11)$$

Because of the symmetry of the switches, we also have

$$P(\overline{A}_1 \mid R) = P(\overline{A}_2 \mid R) = P(\overline{A}_3 \mid R).$$

Repeated Trials

Consider two independent experiments with associated probability models (Ω_1, F_1, P_1) and (Ω_2, F_2, P_2). Let $\xi \in \Omega_1$, $\eta \in \Omega_2$ represent elementary events. A joint performance of the two experiments produces an elementary events $\omega = (\xi, \eta)$. How to characterize an appropriate probability to this "combined event" ? Towards this, consider the Cartesian product space $\Omega = \Omega_1 \times \Omega_2$ generated from Ω_1 and Ω_2 such that if $\xi \in \Omega_1$ and $\eta \in \Omega_2$, then every ω in Ω is an ordered pair of the form $\omega = (\xi, \eta)$. To arrive at a probability model we need to define the combined trio (Ω, F, P).

Suppose $A \in F_1$ and $B \in F_2$. Then $A \times B$ is the set of all pairs (ξ, η), where $\xi \in A$ and $\eta \in B$. Any such subset of Ω appears to be a legitimate event for the combined experiment. Let F denote the field composed of all such subsets $A \times B$ together with their unions and compliments. In this combined experiment, the probabilities of the events $A \times \Omega_2$ and $\Omega_1 \times B$ are such that

$$P(A \times \Omega_2) = P_1(A), \quad P(\Omega_1 \times B) = P_2(B). \qquad \text{(2-12)}$$

Moreover, the events $A \times \Omega_2$ and $\Omega_1 \times B$ are independent for any $A \in F_1$ and $B \in F_2$. Since

$$(A \times \Omega_2) \cap (\Omega_1 \times B) = A \times B, \qquad \text{(2-13)}$$

we conclude using (2-12) that

$$P(A\times B)=P(A\times\Omega_2)\cdot P(\Omega_1\times B)=P_1(A)P_2(B) \tag{2-14}$$

for all $A\in F_1$ and $B\in F_2$. The assignment in (2-14) extends to a unique probability measure $P(\equiv P_1\times P_2)$ on the sets in F and defines the combined trio (Ω, F, P).

Generalization: Given n experiments $\Omega_1,\Omega_2,\cdots,\Omega_n,$ and their associated F_i and $P_i,\ i=1\to n$, let

$$\Omega=\Omega_1\times\Omega_2\times\cdots\times\Omega_n \tag{2-15}$$

represent their Cartesian product whose elementary events are the ordered n-tuples $\xi_1,\xi_2,\cdots,\xi_n,$ where $\xi_i\in\Omega_i$. Events in this combined space are of the form

$$A_1\times A_2\times\cdots\times A_n \tag{2-16}$$

where $A_i\in F_i$, and their unions an intersections.

If all these n experiments are independent, and $P_i(A_i)$ is the probability of the event A_i in F_i then as before

$$P(A_1 \times A_2 \times \cdots \times A_n) = P_1(A_1)P_2(A_2)\cdots P_n(A_n). \tag{2-17}$$

Example 2.3: An event A has probability p of occurring in a single trial. Find the probability that A occurs exactly k times, $k \leq n$ in n trials.

Solution: Let (Ω, F, P) be the probability model for a single trial. The outcome of n experiments is an n-tuple

$$\omega = \{\xi_1, \xi_2, \cdots, \xi_n\} \in \Omega_0, \tag{2-18}$$

where every $\xi_i \in \Omega$ and $\Omega_0 = \Omega \times \Omega \times \cdots \times \Omega$ as in (2-15). The event A occurs at trial # i, if $\xi_i \in A$. Suppose A occurs exactly k times in ω.

Then k of the ξ_i belong to A, say $\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_k}$, and the remaining $n-k$ are contained in its compliment in $\overline{A}$. Using (2-17), the probability of occurrence of such an ω is given by

$$P_0(\omega) = P(\{\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_k}, \cdots, \xi_{i_n}\}) = P(\{\xi_{i_1}\}) P(\{\xi_{i_2}\}) \cdots P(\{\xi_{i_k}\}) \cdots P(\{\xi_{i_n}\})$$

$$= \underbrace{P(A)P(A)\cdots P(A)}_{k} \underbrace{P(\overline{A})P(\overline{A})\cdots P(\overline{A})}_{n-k} = p^k q^{n-k}. \qquad \text{(2-19)}$$

However the k occurrences of A can occur in any particular location inside ω. Let $\omega_1, \omega_2, \cdots, \omega_N$ represent all such events in which A occurs exactly k times. Then

$$\text{"}A \text{ occurs exactly } k \text{ times in } n \text{ trials"} = \omega_1 \cup \omega_2 \cup \cdots \cup \omega_N. \qquad \text{(2-20)}$$

But, all these ω_i s are mutually exclusive, and equiprobable.

Thus

$$P(\text{"}A \text{ occurs exactly } k \text{ times in } n \text{ trials"})$$

$$= \sum_{i=1}^{N} P_0(\omega_i) = NP_0(\omega) = Np^k q^{n-k}, \tag{2-21}$$

where we have used (2-19). Recall that, starting with n possible choices, the first object can be chosen n different ways, and for every such choice the second one in $(n-1)$ ways, … and the kth one $(n-k+1)$ ways, and this gives the total choices for k objects out of n to be $n\,(n-1)\cdots(n-k+1)$. But, this includes the $k!$ choices among the k objects that are indistinguishable for identical objects. As a result

$$N = \frac{n(n-1)\cdots(n-k+1)}{k!} = \frac{n!}{(n-k)!\,k!} \triangleq \binom{n}{k} \tag{2-22}$$

represents the number of combinations, or choices of n identical objects taken k at a time. Using (2-22) in (2-21), we get

$$
\begin{aligned}
P_n(k) &= P(\text{"}A \text{ occurs exactly } k \text{ times in } n \text{ trials"}) \\
&= \binom{n}{k} p^k q^{n-k}, \quad k = 0,1,2,\cdots,n,
\end{aligned}
\tag{2-23}
$$

a formula, due to Bernoulli.

Independent repeated experiments of this nature, where the outcome is either a "success" $(= A)$ or a "failure" $(= \overline{A})$ are characterized as Bernoulli trials, and the probability of k successes in n trials is given by (2-23), where p represents the probability of "success" in any one trial.

Example 2.4: Toss a coin n times. Obtain the probability of getting k heads in n trials ?

Solution: We may identify "head" with "success" (A) and let $p = P(H)$. In that case (2-23) gives the desired probability.

Example 2.5: Consider rolling a fair die eight times. Find the probability that either 3 or 4 shows up five times ?

Solution: In this case we can identify

$$\text{"success"} = A = \{\text{ either } 3 \text{ or } 4 \} = \{f_3\} \cup \{f_4\}.$$

Thus

$$P(A) = P(f_3) + P(f_4) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3},$$

and the desired probability is given by (2-23) with $n = 8$, $k = 5$ and $p = 1/3$. Notice that this is similar to a "biased coin" problem.

Bernoulli trial: consists of repeated independent and identical experiments each of which has only two outcomes A or $\overline{A}$ with $P(A) = p,$ and $P(\overline{A}) = q.$ The probability of exactly k occurrences of A in n such trials is given by (2-23).

Let

$$X_k = \text{"exactly } k \text{ occurrences in } n \text{ trials"}. \tag{2-24}$$

Since the number of occurrences of A in n trials must be an integer $k = 0,1,2,\cdots,n,$ either X_0 or X_1 or X_2 or $\cdots$ or X_n must occur in such an experiment. Thus

$$P(X_0 \cup X_1 \cup \cdots \cup X_n) = 1. \tag{2-25}$$

But $X_i,\ X_j$ are mutually exclusive. Thus

$$P(X_0 \cup X_1 \cup \cdots \cup X_n) = \sum_{k=0}^{n} P(X_k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}. \quad (2\text{-}26)$$

From the relation

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}, \quad (2\text{-}27)$$

(2-26) equals $(p+q)^n = 1,$ and it agrees with (2-25).

For a given *n* and *p* what is the most likely value of *k* ? From Fig.2.2, the most probable value of *k* is that number which maximizes $P_n(k)$ in (2-23). To obtain this value, consider the ratio

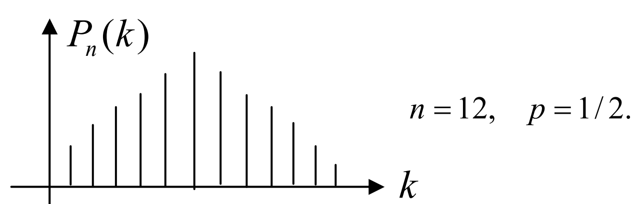


Fig. 2.2

$$\frac{P_n(k-1)}{P_n(k)} = \frac{n!\,p^{k-1}q^{n-k+1}}{(n-k+1)!\,(k-1)!}\,\frac{(n-k)!\,k!}{n!\,p^k q^{n-k}} = \frac{k}{n-k+1}\,\frac{q}{p}. \qquad \text{(2-28)}$$

Thus $P_n(k) \geq P_n(k-1)$, if $k(1-p) \leq (n-k+1)p$ or $k \leq (n+1)p$. Thus $P_n(k)$ as a function of k increases until

$$k = (n+1)p \qquad \text{(2-29)}$$

if it is an integer, or the largest integer $k_{\max}$ less than $(n+1)p$, and (2-29) represents the most likely number of successes (or heads) in n trials.

Example 2.6: In a Bernoulli experiment with n trials, find the probability that the number of occurrences of A is between k_1 and k_2.

Solution: With $X_i, i = 0,1,2,\cdots,n,$ as defined in (2-24), clearly they are mutually exclusive events. Thus

$$P(\text{"Occurrences of } A \text{ is between } k_1 \text{ and } k_2\text{"})$$

$$= P(X_{k_1} \cup X_{k_1+1} \cup \cdots \cup X_{k_2}) = \sum_{k=k_1}^{k_2} P(X_k) = \sum_{k=k_1}^{k_2} \binom{n}{k} p^k q^{n-k}. \quad (2\text{-}30)$$

Example 2.7: Suppose 5,000 components are ordered. The probability that a part is defective equals 0.1. What is the probability that the total number of defective parts does not exceed 400 ?

Solution: Let

$$Y_k = \text{"}k \text{ parts are defective among 5,000 components".}$$

Using (2-30), the desired probability is given by

$$P(Y_0 \cup Y_1 \cup \cdots \cup Y_{400}) = \sum_{k=0}^{400} P(Y_k)$$

$$= \sum_{k=0}^{400} \binom{5000}{k} (0.1)^k (0.9)^{5000-k}. \tag{2-31}$$

Equation (2-31) has too many terms to compute. Clearly, we need a technique to compute the above term in a more efficient manner.

From (2-29), $k_{\max}$ the most likely number of successes in n trials, satisfy

$$(n+1)p - 1 \le k_{\max} \le (n+1)p \tag{2-32}$$

or

$$p - \frac{q}{n} \le \frac{k_{\max}}{n} \le p + \frac{p}{n}, \tag{2-33}$$

so that

$$\lim_{n \to \infty} \frac{k_m}{n} = p. \tag{2-34}$$

From (2-34), as $n \to \infty$, the ratio of the most probable number of successes (A) to the total number of trials in a Bernoulli experiment tends to p, the probability of occurrence of A in a single trial. Notice that (2-34) connects the results of an actual experiment (k_m / n) to the axiomatic definition of p. In this context, it is possible to obtain a more general result as follows:

Bernoulli's theorem: Let A denote an event whose probability of occurrence in a single trial is p. If k denotes the number of occurrences of A in n independent trials, then

$$P\left(\left\{ \left| \frac{k}{n} - p \right| > \varepsilon \right\} \right) < \frac{pq}{n\varepsilon^2}. \tag{2-35}$$

Equation (2-35) states that the frequency definition of probability of an event $\frac{k}{n}$ and its axiomatic definition (p) can be made compatible to any degree of accuracy.

Proof: To prove Bernoulli's theorem, we need two identities. Note that with $P_n(k)$ as in (2-23), direct computation gives

$$\sum_{k=0}^{n} k\, P_n(k) = \sum_{k=1}^{n-1} k \frac{n!}{(n-k)!\,k!} p^k q^{n-k} = \sum_{k=1}^{n} \frac{n!}{(n-k)!\,(k-1)!} p^k q^{n-k}$$

$$= \sum_{i=0}^{n-1} \frac{n!}{(n-i-1)!\,i!} p^{i+1} q^{n-i-1} = np \sum_{i=0}^{n-1} \frac{(n-1)!}{(n-1-i)!\,i!} p^i q^{n-1-i}$$

$$= np(p+q)^{n-1} = np. \tag{2-36}$$

Proceeding in a similar manner, it can be shown that

$$\sum_{k=0}^{n} k^2 P_n(k) = \sum_{k=1}^{n} k \frac{n!}{(n-k)!\,(k-1)!} p^k q^{n-k} = \sum_{k=2}^{n} \frac{n!}{(n-k)!\,(k-2)!} p^k q^{n-k}$$

$$+ \sum_{k=1}^{n} \frac{n!}{(n-k)!\,(k-1)!} p^k q^{n-k} = n^2 p^2 + npq. \tag{2-37}$$

Returning to (2-35), note that

$$\left|\frac{k}{n}-p\right|>\varepsilon \quad \text{is equivalent to} \quad (k-np)^2>n^2\varepsilon^2, \tag{2-38}$$

which in turn is equivalent to

$$\sum_{k=0}^{n}(k-np)^2 P_n(k) > \sum_{k=0}^{n} n^2\varepsilon^2 P_n(k) = n^2\varepsilon^2. \tag{2-39}$$

Using (2-36)-(2-37), the left side of (2-39) can be expanded to give

$$\sum_{k=0}^{n}(k-np)^2 P_n(k) = \sum_{k=0}^{n} k^2 P_n(k) - 2np\sum_{k=0}^{n} k P_n(k) + n^2p^2$$

$$= n^2p^2 + npq - 2np\cdot np + n^2p^2 = npq. \tag{2-40}$$

Alternatively, the left side of (2-39) can be expressed as

$$\sum_{k=0}^{n}(k-np)^2 P_n(k) = \sum_{|k-np|\le n\varepsilon}(k-np)^2 P_n(k) + \sum_{|k-np|>n\varepsilon}(k-np)^2 P_n(k)$$

$$\ge \sum_{|k-np|>n\varepsilon}(k-np)^2 P_n(k) > n^2\varepsilon^2 \sum_{|k-np|>n\varepsilon} P_n(k)$$

$$= n^2\varepsilon^2 P\{|k-np|>n\varepsilon\}. \tag{2-41}$$

Using (2-40) in (2-41), we get the desired result

$$P\left(\left\{\left|\frac{k}{n} - p\right| > \varepsilon\right\}\right) < \frac{pq}{n\varepsilon^2}. \qquad (2\text{-}42)$$

Note that for a given $\varepsilon > 0,\ pq/n\varepsilon^2$ can be made arbitrarily small by letting *n* become large. Thus for very large *n*, we can make the fractional occurrence (relative frequency) $\frac{k}{n}$ of the event *A* as close to the actual probability *p* of the event *A* in a single trial. Thus the theorem states that the probability of event *A* from the axiomatic framework can be computed from the relative frequency definition quite accurately, provided the number of experiments are large enough. Since $k_{\max}$ is the most likely value of k in n trials, from the above discussion, as $n \to \infty$, the plots of $P_n(k)$ tends to concentrate more and more around $k_{\max}$ in (2-32).

Next we present an example that illustrates the usefulness of "simple textbook examples" to practical problems of interest:

Example 2.8 : **Day-trading strategy :** A box contains n *randomly* numbered balls (not 1 through n but arbitrary numbers including numbers greater than n). Suppose a fraction of those balls – say $m = np$; $p < 1$ – are initially drawn one by one with replacement while noting the numbers on those balls. The drawing is allowed to continue *until* a ball is drawn with a number larger than the first m numbers. Determine the fraction p to be initially drawn, so as to maximize the probability of drawing the largest among the n numbers using this strategy.

Solution: Let "$X_k = (k+1)^{st}$ drawn ball has the largest number among all n balls, and the largest among the first k balls is in the group of first m balls, $k > m$." (2.43)

Note that X_k is of the form $A \cap B$,

where

A = "largest among the first k balls is in the group of first m balls drawn"

and

B = "$(k+1)^{st}$ ball has the largest number among all n balls".

Notice that A and B are independent events, and hence

$$P(X_k) = P(A)P(B) = \frac{1}{n}\frac{m}{k} = \frac{1}{n}\frac{np}{k} = \frac{p}{k}. \tag{2-44}$$

Where $m = np$ represents the fraction of balls to be initially drawn. This gives

P ("selected ball has the largest number among all balls")

$$= \sum_{k=m}^{n-1} P(X_k) = p \sum_{k=m}^{n-1} \frac{1}{k} \approx p \int_{np}^{n} \frac{1}{k} = p \ln k \Big|_{np}^{n}$$

$$= -p \ln p. \tag{2-45}$$

Maximization of the desired probability in (2-45) with respect to p gives

$$\frac{d}{dp}(-p\ln p) = -(1+\ln p) = 0$$

or

$$p = e^{-1} \simeq 0.3679. \qquad (2\text{-}46)$$

From (2-45), the maximum value for the desired probability of drawing the largest number equals 0.3679 also. Interestingly the above strategy can be used to "play the stock market".

Suppose one gets into the market and decides to stay up to 100 days. The stock values fluctuate day by day, and the important question is when to get out?

According to the above strategy, one should get out at the first opportunity after 37 days, when the stock value exceeds the maximum among the first 37 days. In that case the probability of hitting the top value over 100 days for the stock is also about 37%. Of course, the above argument assumes that the stock values over the period of interest are randomly fluctuating without exhibiting any other trend. Interestingly, such is the case if we consider shorter time frames such as inter-day trading.

In summary if one must day-trade, then a possible strategy might be to get in at 9.30 AM, and get out any time after 12 noon (9.30 AM $+ 0.3679 \times 6.5$ hrs $= 11.54$ AM to be precise) at the first peak that exceeds the peak value between 9.30 AM and 12 noon. In that case chances are about 37% that one hits the absolute top value for that day! (disclaimer : Trade at your own risk)

We conclude this lecture with a variation of the *Game of craps* discussed in Example 3-16, Text.

Example 2.9: **Game of craps using biased dice:**

From Example 3.16, Text, the probability of winning the game of craps is $0.492929\ \cdots$ for the player. Thus the game is slightly advantageous to the house. This conclusion of course assumes that the two dice in question are perfect cubes. Suppose that is not the case.

Let us assume that the two dice are slightly loaded in such a manner so that the faces 1, 2 and 3 appear with probability $\frac{1}{6}-\varepsilon$ and faces 4, 5 and 6 appear with probability $\frac{1}{6}+\varepsilon, \ \ \varepsilon > 0$ for each dice. If T represents the combined total for the two dice (following Text notation), we get

$$p_4 = P\{T = 4\} = P\{(1,3),(2,2),(1,3)\} = 3(\tfrac{1}{6} - \varepsilon)^2$$

$$p_5 = P\{T = 5\} = P\{(1,4),(2,3),(3,2),(4,1)\} = 2(\tfrac{1}{36} - \varepsilon^2) + 2(\tfrac{1}{6} - \varepsilon)^2$$

$$p_6 = P\{T = 6\} = P\{(1,5),(2,4),(3,3),(4,2),(5,1)\} = 4(\tfrac{1}{36} - \varepsilon^2) + (\tfrac{1}{6} - \varepsilon)^2$$

$$p_7 = P\{T = 7\} = P\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} = 6(\tfrac{1}{36} - \varepsilon^2)$$

$$p_8 = P\{T = 8\} = P\{(2,6),(3,5),(4,4),(5,3),(6,2)\} = 4(\tfrac{1}{36} - \varepsilon^2) + (\tfrac{1}{6} + \varepsilon)^2$$

$$p_9 = P\{T = 9\} = P\{(3,6),(4,5),(5,4),(6,3)\} = 2(\tfrac{1}{36} - \varepsilon^2) + 2(\tfrac{1}{6} + \varepsilon)^2$$

$$p_{10} = P\{T = 10\} = P\{(4,6),(5,5),(6,4)\} = 3(\tfrac{1}{6} + \varepsilon)^2$$

$$p_{11} = P\{T = 11\} = P\{(5,6),(6,5)\} = 2(\tfrac{1}{6} + \varepsilon)^2.$$

(Note that "(1,3)" above represents the event "the first dice shows face 1, and the second dice shows face 3" etc.)
For $\varepsilon = 0.01,$ we get the following Table:

$T = k$	4	5	6	7	8	9	10	11
$p_k = P\{T = k\}$	0.0706	0.1044	0.1353	0.1661	0.1419	0.1178	0.0936	0.0624

This gives the probability of win on the first throw to be (use (3-56), Text)

$$P_1 = P(T = 7) + P(T = 11) = 0.2285 \tag{2-47}$$

and the probability of win by throwing a carry-over to be (use (3-58)-(3-59), Text)

$$P_2 = \sum_{\substack{k=4 \\ k \neq 7}}^{10} \frac{p_k^2}{p_k + p_7} = 0.2717 \tag{2-48}$$

Thus

$$P\{\text{winning the game}\} = P_1 + P_2 = 0.5002 \tag{2-49}$$

Although perfect dice gives rise to an unfavorable game, a slight loading of the dice turns the fortunes around in favor of the player! (Not an exciting conclusion as far as the casinos are concerned).

Even if we let the two dice to have different loading factors ε_1 and ε_2 (for the situation described above), similar conclusions do follow. For example, $\varepsilon_1 = 0.01$ and $\varepsilon_2 = 0.005$ gives (show this)

$$P\{\text{winning the game}\} = 0.5015. \tag{2-50}$$

Once again the game is in favor of the player!

Although the advantage is very modest in each play, from Bernoulli's theorem the cumulative effect can be quite significant when a large number of game are played. All the more reason for the casinos to keep the dice in perfect shape.

In summary, small chance variations in each game of craps can lead to significant counter-intuitive changes when a large number of games are played. What appears to be a favorable game for the house may indeed become an unfavorable game, and when played repeatedly can lead to unpleasant outcomes.

Appendix: Euler's Identity

S. Ramanujan in one of his early papers (*J. of Indian Math Soc*; V, 1913) starts with the clever observation that if $a_2, a_3, a_5, a_7, a_{11}, \cdots$ are numbers less than unity where the subscripts $2,3,5,7,11,\cdots$ are the series of prime numbers, then[1]

$$\frac{1}{1-a_2}\cdot\frac{1}{1-a_3}\cdot\frac{1}{1-a_5}\cdot\frac{1}{1-a_7}\cdots = 1 + a_2 + a_3 + a_2\cdot a_2 + a_5$$
$$+ a_2\cdot a_3 + a_7 + a_2\cdot a_2\cdot a_2 + a_3\cdot a_3 + \cdots. \quad \text{(2-A)}$$

Notice that the terms in (2-A) are arranged in such a way that the product obtained by multiplying the subscripts are the series of all natural numbers $2,3,4,5,6,7,8,9,\cdots$. Clearly, (2-A) follows by observing that the natural numbers

[1]The relation (2-A) is ancient.

are formed by multiplying primes and their powers.

Ramanujan uses (2-A) to derive a variety of interesting identities including the Euler's identity that follows by letting $a_2 = 1/2^s, a_3 = 1/3^s, a_5 = 1/5^s, \cdots$ in (2-A). This gives the Euler identity

$$\prod_{p \text{ prime}} \left(1 - \tfrac{1}{p^s}\right)^{-1} = \sum_{n=1}^{\infty} 1/n^s. \tag{2-B}$$

The sum on the right side in (2-B) can be related to the Bernoulli numbers (for *s* even).

Bernoulli numbers are positive rational numbers defined through the power series expansion of the even function $\frac{x}{2}\cot(x/2)$. Thus if we write

$$\frac{x}{2}\cot(x/2) \triangleq 1 - B_1 \frac{x^2}{2!} - B_2 \frac{x^4}{4!} - B_3 \frac{x^6}{6!} - \cdots \tag{2-C}$$

then $B_1 = \frac{1}{6}, B_2 = \frac{1}{30}, B_3 = \frac{1}{42}, B_4 = \frac{1}{30}, B_5 = \frac{1}{66}, \cdots.$

By direct manipulation of (2-C) we also obtain

$$\frac{x}{e^x-1}=1-\frac{x}{2}+\frac{B_1x^2}{2!}-\frac{B_2x^4}{4!}+\frac{B_3x^6}{6!}-\cdots \qquad \text{(2-D)}$$

so that the Bernoulli numbers may be defined through (2-D) as well. Further

$$B_n=4n\int_0^\infty \frac{x^{2n-1}}{e^{2\pi x}-1}\,dx=\int_0^\infty x^{2n-1}(e^{-2\pi x}+e^{-4\pi x}+\cdots)dx$$

$$=\frac{2(2n)!}{(2\pi)^{2n}}\left(\frac{1}{1^{2n}}+\frac{1}{2^{2n}}+\frac{1}{3^{2n}}+\frac{1}{4^{2n}}+\cdots\right)$$

which gives

$$S_{2n}\triangleq\sum_{k=1}^{\infty}1/k^{2n}=\frac{(2\pi)^{2n}B_n}{2(2n)!} \qquad \text{(2-E)}$$

Thus[1]

$$\sum_{k=1}^{\infty}1/k^2=\frac{\pi^2}{6};\qquad \sum_{k=1}^{\infty}1/k^4=\frac{\pi^4}{90}\quad \text{etc.}$$

[1]The series $\sum_{k=1}^{\infty}1/k^2$ can be summed using the Fourier series expansion of a periodic ramp signal as well.

3. Random Variables

Let (Ω, F, P) be a probability model for an experiment, and X a function that maps every $\xi \in \Omega$, to a unique point $x \in R$, the set of real numbers. Since the outcome ξ is not certain, so is the value $X(\xi) = x$. Thus if B is some subset of R, we may want to determine the probability of "$X(\xi) \in B$". To determine this probability, we can look at the set $A = X^{-1}(B) \in \Omega$ that contains all $\xi \in \Omega$ that maps into B under the function X.

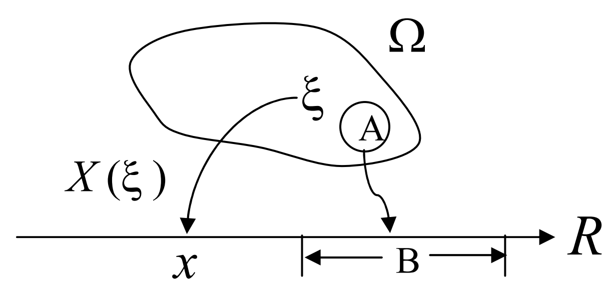


Fig. 3.1

Obviously, if the set $A = X^{-1}(B)$ also belongs to the associated field F, then it is an event and the probability of A is well defined; in that case we can say

$$\text{Probability of the event } "X(\xi) \in B" = P(X^{-1}(B)). \qquad (3\text{-}1)$$

However, $X^{-1}(B)$ may not always belong to F for all B, thus creating difficulties. The notion of random variable (r.v) makes sure that the inverse mapping always results in an event so that we are able to determine the probability for any $B \in R$.

Random Variable (r.v): A finite single valued function $X(\cdot)$ that maps the set of all experimental outcomes Ω into the set of real numbers R is said to be a r.v, if the set $\{\xi \mid X(\xi) \leq x\}$ is an event $(\in F)$ for every x in R.

Alternatively X is said to be a r.v, if $X^{-1}(B) \in F$ where B represents semi-definite intervals of the form $\{-\infty < x \leq a\}$ and all other sets that can be constructed from these sets by performing the set operations of union, intersection and negation any number of times. The Borel collection B of such subsets of R is the smallest σ-field of subsets of R that includes all semi-infinite intervals of the above form. Thus if X is a r.v, then

$$\{\xi \mid X(\xi) \leq x\} = \{X \leq x\} \tag{3-2}$$

is an event for every x. What about $\{a < X \leq b\}$, $\{X = a\}$? Are they also events ? In fact with $b > a$ since $\{X \leq a\}$ and $\{X \leq b\}$ are events, $\{X \leq a\}^c = \{X > a\}$ is an event and hence $\{X > a\} \cap \{X \leq b\} = \{a < X \leq b\}$ is also an event.

Thus, $\left\{ a - \frac{1}{n} < X \le a \right\}$ is an event for every n. Consequently

$$\bigcap_{n=1}^{\infty} \left\{ a - \frac{1}{n} < X \le a \right\} = \{ X = a \} \tag{3-3}$$

is also an event. All events have well defined probability. Thus the probability of the event $\{ \xi \mid X(\xi) \le x \}$ must depend on x. Denote

$$P\{ \xi \mid X(\xi) \le x \} = F_X(x) \ge 0. \tag{3-4}$$

The role of the subscript X in (3-4) is only to identify the actual r.v. $F_X(x)$ is said to the Probability Distribution Function (PDF) associated with the r.v X.

Distribution Function: Note that a distribution function $g(x)$ is nondecreasing, right-continuous and satisfies

$$g(+\infty) = 1, \quad g(-\infty) = 0, \tag{3-5}$$

i.e., if $g(x)$ is a distribution function, then

(i) $g(+\infty) = 1, \quad g(-\infty) = 0,$

(ii) if $x_1 < x_2$, then $g(x_1) \le g(x_2)$, (3-6)

and

(iii) $g(x^+) = g(x)$, for all x.

We need to show that $F_X(x)$ defined in (3-4) satisfies all properties in (3-6). In fact, for any r.v X,

(i) $$F_X(+\infty) = P\{\xi \mid X(\xi) \le +\infty\} = P(\Omega) = 1 \tag{3-7}$$

and $$F_X(-\infty) = P\{\xi \mid X(\xi) \le -\infty\} = P(\phi) = 0. \tag{3-8}$$

(ii) If $x_1 < x_2,$ then the subset $(-\infty, x_1) \subset (-\infty, x_2).$ Consequently the event $\{\xi \mid X(\xi) \le x_1\} \subset \{\xi \mid X(\xi) \le x_2\},$ since $X(\xi) \le x_1$ implies $X(\xi) \le x_2.$ As a result

$$F_X(x_1) \triangleq P\big(X(\xi) \le x_1\big) \le P\big(X(\xi) \le x_2\big) \triangleq F_X(x_2), \tag{3-9}$$

implying that the probability distribution function is nonnegative and monotone nondecreasing.

(iii) Let $x < x_n < x_{n-1} < \cdots < x_2 < x_1,$ and consider the event

$$A_k = \{\xi \mid x < X(\xi) \le x_k\}. \tag{3-10}$$

since

$$\{x < X(\xi) \le x_k\} \cup \{X(\xi) \le x\} = \{X(\xi) \le x_k\}, \tag{3-11}$$

using mutually exclusive property of events we get

$$P(A_k) = P\big(x < X(\xi) \le x_k\big) = F_X(x_k) - F_X(x). \tag{3-12}$$

But $\cdots A_{k+1} \subset A_k \subset A_{k-1} \cdots,$ and hence

$$\lim_{k\to\infty} A_k = \bigcap_{k=1}^{\infty} A_k = \phi \quad \text{and hence} \quad \lim_{k\to\infty} P(A_k) = 0. \tag{3-13}$$

Thus

$$\lim_{k\to\infty} P(A_k) = \lim_{k\to\infty} F_X(x_k) - F_X(x) = 0.$$

But $\lim_{k\to\infty} x_k = x^+,$ the right limit of x, and hence

$$F_X(x^+) = F_X(x), \tag{3-14}$$

i.e., $F_X(x)$ is right-continuous, justifying all properties of a distribution function.

Additional Properties of a PDF

(iv) If $F_X(x_0) = 0$ for some x_0, then $F_X(x) = 0, \quad x \le x_0.$ (3-15)

This follows, since $F_X(x_0) = P\left(X(\xi) \le x_0\right) = 0$ implies $\{X(\xi) \le x_0\}$ is the null set, and for any $x \le x_0$, $\{ X(\xi) \le x \}$ will be a subset of the null set.

(v) $P\{ X(\xi) > x \} = 1 - F_X(x).$ (3-16)

We have $\{ X(\xi) \le x \} \cup \{ X(\xi) > x \} = \Omega$, and since the two events are mutually exclusive, (16) follows.

(vi) $P\{ x_1 < X(\xi) \le x_2 \} = F_X(x_2) - F_X(x_1), \quad x_2 > x_1.$ (3-17)

The events $\{ X(\xi) \le x_1 \}$ and $\{x_1 < X(\xi) \le x_2\}$ are mutually exclusive and their union represents the event $\{ X(\xi) \le x_2 \}$.

(vii) $P(X(\xi) = x) = F_X(x) - F_X(x^-).$ (3-18)

Let $x_1 = x - \varepsilon, \;\; \varepsilon > 0,$ and $x_2 = x.$ From (3-17)

$$\lim_{\varepsilon \to 0} P\{ x - \varepsilon < X(\xi) \le x \} = F_X(x) - \lim_{\varepsilon \to 0} F_X(x - \varepsilon), \tag{3-19}$$

or

$$P\{ X(\xi) = x \} = F_X(x) - F_X(x^-). \tag{3-20}$$

According to (3-14), $F_X(x_0^+)$, the limit of $F_X(x)$ as $x \to x_0$ from the right always exists and equals $F_X(x_0)$. However the left limit value $F_X(x_0^-)$ need not equal $F_X(x_0)$. Thus $F_X(x)$ need not be continuous from the left. At a discontinuity point of the distribution, the left and right limits are different, and from (3-20)

$$P\{ X(\xi) = x_0 \} = F_X(x_0) - F_X(x_0^-) > 0. \tag{3-21}$$

Thus the only discontinuities of a distribution function $F_X(x)$ are of the jump type, and occur at points x_0 where (3-21) is satisfied. These points can always be enumerated as a sequence, and moreover they are at most countable in number.

Example 3.1: X is a r.v such that $X(\xi) = c,\ \xi \in \Omega$. Find $F_X(x)$.
Solution: For $x < c,\ \{X(\xi) \le x\} = \{\phi\}$, so that $F_X(x) = 0$, and for $x > c,\ \{X(\xi) \le x\} = \Omega$, so that $F_X(x) = 1$. (Fig.3.2)

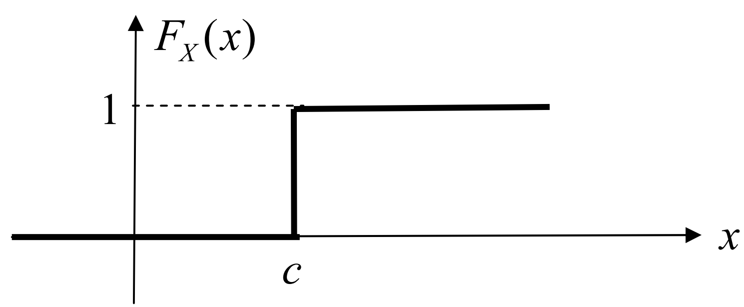


Fig. 3.2

Example 3.2: Toss a coin. $\Omega = \{H, T\}$. Suppose the r.v X is such that $X(T) = 0,\ X(H) = 1$. Find $F_X(x)$.

Solution: For $x < 0, \{ X(\xi) \le x \} = \{ \phi \}$, so that $F_X(x) = 0$.

$0 \le x < 1$, $\{ X(\xi) \le x \} = \{ T \}$, so that $F_X(x) = P\{ T \} = 1 - p$,

$x \ge 1$, $\{ X(\xi) \le x \} = \{ H, T \} = \Omega$, so that $F_X(x) = 1$. (Fig. 3.3)

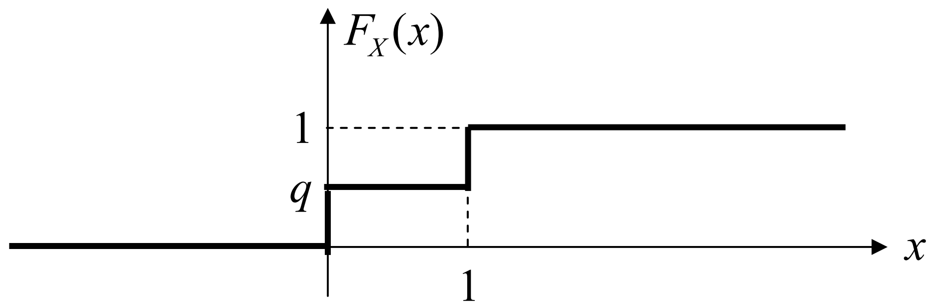


Fig.3.3

•X is said to be a continuous-type r.v if its distribution function $F_X(x)$ is continuous. In that case $F_X(x^-) = F_X(x)$ for all x, and from (3-21) we get $P\{X = x\} = 0$.

•If $F_X(x)$ is constant except for a finite number of jump discontinuities(piece-wise constant; step-type), then X is said to be a discrete-type r.v. If x_i is such a discontinuity point, then from (3-21)

$$p_i = P\{X = x_i\} = F_X(x_i) - F_X(x_i^-). \qquad (3\text{-}22)$$

From Fig.3.2, at a point of discontinuity we get

$$P\{ X = c \} = F_X(c) - F_X(c^-) = 1 - 0 = 1.$$

and from Fig.3.3,

$$P\{ X = 0 \} = F_X(0) - F_X(0^-) = q - 0 = q.$$

Example:3.3 A fair coin is tossed twice, and let the r.v X represent the number of heads. Find $F_X(x)$.
Solution: In this case $\Omega = \{ HH, HT, TH, TT \}$, and

$$X(HH) = 2, X(HT) = 1, X(TH) = 1, X(TT) = 0.$$

$$x < 0, \quad \{X(\xi) \le x\} = \phi \Rightarrow F_X(x) = 0,$$

$$0 \le x < 1, \{X(\xi) \le x\} = \{ TT \} \Rightarrow F_X(x) = P\{ TT \} = P(T)P(T) = \frac{1}{4},$$

$$1 \le x < 2, \{X(\xi) \le x\} = \{ TT, HT, TH \} \Rightarrow F_X(x) = P\{ TT, HT, TH \} = \frac{3}{4},$$

$$x \ge 2, \quad \{X(\xi) \le x\} = \Omega \Rightarrow F_X(x) = 1. \text{ (Fig. 3.4)}$$

From Fig.3.4, $P\{X=1\}=F_X(1)-F_X(1^-)=3/4-1/4=1/2.$

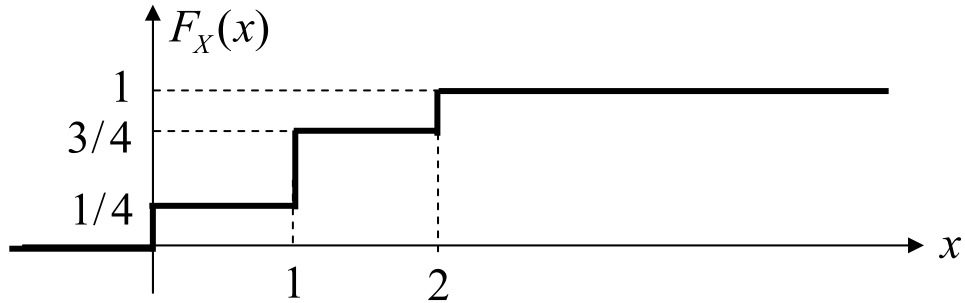


Fig. 3.4

Probability density function (p.d.f)

The derivative of the distribution function $F_X(x)$ is called the probability density function $f_X(x)$ of the r.v X. Thus

$$f_X(x) \stackrel{\Delta}{=} \frac{dF_X(x)}{dx}. \tag{3-23}$$

Since

$$\frac{dF_X(x)}{dx} = \lim_{\Delta x \to 0} \frac{F_X(x+\Delta x)-F_X(x)}{\Delta x} \geq 0, \tag{3-24}$$

from the monotone-nondecreasing nature of $F_X(x)$, it follows that $f_X(x) \geq 0$ for all x. $f_X(x)$ will be a continuous function, if X is a continuous type r.v. However, if X is a discrete type r.v as in (3-22), then its p.d.f has the general form (Fig. 3.5)

$$f_X(x) = \sum_i p_i \delta(x - x_i), \qquad (3\text{-}25)$$

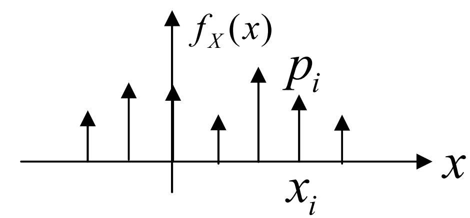


Fig. 3.5

where x_i represent the jump-discontinuity points in $F_X(x)$. As Fig. 3.5 shows $f_X(x)$ represents a collection of positive discrete masses, and it is known as the probability mass function (p.m.f) in the discrete case. From (3-23), we also obtain by integration

$$F_X(x) = \int_{-\infty}^{x} f_x(u)\,du . \qquad (3\text{-}26)$$

Since $F_X(+\infty) = 1,$ (3-26) yields

$$\int_{-\infty}^{+\infty} f_x(x)\,dx = 1, \qquad (3\text{-}27)$$

which justifies its name as the density function. Further, from (3-26), we also get (Fig. 3.6b)

$$P\{x_1 < X(\xi) \le x_2\} = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)\,dx. \qquad (3\text{-}28)$$

Thus the area under $f_X(x)$ in the interval (x_1, x_2) represents the probability in (3-28).

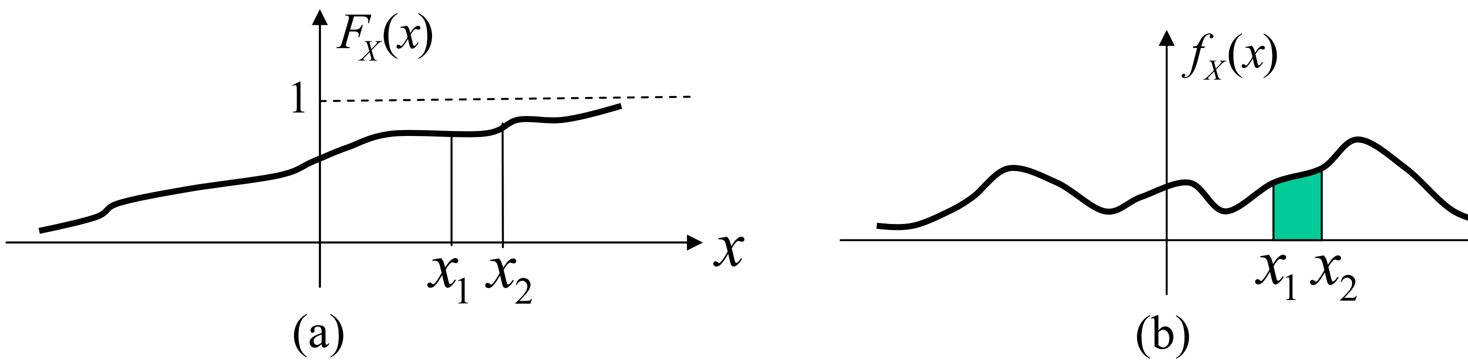


Fig. 3.6

Often, r.vs are referred by their specific density functions - both in the continuous and discrete cases - and in what follows we shall list a number of them in each category.

Continuous-type random variables

1. Normal (Gaussian): X is said to be normal or Gaussian r.v, if

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}. \qquad (3\text{-}29)$$

This is a bell shaped curve, symmetric around the parameter μ, and its distribution function is given by

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2} dy \overset{\Delta}{=} G\left(\frac{x-\mu}{\sigma}\right), \qquad (3\text{-}30)$$

where $G(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$ is often tabulated. Since $f_X(x)$ depends on two parameters μ and σ^2, the notation $X \sim N(\mu, \sigma^2)$ will be used to represent (3-29).

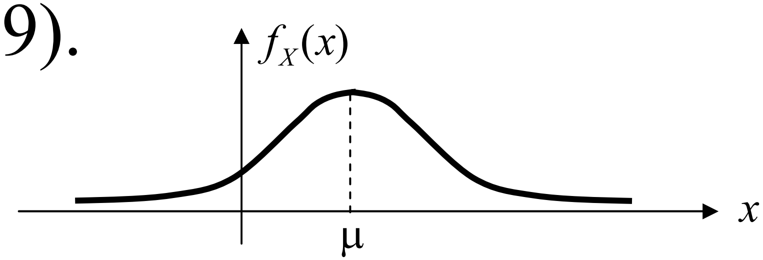


Fig. 3.7

2. Uniform: $X \sim U(a,b),\ a<b,$ if (Fig. 3.8)

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b, \\ 0, & \text{otherwise.} \end{cases} \tag{3.31}$$

3. Exponential: $X \sim \varepsilon(\lambda)$ if (Fig. 3.9)

$$f_X(x) = \begin{cases} \dfrac{1}{\lambda} e^{-x/\lambda}, & x \ge 0, \\ 0, & \text{otherwise.} \end{cases} \tag{3-32}$$

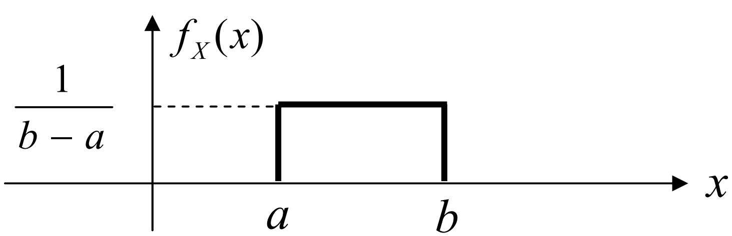


Fig. 3.8

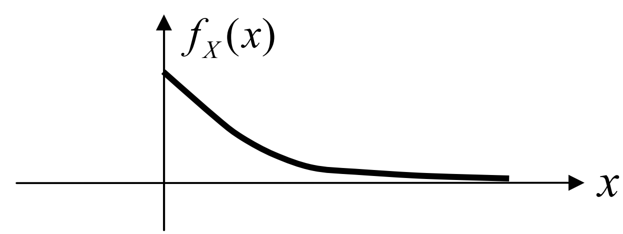


Fig. 3.9

4. Gamma: $X \sim G(\alpha, \beta)$ if $(\alpha > 0,\ \beta > 0)$ (Fig. 3.10)

$$f_X(x) = \begin{cases} \dfrac{x^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}} e^{-x/\beta}, & x \ge 0, \\ 0, & \text{otherwise.} \end{cases} \qquad (3\text{-}33)$$

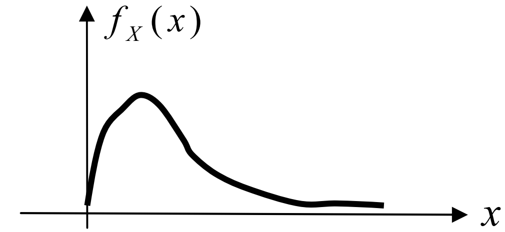


Fig. 3.10

If $\alpha = n$ an integer $\Gamma(n) = (n-1)!$.

5. Beta: $X \sim \beta(a,b)$ if $(a > 0,\ b > 0)$ (Fig. 3.11)

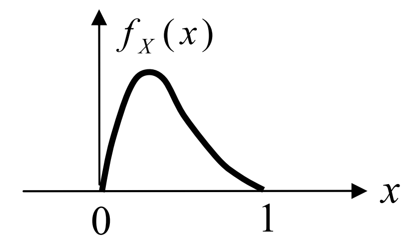


Fig. 3.11

$$f_X(x) = \begin{cases} \dfrac{1}{\beta(a,b)} x^{a-1}(1-x)^{b-1}, & 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases} \qquad (3\text{-}34)$$

where the Beta function $\beta(a,b)$ is defined as

$$\beta(a,b) = \int_0^1 u^{a-1}(1-u)^{b-1}\,du. \qquad (3\text{-}35)$$

6. Chi-Square: $X \sim \chi^2(n)$, if (Fig. 3.12)

$$f_X(x) = \begin{cases} \dfrac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}, & x \geq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{3-36}$$

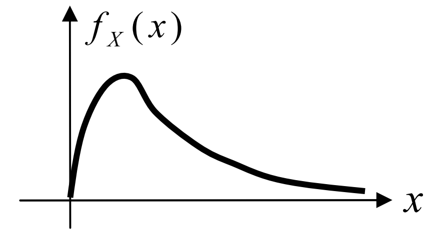


Fig. 3.12

Note that $\chi^2(n)$ is the same as Gamma $(n/2, 2)$.

7. Rayleigh: $X \sim R(\sigma^2)$, if (Fig. 3.13)

$$f_X(x) = \begin{cases} \dfrac{x}{\sigma^2} e^{-x^2/2\sigma^2}, & x \geq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{3-37}$$

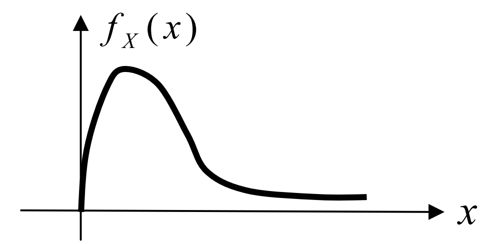


Fig. 3.13

8. Nakagami – m distribution:

$$f_X(x) = \begin{cases} \dfrac{2}{\Gamma(m)} \left(\dfrac{m}{\Omega}\right)^m x^{2m-1} e^{-mx^2/\Omega}, & x \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{3-38}$$

9. Cauchy: $X \sim C(\alpha, \mu)$, if (Fig. 3.14)

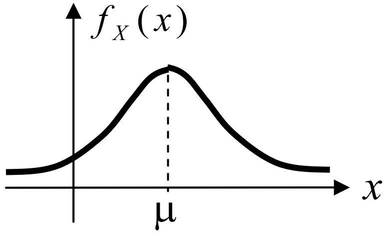


Fig. 3.14

$$f_X(x) = \frac{\alpha / \pi}{\alpha^2 + (x - \mu)^2}, \quad -\infty < x < +\infty. \qquad (3\text{-}39)$$

10. Laplace: (Fig. 3.15)

$$f_X(x) = \frac{1}{2\lambda} e^{-|x|/\lambda}, \quad -\infty < x < +\infty. \qquad (3\text{-}40)$$

11. Student's *t*-distribution with n degrees of freedom (Fig 3.16)

$$f_T(t) = \frac{\Gamma\left((n+1)/2\right)}{\sqrt{\pi n}\,\Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}, \quad -\infty < t < +\infty. \qquad (3\text{-}41)$$

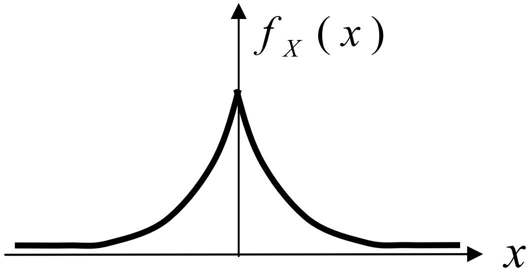


Fig. 3.15

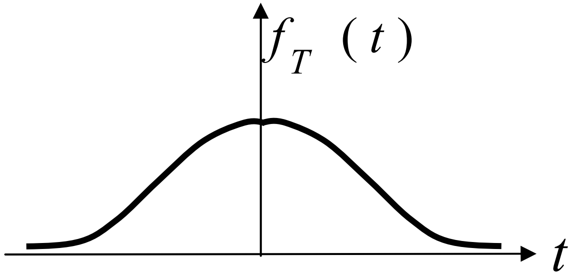


Fig. 3.16

12. Fisher's F-distribution

$$f_z(z) = \begin{cases} \dfrac{\Gamma\{(m+n)/2\}\, m^{m/2} n^{n/2}}{\Gamma(m/2)\,\Gamma(n/2)} \dfrac{z^{m/2-1}}{(n+mz)^{(m+n)/2}}, & z \geq 0 \\ 0 & \text{otherwise} \end{cases} \qquad (3\text{-}42)$$

Discrete-type random variables

1. Bernoulli: X takes the values (0,1), and

$$P(X=0)=q, \quad P(X=1)=p. \tag{3-43}$$

2. Binomial: $X \sim B(n,p)$, if (Fig. 3.17)

$$P(X=k)=\binom{n}{k} p^k q^{n-k}, \quad k=0,1,2,\cdots,n. \tag{3-44}$$

3. Poisson: $X \sim P(\lambda)$, if (Fig. 3.18)

$$P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!}, \quad k=0,1,2,\cdots,\infty. \tag{3-45}$$

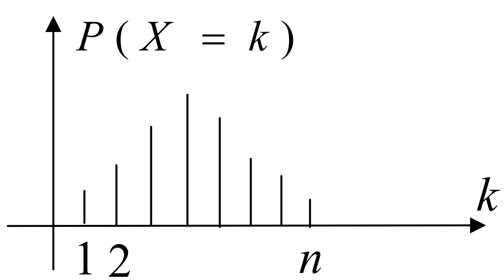


Fig. 3.17

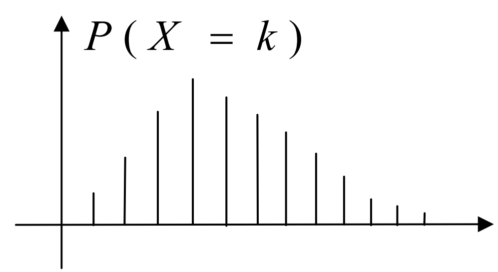


Fig. 3.18

4. Hypergeometric:

$$P(X=k)=\frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}, \qquad \max(0, m+n-N) \le k \le \min(m,n) \tag{3-46}$$

5. Geometric: $X \sim g(p)$ if

$$P(X=k)=pq^k, \quad k=0,1,2,\cdots,\infty, \qquad q=1-p. \tag{3-47}$$

6. Negative Binomial: $X \sim NB(r,p)$, if

$$P(X=k)=\binom{k-1}{r-1}p^r q^{k-r}, \qquad k=r, r+1, \cdots. \tag{3-48}$$

7. Discrete-Uniform:

$$P(X=k)=\frac{1}{N}, \quad k=1,2,\cdots,N. \tag{3-49}$$

We conclude this lecture with a general distribution due to Polya that includes both binomial and hypergeometric as special cases.

Polya's distribution: A box contains *a* white balls and *b* black balls. A ball is drawn at random, and it is replaced along with *c* balls of the same color. If X represents the number of white balls drawn in n such draws, $X = 0,\ 1,\ 2,\ \cdots,\ n,$ find the probability mass function of X.

Solution: Consider the specific sequence of draws where k white balls are first drawn, followed by $n - k$ black balls. The probability of drawing k successive white balls is given by

$$p_W = \frac{a}{a+b}\ \frac{a+c}{a+b+c}\ \frac{a+2c}{a+b+2c}\ \cdots\ \frac{a+(k-1)c}{a+b+(k-1)c} \qquad (3\text{-}50)$$

Similarly the probability of drawing k white balls followed by $n-k$ black balls is given by

$$p_k = p_w \frac{b}{a+b+kc} \frac{b+c}{a+b+(k+1)c} \cdots \frac{b+(n-k-1)c}{a+b+(n-1)c}$$

$$= \prod_{i=0}^{k-1} \frac{a+ic}{a+b+ic} \prod_{j=0}^{n-k-1} \frac{b+jc}{a+b+(j+k)c}. \qquad (3\text{-}51)$$

Interestingly, p_k in (3-51) also represents the probability of drawing k white balls and $(n-k)$ black balls in *any other specific order* (*i.e.*, The same set of numerator and denominator terms in (3-51) contribute to *all other* sequences as well.) But there are $\binom{n}{k}$ such distinct mutually exclusive sequences and summing over all of them, we obtain the Polya distribution (probability of getting k white balls in n draws) to be

$$P(X=k) = \binom{n}{k} p_k = \binom{n}{k} \prod_{i=0}^{k-1} \frac{a+ic}{a+b+ic} \prod_{j=0}^{n-k-1} \frac{b+jc}{a+b+(j+k)c}, \qquad k=0,1,2,\cdots,n. \qquad (3\text{-}52)$$

Both binomial distribution as well as the hypergeometric distribution are special cases of (3-52).

For example if draws are done with replacement, then $c = 0$ and (3-52) simplifies to the binomial distribution

$$P(X=k) = \binom{n}{k} p^k q^{n-k}, \qquad k = 0,1,2,\cdots,n \tag{3-53}$$

where

$$p = \frac{a}{a+b}, \qquad q = \frac{b}{a+b} = 1-p.$$

Similarly if the draws are conducted without replacement, Then $c = -1$ in (3-52), and it gives

$$P(X=k) = \binom{n}{k} \frac{a(a-1)(a-2)\cdots(a-k+1)}{(a+b)(a+b-1)\cdots(a+b-k+1)} \frac{b(b-1)\cdots(b-n+k+1)}{(a+b-k)\cdots(a+b-n+1)}$$

$$P(X=k)=\frac{n!}{k!(n-k)!}\ \frac{a!(a+b-k)!}{(a-k)!(a+b)!}\ \frac{b!(a+b-n)!}{(b-n+k)!(a+b-k)!}=\frac{\binom{a}{k}\binom{b}{n-k}}{\binom{a+b}{n}}$$

(3-54)

which represents the hypergeometric distribution. Finally $c=+1$ gives (replacements are doubled)

$$P(X=k)=\binom{n}{k}\frac{(a+k-1)!\ (a+b+1)!}{(a-1)!\ (a+b+k-1)!}\ \frac{(b+n-k-1)!\ (a+b+k-1)!}{(b-1)!\ (a+b+n-1)!}$$

$$=\frac{\binom{a+k-1}{k}\binom{b+n-k-1}{n-k}}{\binom{a+b+n-1}{n}}. \tag{3-55}$$

we shall refer to (3-55) as Polya's +1 distribution. the general Polya distribution in (3-52) has been used to study the spread of contagious diseases (epidemic modeling).

4. Binomial Random Variable Approximations, Conditional Probability Density Functions and Stirling's Formula

Let X represent a Binomial r.v as in (3-42). Then from (2-30)

$$P(k_1 \leq X \leq k_2) = \sum_{k=k_1}^{k_2} P_n(k) = \sum_{k=k_1}^{k_2} \binom{n}{k} p^k q^{n-k}. \tag{4-1}$$

Since the binomial coefficient $\binom{n}{k} = \frac{n!}{(n-k)!\,k!}$ grows quite rapidly with n, it is difficult to compute (4-1) for large n. In this context, two approximations are extremely useful.

4.1 **The Normal Approximation** (Demoivre-Laplace Theorem) Suppose $n \to \infty$ with p held fixed. Then for k in the $\sqrt{npq}$ neighborhood of np, we can approximate

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} e^{-(k-np)^2/2npq} . \tag{4-2}$$

Thus if k_1 and k_2 in (4-1) are within or around the neighborhood of the interval $\left(np - \sqrt{npq},\ np + \sqrt{npq}\right)$, we can approximate the summation in (4-1) by an integration. In that case (4-1) reduces to

$$P(k_1 \le X \le k_2) = \int_{k_1}^{k_2} \frac{1}{\sqrt{2\pi npq}} e^{-(x-np)^2/2npq} dx = \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy, \tag{4-3}$$

where

$$x_1 = \frac{k_1 - np}{\sqrt{npq}}, \quad x_2 = \frac{k_2 - np}{\sqrt{npq}}.$$

We can express (4-3) in terms of the normalized integral

$$erf(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-y^2/2} dy = erf(-x) \tag{4-4}$$

that has been tabulated extensively (See Table 4.1).

For example, if x_1 and x_2 are both positive ,we obtain

$$P\left(k_1 \le X \le k_2\right) = erf\,(x_2) - erf\,(x_1). \tag{4-5}$$

Example 4.1: A fair coin is tossed 5,000 times. Find the probability that the number of heads is between 2,475 to 2,525.

Solution: We need $P(2{,}475 \le X \le 2{,}525)$. Here *n* is large so that we can use the normal approximation. In this case $p = \frac{1}{2}$, so that $np = 2{,}500$ and $\sqrt{npq} \approx 35$. Since $np - \sqrt{npq} = 2{,}465$, and $np + \sqrt{npq} = 2{,}535$, the approximation is valid for $k_1 = 2{,}475$ and $k_2 = 2{,}525$. Thus

$$P\left(k_1 \le X \le k_2\right) = \int_{x_1}^{x_2} \frac{1}{2\pi} e^{-y^2/2} dy.$$

Here $x_1 = \dfrac{k_1 - np}{\sqrt{npq}} = -\dfrac{5}{7}, \quad x_2 = \dfrac{k_2 - np}{\sqrt{npq}} = \dfrac{5}{7}.$

$$\text{erf}\,(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-y^2/2}\,dy = G(x) - \frac{1}{2}$$

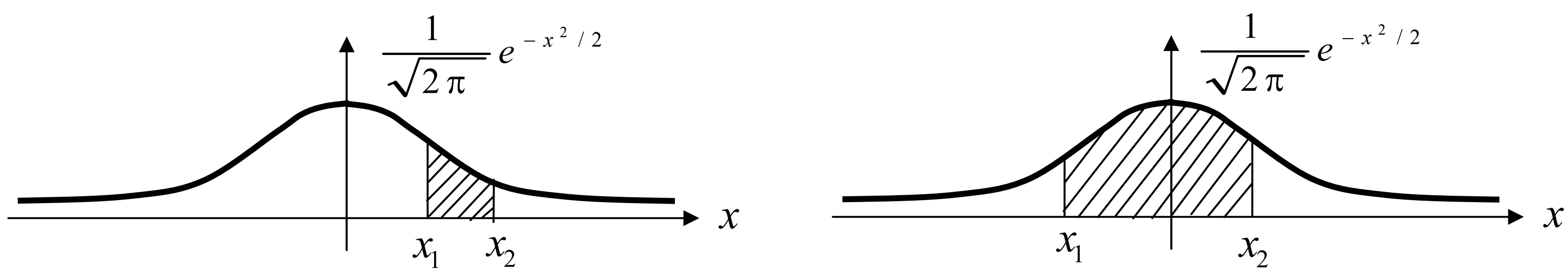
x	erf(x)	x	erf(x)	x	erf(x)	x	erf(x)
0.05	0.01994	0.80	0.28814	1.55	0.43943	2.30	0.48928
0.10	0.03983	0.85	0.30234	1.60	0.44520	2.35	0.49061
0.15	0.05962	0.90	0.31594	1.65	0.45053	2.40	0.49180
0.20	0.07926	0.95	0.32894	1.70	0.45543	2.45	0.49286
0.25	0.09871	1.00	0.34134	1.75	0.45994	2.50	0.49379
0.30	0.11791	1.05	0.35314	1.80	0.46407	2.55	0.49461
0.35	0.13683	1.10	0.36433	1.85	0.46784	2.60	0.49534
0.40	0.15542	1.15	0.37493	1.90	0.47128	2.65	0.49597
0.45	0.17364	1.20	0.38493	1.95	0.47441	2.70	0.49653
0.50	0.19146	1.25	0.39435	2.00	0.47726	2.75	0.49702
0.55	0.20884	1.30	0.40320	2.05	0.47982	2.80	0.49744
0.60	0.22575	1.35	0.41149	2.10	0.48214	2.85	0.49781
0.65	0.24215	1.40	0.41924	2.15	0.48422	2.90	0.49813
0.70	0.25804	1.45	0.42647	2.20	0.48610	2.95	0.49841
0.75	0.27337	1.50	0.43319	2.25	0.48778	3.00	0.49865

Table 4.1

Since $x_1 < 0$, from Fig. 4.1(b), the above probability is given by $P(2{,}475 \le X \le 2{,}525) = \text{erf}\,(x_2) - \text{erf}\,(x_1) = \text{erf}\,(x_2) + \text{erf}\,(|\,x_1\,|)$

$$= 2\,\text{erf}\left(\frac{5}{7}\right) = 0.516\,,$$

where we have used Table 4.1 $(\text{erf}(0.7) = 0.258)$.



(a) $x_1 > 0,\ \ x_2 > 0$ (b) $x_1 < 0,\ \ x_2 > 0$

Fig. 4.1

4.2. **The Poisson Approximation**

As we have mentioned earlier, for large *n*, the Gaussian approximation of a binomial r.v is valid only if *p* is fixed, i.e., only if $np \gg 1$ and $npq \gg 1$. what if *np* is small, or if it does not increase with *n*?

Obviously that is the case if, for example, $p \to 0$ as $n \to \infty$, such that $np = \lambda$ is a fixed number.

Many random phenomena in nature in fact follow this pattern. Total number of calls on a telephone line, claims in an insurance company etc. tend to follow this type of behavior. Consider random arrivals such as telephone calls over a line. Let n represent the total number of calls in the interval $(0, T)$. From our experience, as $T \to \infty$ we have $n \to \infty$ so that we may assume $n = \mu T$. Consider a small interval of duration Δ as in Fig. 4.2. If there is only a single call coming in, the probability p of that single call occurring in that interval must depend on its relative size with respect to T.

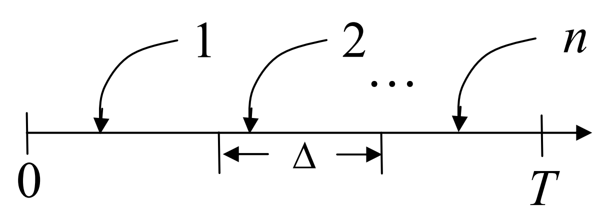


Fig. 4.2

Hence we may assume $p = \frac{\Delta}{T}.$ Note that $p \to 0$ as $T \to \infty$. However in this case $np = \mu T \cdot \frac{\Delta}{T} = \mu \Delta = \lambda$ is a constant, and the normal approximation is invalid here.

Suppose the interval Δ in Fig. 4.2 is of interest to us. A call inside that interval is a "success" (H), whereas one outside is a "failure" (T). This is equivalent to the coin tossing situation, and hence the probability $P_n(k)$ of obtaining k calls (in any order) in an interval of duration Δ is given by the binomial p.m.f. Thus

$$P_n(k) = \frac{n!}{(n-k)!\,k!} p^k (1-p)^{n-k}, \tag{4-6}$$

and here as $n \to \infty$, $p \to 0$ such that $np = \lambda$. It is easy to obtain an excellent approximation to (4-6) in that situation. To see this, rewrite (4-6) as

$$P_n(k) = \frac{n(n-1)\cdots(n-k+1)}{n^k}\frac{(np)^k}{k!}(1-np/n)^{n-k}$$
$$= \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\frac{\lambda^k}{k!}\frac{(1-\lambda/n)^n}{(1-\lambda/n)^k}. \tag{4-7}$$

Thus

$$\lim_{n\to\infty,\, p\to 0,\, np=\lambda} P_n(k) = \frac{\lambda^k}{k!}e^{-\lambda}, \tag{4-8}$$

since the finite products $\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)$ as well as $\left(1-\frac{\lambda}{n}\right)^k$ tend to unity as $n \to \infty$, and

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

The right side of (4-8) represents the Poisson p.m.f and the Poisson approximation to the binomial r.v is valid in situations where the binomial r.v parameters n and p diverge to two extremes $(n \to \infty, p \to 0)$ such that their product np is a constant.

Example 4.2: Winning a Lottery: Suppose two million lottery tickets are issued with 100 winning tickets among them. (a) If a person purchases 100 tickets, what is the probability of winning? (b) How many tickets should one buy to be 95% confident of having a winning ticket?

Solution: The probability of buying a winning ticket

$$p = \frac{\text{No. of winning tickets}}{\text{Total no. of tickets}} = \frac{100}{2\times 10^{6}} = 5\times 10^{-5}.$$

Here $n = 100$, and the number of winning tickets X in the n purchased tickets has an approximate Poisson distribution with parameter $\lambda = np = 100 \times 5 \times 10^{-5} = 0.005$. Thus

$$P(X = k) = e^{-\lambda} \frac{\lambda^{k}}{k!},$$

and (a) Probability of winning $= P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-\lambda} \approx 0.005.$

(b) In this case we need $P(X \geq 1) \geq 0.95.$

$$P(X \geq 1) = 1 - e^{-\lambda} \geq 0.95 \quad \text{implies} \quad \lambda \geq \ln 20 = 3.$$

But $\lambda = np = n \times 5 \times 10^{-5} \geq 3$ or $n \geq 60{,}000.$ Thus one needs to buy about 60,000 tickets to be 95% confident of having a winning ticket!

Example 4.3: A space craft has 100,000 components $(n \to \infty)$ The probability of any one component being defective is $2 \times 10^{-5} \, (p \to 0).$ The mission will be in danger if five or more components become defective. Find the probability of such an event.

Solution: Here *n* is large and *p* is small, and hence Poisson approximation is valid. Thus $np = \lambda = 100{,}000 \times 2 \times 10^{-5} = 2,$ and the desired probability is given by

$$P(X \geq 5) = 1 - P(X \leq 4) = 1 - \sum_{k=0}^{4} e^{-\lambda} \frac{\lambda^k}{k!} = 1 - e^{-2} \sum_{k=0}^{4} \frac{\lambda^k}{k!}$$

$$= 1 - e^{-2}\left(1 + 2 + 2 + \frac{4}{3} + \frac{2}{3}\right) = 0.052.$$

Conditional Probability Density Function

For any two events *A* and *B*, we have defined the conditional probability of *A* given *B* as

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) \neq 0. \tag{4-9}$$

Noting that the probability distribution function $F_X(x)$ is given by

$$F_X(x) = P\{ X(\xi) \leq x \}, \tag{4-10}$$

we may define the conditional distribution of the r.v *X* given the event *B* as

$$F_X(x \mid B) = P\{X(\xi) \le x \mid B\} = \frac{P\{(X(\xi) \le x) \cap B\}}{P(B)}. \quad (4\text{-}11)$$

Thus the definition of the conditional distribution depends on conditional probability, and since it obeys all probability axioms, it follows that the conditional distribution has the same properties as any distribution function. In particular

$$F_X(+\infty \mid B) = \frac{P\{(X(\xi) \le +\infty) \cap B\}}{P(B)} = \frac{P(B)}{P(B)} = 1,$$
$$F_X(-\infty \mid B) = \frac{P\{(X(\xi) \le -\infty) \cap B\}}{P(B)} = \frac{P(\phi)}{P(B)} = 0. \quad (4\text{-}12)$$

Further

$$P(x_1 < X(\xi) \le x_2 \mid B) = \frac{P\{(x_1 < X(\xi) \le x_2) \cap B\}}{P(B)}$$
$$= F_X(x_2 \mid B) - F_X(x_1 \mid B), \quad (4\text{-}13)$$

Since for $x_2 \ge x_1,$

$$\big(X(\xi) \le x_2\big) = \big(X(\xi) \le x_1\big) \cup \big(x_1 < X(\xi) \le x_2\big). \tag{4-14}$$

The conditional density function is the derivative of the conditional distribution function. Thus

$$f_X(x \mid B) = \frac{dF_X(x \mid B)}{dx}, \tag{4-15}$$

and proceeding as in (3-26) we obtain

$$F_X(x \mid B) = \int_{-\infty}^{x} f_X(u \mid B)\, du. \tag{4-16}$$

Using (4-16), we can also rewrite (4-13) as

$$P\big(x_1 < X(\xi) \le x_2 \mid B\big) = \int_{x_1}^{x_2} f_X(x \mid B)\, dx. \tag{4-17}$$

Example 4.4: Refer to example 3.2. Toss a coin and $X(T)=0$, $X(H)=1$. Suppose $B = \{H\}$. Determine $F_X(x \mid B)$.

Solution: From Example 3.2, $F_X(x)$ has the following form. We need $F_X(x \mid B)$ for all x.

For $x < 0$, $\{X(\xi) \le x\} = \phi$, so that $\{(X(\xi) \le x) \cap B\} = \phi$, and $F_X(x \mid B) = 0$.

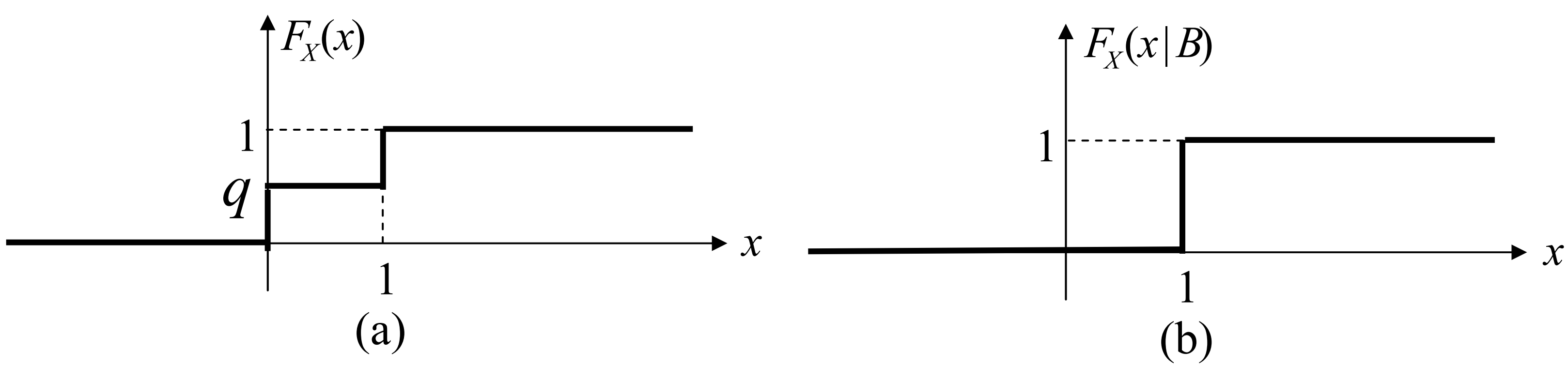


Fig. 4.3

For $0 \le x < 1, \quad \{X(\xi) \le x\} = \{T\},$ so that

$\{(X(\xi) \le x) \cap B\} = \{T\} \cap \{H\} = \phi$ and $F_X(x \mid B) = 0.$

For $x \ge 1, \quad \{X(\xi) \le x\} = \Omega,$ and

$\{(X(\xi) \le x) \cap B\} = \Omega \cap \{B\} = \{B\}$ and $F_X(x \mid B) = \dfrac{P(B)}{P(B)} = 1$

(see Fig. 4.3(b)).

Example 4.5: Given $F_X(x),$ suppose $B = \{X(\xi) \le a\}.$ Find $f_X(x \mid B).$
Solution: We will first determine $F_X(x \mid B).$ From (4-11) and B as given above, we have

$$F_X(x \mid B) = \frac{P\{(X \le x) \cap (X \le a)\}}{P(X \le a)}. \tag{4-18}$$

For $x < a,\ (X \le x) \cap (X \le a) = (X \le x)$ so that

$$F_X(x \mid B) = \frac{P(X \le x)}{P(X \le a)} = \frac{F_X(x)}{F_X(a)}. \tag{4-19}$$

For $x \ge a,\ (X \le x) \cap (X \le a) = (X \le a)$ so that $F_X(x \mid B) = 1$.

Thus

$$F_X(x \mid B) = \begin{cases} \dfrac{F_X(x)}{F_X(a)}, & x < a, \\ 1, & x \ge a, \end{cases} \tag{4-20}$$

and hence

$$f_X(x \mid B) = \frac{d}{dx} F_X(x \mid B) = \begin{cases} \dfrac{f_X(x)}{F_X(a)}, & x < a, \\ 0, & \text{otherwise.} \end{cases} \tag{4-21}$$

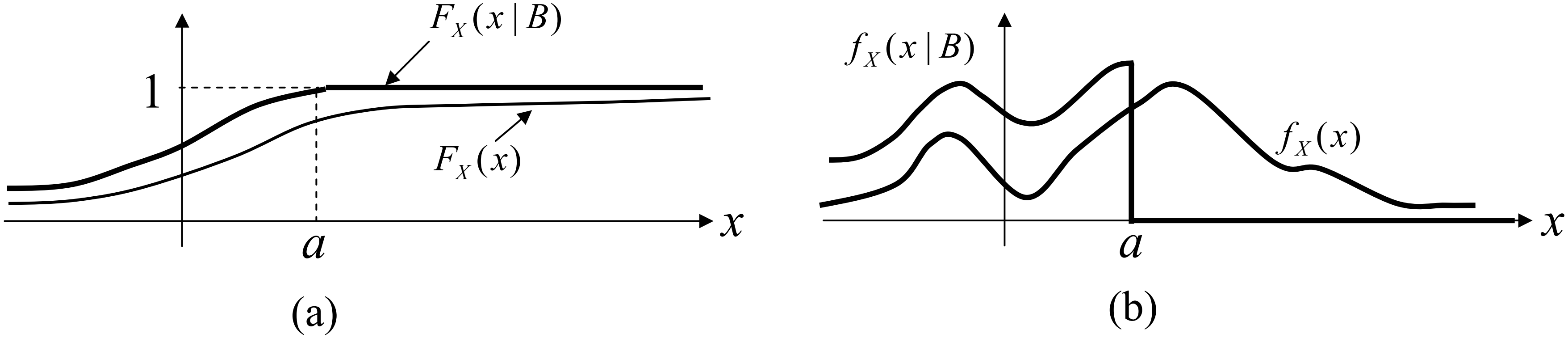


Fig. 4.4

Example 4.6: Let B represent the event $\{a < X(\xi) \le b\}$ with $b > a$. For a given $F_X(x)$, determine $F_X(x \mid B)$ and $f_X(x \mid B)$.

Solution:

$$F_X(x \mid B) = P\{X(\xi) \le x \mid B\} = \frac{P\{(X(\xi) \le x) \cap (a < X(\xi) \le b)\}}{P(a < X(\xi) \le b)}$$

$$= \frac{P\{(X(\xi) \le x) \cap (a < X(\xi) \le b)\}}{F_X(b) - F_X(a)}. \tag{4-22}$$

For $x < a$, we have $\{X(\xi) \le x\} \cap \{a < X(\xi) \le b\} = \phi$, and hence $F_X(x \mid B) = 0$. (4-23)

For $a \le x < b$, we have $\{X(\xi) \le x\} \cap \{a < X(\xi) \le b\} = \{a < X(\xi) \le x\}$ and hence

$$F_X(x \mid B) = \frac{P\big(a < X(\xi) \le x\big)}{F_X(b) - F_X(a)} = \frac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)}. \qquad (4\text{-}24)$$

For $x \ge b$, we have $\{X(\xi) \le x\} \cap \{a < X(\xi) \le b\} = \{a < X(\xi) \le b\}$ so that $F_X(x \mid B) = 1.$ (4-25)

Using (4-23)-(4-25), we get (see Fig. 4.5)

$$f_X(x \mid B) = \begin{cases} \dfrac{f_X(x)}{F_X(b) - F_X(a)}, & a < x \le b, \\ 0, & \text{otherwise.} \end{cases} \qquad (4\text{-}26)$$

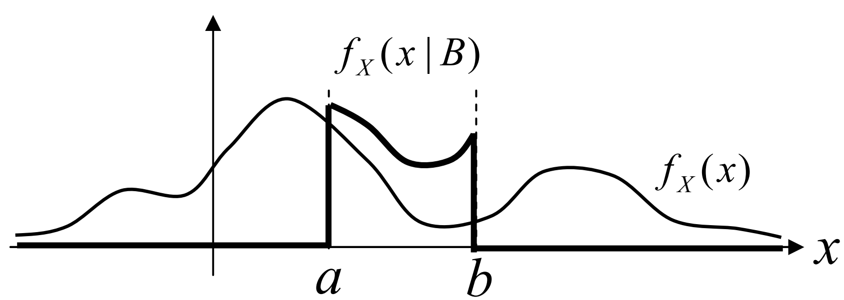


Fig. 4.5

We can use the conditional p.d.f together with the Bayes' theorem to update our a-priori knowledge about the probability of events in presence of new observations. Ideally, any new information should be used to update our knowledge. As we see in the next example, conditional p.d.f together with Bayes' theorem allow systematic updating. For any two events A and B, Bayes' theorem gives

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}. \tag{4-27}$$

Let $B = \{x_1 < X(\xi) \le x_2\}$ so that (4-27) becomes (see (4-13) and (4-17))

$$P\{A \mid (x_1 < X(\xi) \le x_2)\} = \frac{P\big((x_1 < X(\xi) \le x_2) \mid A\big)P(A)}{P\big(x_1 < X(\xi) \le x_2\big)}$$

$$= \frac{F_X(x_2 \mid A) - F_X(x_1 \mid A)}{F_X(x_2) - F_X(x_1)} P(A) = \frac{\int_{x_1}^{x_2} f_X(x \mid A)\,dx}{\int_{x_1}^{x_2} f_X(x)\,dx} P(A). \tag{4-28}$$

Further, let $x_1 = x,\ x_2 = x+\varepsilon,\ \varepsilon > 0,$ so that in the limit as $\varepsilon \to 0,$

$$\lim_{\varepsilon \to 0} P\{A \mid (x < X(\xi) \le x+\varepsilon)\} = P\big(A \mid X(\xi) = x\big) = \frac{f_X(x \mid A)}{f_X(x)} P(A). \tag{4-29}$$

or

$$f_{X|A}(x \mid A) = \frac{P(A \mid X = x) f_X(x)}{P(A)}. \tag{4-30}$$

From (4-30), we also get

$$P(A) \underbrace{\int_{-\infty}^{+\infty} f_X(x \mid A)\,dx}_{1} = \int_{-\infty}^{+\infty} P(A \mid X = x) f_X(x)\,dx, \tag{4-31}$$

or

$$P(A) = \int_{-\infty}^{+\infty} P(A \mid X = x) f_X(x)\,dx \tag{4-32}$$

and using this in (4-30), we get the desired result

$$f_{X|A}(x \mid A) = \frac{P(A \mid X = x) f_X(x)}{\int_{-\infty}^{+\infty} P(A \mid X = x) f_X(x)\,dx}. \tag{4-33}$$

To illustrate the usefulness of this formulation, let us reexamine the coin tossing problem.

Example 4.7: Let $p = P(H)$ represent the probability of obtaining a head in a toss. For a given coin, a-priori *p* can possess any value in the interval (0,1). In the absence of any additional information, we may assume the a-priori p.d.f $f_P(p)$ to be a uniform distribution in that interval. Now suppose we actually perform an experiment of tossing the coin *n* times, and *k* heads are observed. This is new information. How can we update $f_P(p)$?

Solution: Let *A*= "*k* heads in *n* specific tosses". Since these tosses result in a specific sequence,

$$P(A \mid P = p) = p^k q^{n-k}, \qquad (4\text{-}34)$$

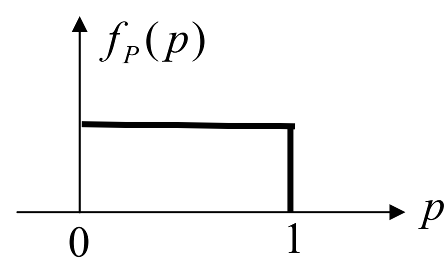


Fig.4.6

and using (4-32) we get

$$P(A)=\int_0^1 P(A\,|\,P=p)f_P(p)dp=\int_0^1 p^k(1-p)^{n-k}dp=\frac{(n-k)!\,k!}{(n+1)!}. \qquad (4\text{-}35)$$

The a-posteriori p.d.f $f_{P|A}(p\,|\,A)$ represents the updated information given the event A, and from (4-30)

$$f_{P|A}(p\,|\,A)=\frac{P(A\,|\,P=p)f_P(p)}{P(A)}$$

$$=\frac{(n+1)!}{(n-k)!\,k!}p^k q^{n-k},\quad 0<p<1 \sim \beta(n,k). \qquad (4\text{-}36)$$

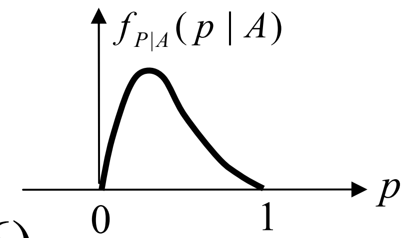


Fig. 4.7

Notice that the a-posteriori p.d.f of p in (4-36) is not a uniform distribution, but a beta distribution. We can use this a-posteriori p.d.f to make further predictions, For example, in the light of the above experiment, what can we say about the probability of a head occurring in the next $(n+1)$th toss?

Let B= "head occurring in the (n+1)th toss, given that k heads have occurred in n previous tosses".

Clearly $P(B \mid P = p) = p,$ and from (4-32)

$$P(B) = \int_0^1 P(B \mid P = p) f_P(p \mid A) dp. \tag{4-37}$$

Notice that unlike (4-32), we have used the a-posteriori p.d.f in (4-37) to reflect our knowledge about the experiment already performed. Using (4-36) in (4-37), we get

$$P(B) = \int_0^1 p \cdot \frac{(n+1)!}{(n-k)!\,k!} p^k q^{n-k} dp = \frac{k+1}{n+2}. \tag{4-38}$$

Thus, if n =10, and k = 6, then

$$P(B) = \frac{7}{12} = 0.58,$$

which is more realistic compare to p = 0.5.

To summarize, if the probability of an event X is unknown, one should make noncommittal judgement about its a-priori probability density function $f_X(x)$. Usually the uniform distribution is a reasonable assumption in the absence of any other information. Then experimental results (A) are obtained, and out knowledge about X must be updated reflecting this new information. Bayes' rule helps to obtain the a-posteriori p.d.f of X given A. From that point on, this a-posteriori p.d.f $f_{X|A}(x \mid A)$ should be used to make further predictions and calculations.

Stirling's Formula : What is it?

Stirling's formula gives an accurate approximation for $n!$ as follows:

$$n! \sim \sqrt{2\pi}\; n^{n+\frac{1}{2}} e^{-n} \qquad (4\text{-}39)$$

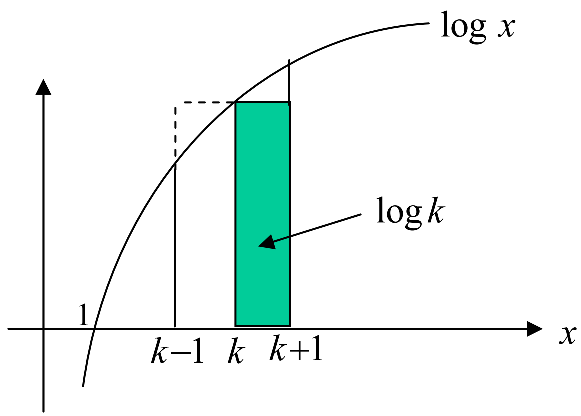
in the sense that the ratio of the two sides in (4-39) is near to one; i.e., their relative error is small, or the percentage error decreases steadily as n increases. The approximation is remarkably accurate even for small n. Thus $1! = 1$ is approximated as $\sqrt{2\pi}\,/\,e \simeq 0.9221,$ and $3! = 6$ is approximated as 5.836.

Prior to Stirling's work, DeMoivre had established the same formula in (4-39) in connection with binomial distributions in probability theory. However DeMoivre did not establish the constant term $\sqrt{2\pi}$ in (4-39); that was done by James Stirling ($\simeq$1730).

How to prove it?

We start with a simple observation: The function $\log x$ is a monotone increasing function, and hence we have

$$\int_{k-1}^{k} \log x \, dx < \log k < \int_{k}^{k+1} \log x \, dx.$$



Summing over $k = 1, 2, \cdots, n$ we get

$$\int_{0}^{n} \log x \, dx < \log n! < \int_{1}^{n+1} \log x \, dx$$

or $\left(\int \log x dx = x \log x - x\right)$

$$n \log n - n < \log n! < (n+1)\log(n+1) - n. \qquad \text{(4-40)}$$

The double inequality in (4-40) clearly suggests that $\log n!$ is close to the arithmetic mean of the two extreme numbers there. However the actual arithmetic mean is complicated and it involves several terms. Since $(n+\frac{1}{2})\log n - n$ is quite close to the above arithmetic mean, we consider the difference[1]

$$a_n \triangleq \log n! - (n+\tfrac{1}{2})\log n + n. \qquad (4\text{-}41)$$

This gives

$$\begin{aligned} a_n - a_{n+1} &= \log n! - (n+\tfrac{1}{2})\log n + n - \log(n+1)! \\ &\quad + (n+\tfrac{3}{2})\log(n+1) - (n+1) \\ &= -\log(n+1) - (n+\tfrac{1}{2})\log n + (n+\tfrac{3}{2})\log(n+1) - 1 \\ &= (n+\tfrac{1}{2})\log\tfrac{n+1}{n} - 1 = (n+\tfrac{1}{2})\log\frac{n+\frac{1}{2}+\frac{1}{2}}{n+\frac{1}{2}-\frac{1}{2}} - 1. \end{aligned}$$

[1]According to W. Feller this clever idea to use the approximate mean $(n+\frac{1}{2})\log n - n$ is due to H.E. Robbins, and it leads to an elementary proof.

Hence[1]

$$a_n - a_{n+1} = (n+\tfrac{1}{2})\log\left(\frac{1+\frac{1}{2n+1}}{1-\frac{1}{2n+1}}\right) - 1$$

$$= 2(n+\tfrac{1}{2})\left(\frac{1}{2n+1} + \frac{1}{3(2n+1)^3} + \frac{1}{5(2n+1)^5} + \cdots\right) - 1$$

$$= \frac{1}{3(2n+1)^2} + \frac{1}{5(2n+1)^4} + \frac{1}{7(2n+1)^6} + \cdots > 0 \qquad (4\text{-}42)$$

Thus $\{a_n\}$ is a monotone decreasing sequence and let c represent its limit, i.e.,

$$\lim_{n\to\infty} a_n = c \qquad (4\text{-}43)$$

From (4-41), as $n \to \infty$ this is equivalent to

$$n! \sim e^c n^{n+\frac{1}{2}} e^{-n}. \qquad (4\text{-}44)$$

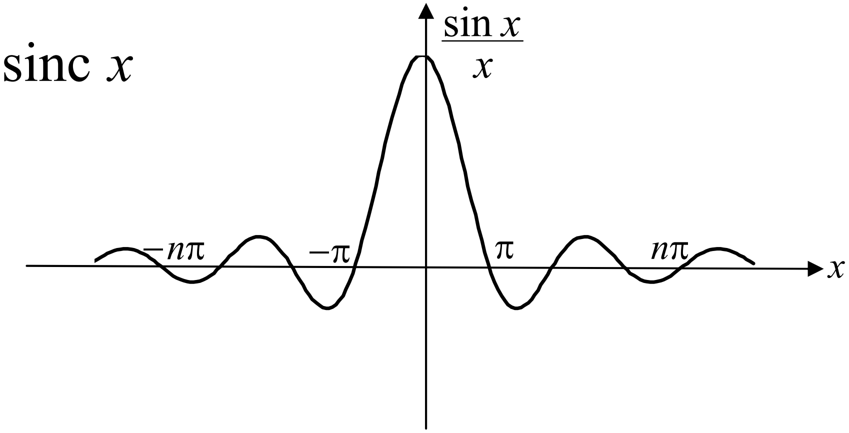
To find the constant term c in (4-44), we can make use of a formula due to Wallis ($\simeq$ 1655).

---

[1]By Taylor series expansion

$\log(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \cdots;\ \log\frac{1}{(1-x)} = x + \frac{1}{2}x^2 + \frac{1}{3}x^3 + \frac{1}{4}x^4 + \cdots$

The well known function $\frac{\sin x}{x} = \text{sinc } x$ goes to zero at $x = \pm n\pi$; moreover these are the *only* zeros of this function. Also $\frac{\sin x}{x}$ has no finite poles. (All poles are at infinity). As a result we can write



$$\frac{\sin x}{x} = (1 - \tfrac{x^2}{\pi^2})(1 - \tfrac{x^2}{4\pi^2}) \cdots (1 - \tfrac{x^2}{n^2\pi^2}) \cdots$$

or [for a proof of this formula, see chapter 4 of Dienes, *The Taylor Series*]

$$\sin x = x(1 - \tfrac{x^2}{\pi^2})(1 - \tfrac{x^2}{4\pi^2}) \cdots (1 - \tfrac{x^2}{n^2\pi^2}) \cdots,$$

which for $x = \pi/2$ gives the Wallis' formula

$$1 = \frac{\pi}{2}(1 - \tfrac{1}{2^2})(1 - \tfrac{1}{4^2}) \cdots (1 - \tfrac{1}{(2n)^2}) = \frac{\pi}{2}(\tfrac{1\cdot 3}{2^2})(\tfrac{3\cdot 5}{4^2})(\tfrac{5\cdot 7}{6^2}) \cdots (\tfrac{(2n-1)\cdot(2n+1)}{(2n)^2}) \cdots$$

or $$\frac{\pi}{2}=\prod_{n=1}^{\infty}\frac{4n^2}{(2n-1)(2n+1)}=\left(\frac{2^2}{1\cdot 3}\right)\left(\frac{4^2}{3\cdot 5}\right)\left(\frac{6^2}{5\cdot 7}\right)\cdots\left(\frac{(2n)^2}{(2n-1)\cdot(2n+1)}\right)\cdots$$

$$=\left(\frac{2\cdot 4\cdot 6\cdots 2n\cdots}{1\cdot 3\cdot 5\cdots(2n-1)\cdots}\right)^2\frac{1}{2n+1}\cdots.$$

Thus as $n\to\infty$, this gives

$$\sqrt{\pi}=\frac{2\cdot 4\cdot 6\cdots 2n}{1\cdot 3\cdot 5\cdots(2n-1)}\frac{1}{\sqrt{n+\frac{1}{2}}}\cdots$$

$$=\frac{(2\cdot 4\cdot 6\cdots 2n)^2}{(2n)!}\frac{1}{\sqrt{n+\frac{1}{2}}}\cdots=2^{2n}\frac{(n!)^2}{(2n)!}\frac{1}{\sqrt{n+\frac{1}{2}}}.$$

Thus as $n\to\infty$

$$\log\sqrt{\pi}=2n\log 2+2\log n!-\log(2n)!-\frac{1}{2}\log(n+\tfrac{1}{2}) \quad (4\text{-}45)$$

But from (4-41) and (4-43)

$$\lim_{n\to\infty}\log n!=c+\lim_{n\to\infty}\{(n+\tfrac{1}{2})\log n-n\} \quad (4\text{-}46)$$

and hence letting $n\to\infty$ in (4-45) and making use of (4-46) we get

$$\log\sqrt{\pi} = c - \tfrac{1}{2}\log 2 - \lim_{n\to\infty}\tfrac{1}{2}\log(1+\tfrac{1}{2n}) = c - \tfrac{1}{2}\log 2$$

which gives

$$e^c = \sqrt{2\pi}. \tag{4-47}$$

With (4-47) in (4-44) we obtain (4-39), and this proves the Stirling's formula.

Upper and Lower Bounds

It is possible to obtain reasonably good upper and lower bounds for $n!$ by elementary reasoning as well. To see this, note that from (4-42) we get

$$a_n - a_{n+1} < \frac{1}{3}\left(\frac{1}{(2n+1)^2} + \frac{1}{(2n+1)^4}\cdots\right)$$

$$= \frac{1}{3[(2n+1)^2 - 1]} = \frac{1}{12n(n+1)} = \frac{1}{12n} - \frac{1}{12(n+1)}$$

so that $\{a_n - 1/12n\}$ is a monotonically increasing sequence whose limit is also c. Hence for any finite n

$$a_n - \frac{1}{12n} < c \quad \Rightarrow \quad a_n < c + \frac{1}{12n}$$

and together with (4-41) and (4-47) this gives

$$n! < \sqrt{2\pi}\; n^{n+\frac{1}{2}} e^{-n\left(1-\frac{1}{12n^2}\right)} \tag{4-48}$$

Similarly from (4-42) we also have

$$a_n - a_{n+1} > \frac{1}{3(2n+1)^2} > \frac{1}{12n+1} - \frac{1}{12(n+1)+1} > 0$$

so that $\{a_n - 1/(12n+1)\}$ is a monotone decreasing sequence whose limit also equals c. Hence

$$a_n - \frac{1}{12n+1} > c \quad \Rightarrow \quad a_n > c + \frac{1}{12n+1}$$

or

$$n! > \sqrt{2\pi}\; n^{n+\frac{1}{2}} e^{-n\left(1-\frac{1}{12n(n+1)}\right)}. \tag{4-49}$$

Together with (4-48)-(4-49) we obtain

$$\sqrt{2\pi}\; n^{n+\frac{1}{2}} e^{-n} e^{\frac{1}{12n+1}} < n! < \sqrt{2\pi}\; n^{n+\frac{1}{2}} e^{-n} e^{\frac{1}{12n}}. \tag{4-50}$$

Stirling's formula also follows from the asymptotic expansion for the Gamma function given by

$$\Gamma(x) = \sqrt{2\pi}\; e^{-x} x^{x-\frac{1}{2}} \left\{ 1 + \frac{1}{12x} + \frac{1}{286x^2} - \frac{139}{51840x^3} + o\left(\frac{1}{x^4}\right) \right\} \tag{4-51}$$

Together with $\Gamma(x+1) = x\Gamma(x)$, the above expansion can be used to compute numerical values for real x. For a derivation of (4-51), one may look into Chapter 2 of the classic text by Whittaker and Watson (*Modern Analysis*).

We can use Stirling's formula to obtain yet another approximation to the binomial probability mass function. Since

$$\binom{n}{k} p^k q^{n-k} = \frac{n!}{(n-k)!\,k!} p^k q^{n-k}, \qquad \text{(4-52)}$$

using (4-50) on the right side of (4-52) we obtain

$$\binom{n}{k} p^k q^{n-k} > c_1 \sqrt{\frac{n}{2\pi(n-k)k}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}$$

and

$$\binom{n}{k} p^k q^{n-k} < c_2 \sqrt{\frac{n}{2\pi(n-k)k}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}$$

where

$$c_1 = e^{\left\{\frac{1}{12n+1} - \frac{1}{12(n-k)} - \frac{1}{12k}\right\}}$$

and

$$c_2 = e^{\left\{\frac{1}{12n} - \frac{1}{12(n-k)+1} - \frac{1}{12k+1}\right\}}.$$

Notice that the constants c_1 and c_2 are quite close to each other.

5. Functions of a Random Variable

Let X be a r.v defined on the model $(\Omega, F, P),$ and suppose $g(x)$ is a function of the variable x. Define

$$Y = g(X). \tag{5-1}$$

Is Y necessarily a r.v? If so what is its PDF $F_Y(y),$ pdf $f_Y(y)$?

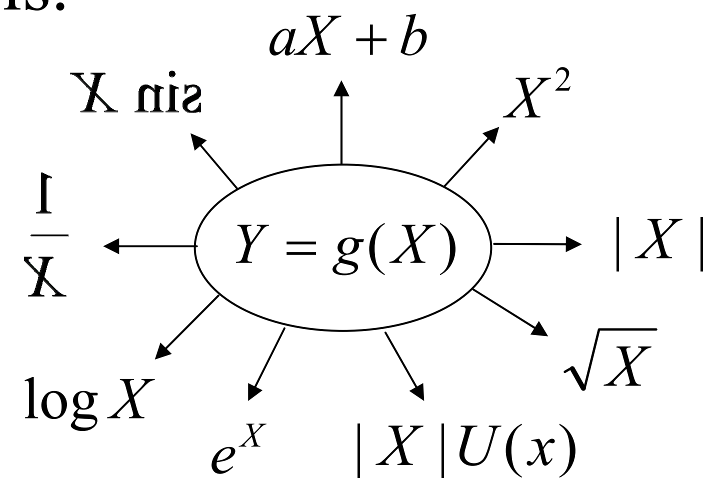
Clearly if Y is a r.v, then for every Borel set B, the set of ξ for which $Y(\xi) \in B$ must belong to F. Given that X is a r.v, this is assured if $g^{-1}(B)$ is also a Borel set, i.e., if $g(x)$ is a Borel function. In that case if X is a r.v, so is Y, and for every Borel set B

$$P(Y \in B) = P(X \in g^{-1}(B)). \tag{5-2}$$

In particular

$$F_Y(y) = P(Y(\xi) \le y) = P\big(g(X(\xi)) \le y\big) = P\big(X(\xi) \le g^{-1}(-\infty, y]\big). \quad (5\text{-}3)$$

Thus the distribution function as well of the density function of Y can be determined in terms of that of X. To obtain the distribution function of Y, we must determine the Borel set on the x-axis such that $X(\xi) \le g^{-1}(y)$ for every given y, and the probability of that set. At this point, we shall consider some of the following functions to illustrate the technical details.



Example 5.1: $Y = aX + b$ (5-4)

Solution: Suppose $a > 0$.

$$F_Y(y) = P(Y(\xi) \le y) = P(aX(\xi) + b \le y) = P\left(X(\xi) \le \frac{y-b}{a} \right) = F_X\left(\frac{y-b}{a} \right). \quad (5\text{-}5)$$

and

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a} \right). \quad (5\text{-}6)$$

On the other hand if $a < 0$, then

$$F_Y(y) = P(Y(\xi) \le y) = P(aX(\xi) + b \le y) = P\left(X(\xi) > \frac{y-b}{a} \right)$$

$$= 1 - F_X\left(\frac{y-b}{a} \right), \quad (5\text{-}7)$$

and hence

$$f_Y(y) = -\frac{1}{a} f_X\left(\frac{y-b}{a} \right). \quad (5\text{-}8)$$

From (5-6) and (5-8), we obtain (for all a)

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right). \tag{5-9}$$

Example 5.2: $Y = X^2.$ (5-10)

$$F_Y(y) = P\big(Y(\xi) \le y\big) = P\big(X^2(\xi) \le y\big). \tag{5-11}$$

If $y < 0,$ then the event $\{X^2(\xi) \le y\} = \phi$, and hence

$$F_Y(y) = 0, \quad y < 0. \tag{5-12}$$

For $y > 0,$ from Fig. 5.1, the event $\{Y(\xi) \le y\} = \{X^2(\xi) \le y\}$ is equivalent to $\{x_1 < X(\xi) \le x_2\}.$

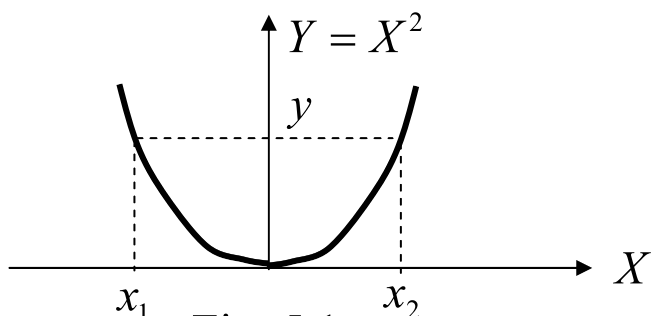


Fig. 5.1

Hence

$$F_Y(y) = P\left(x_1 < X(\xi) \le x_2\right) = F_X(x_2) - F_X(x_1)$$
$$= F_X(\sqrt{y}) - F_X(-\sqrt{y}), \quad y > 0. \tag{5-13}$$

By direct differentiation, we get

$$f_Y(y) = \begin{cases} \dfrac{1}{2\sqrt{y}}\left(f_X(\sqrt{y}) + f_X(-\sqrt{y})\right), & y > 0, \\ 0, & \text{otherwise}\;. \end{cases} \tag{5-14}$$

If $f_X(x)$ represents an even function, then (5-14) reduces to

$$f_Y(y) = \frac{1}{\sqrt{y}} f_X\left(\sqrt{y}\right) U(y). \tag{5-15}$$

In particular if $X \sim N(0,1)$, so that

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \tag{5-16}$$

and substituting this into (5-14) or (5-15), we obtain the p.d.f of $Y = X^2$ to be

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2} U(y). \tag{5-17}$$

On comparing this with (3-36), we notice that (5-17) represents a Chi-square r.v with $n = 1$, since $\Gamma(1/2) = \sqrt{\pi}$. Thus, if X is a Gaussian r.v with $\mu = 0$, then $Y = X^2$ represents a Chi-square r.v with one degree of freedom ($n = 1$).

Example 5.3: Let

$$Y = g(X) = \begin{cases} X - c, & X > c, \\ 0, & -c < X \le c, \\ X + c, & X \le -c. \end{cases}$$

In this case

$$P(Y=0) = P(-c < X(\xi) \le c) = F_X(c) - F_X(-c). \tag{5-18}$$

For $y > 0,$ we have $x > c,$ and $Y(\xi) = X(\xi) - c$ so that

$$\begin{aligned} F_Y(y) &= P\big(Y(\xi) \le y\big) = P(X(\xi) - c \le y) \\ &= P\big(X(\xi) \le y + c\big) = F_X(y+c), \quad y > 0. \end{aligned} \tag{5-19}$$

Similarly $y < 0,$ if $x < -c,$ and $Y(\xi) = X(\xi) + c$ so that

$$\begin{aligned} F_Y(y) &= P\big(Y(\xi) \le y\big) = P(X(\xi) + c \le y) \\ &= P\big(X(\xi) \le y - c\big) = F_X(y-c), \quad y < 0. \end{aligned} \tag{5-20}$$

Thus

$$f_Y(y) = \begin{cases} f_X(y+c), & y > 0, \\ [F_X(c) - F_X(-c)]\delta(y), & \\ f_X(y-c), & y < 0. \end{cases} \tag{5-21}$$

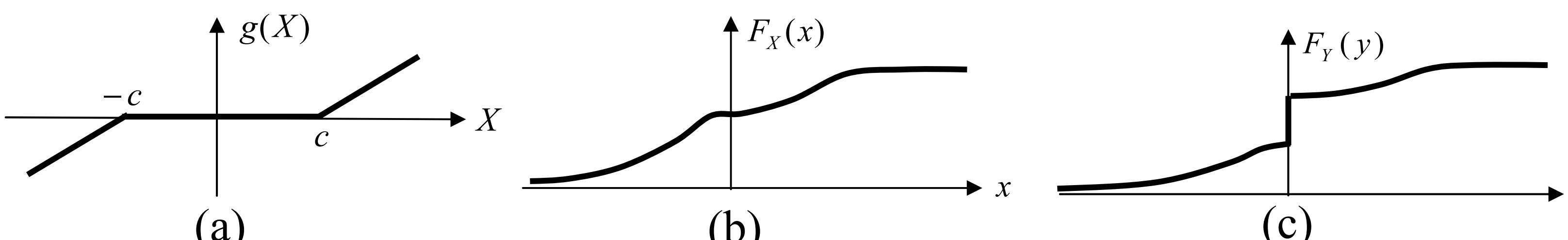


Fig. 5.2

Example 5.4: Half-wave rectifier

$$Y = g(X); \quad g(x) = \begin{cases} x, & x > 0, \\ 0, & x \le 0. \end{cases} \qquad (5\text{-}22)$$

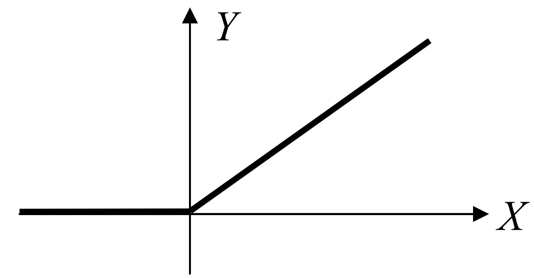


Fig. 5.3

In this case

$$P(Y = 0) = P(X(\xi) \le 0) = F_X(0). \qquad (5\text{-}23)$$

and for $y > 0$, since $Y = X$,

$$F_Y(y) = P\big(Y(\xi) \le y\big) = P\big(X(\xi) \le y\big) = F_X(y). \qquad (5\text{-}24)$$

Thus

$$f_Y(y) = \begin{cases} f_X(y), & y > 0, \\ F_X(0)\delta(y) & y = 0, \\ 0, & y < 0, \end{cases} \quad = \quad f_X(y)U(y) + F_X(0)\delta(y). \qquad (5\text{-}25)$$

Note: As a general approach, given $Y = g(X)$, first sketch the graph $y = g(x)$, and determine the range space of y. Suppose $a < y < b$ is the range space of $y = g(x)$. Then clearly for $y < a$, $F_Y(y) = 0$, and for $y > b$, $F_Y(y) = 1$, so that $F_Y(y)$ can be nonzero only in $a < y < b$. Next, determine whether there are discontinuities in the range space of y. If so evaluate $P(Y(\xi) = y_i)$ at these discontinuities. In the continuous region of y, use the basic approach

$$F_Y(y) = P\big(g(X(\xi)) \le y\big)$$

and determine appropriate events in terms of the r.v X for every y. Finally, we must have $F_Y(y)$ for $-\infty < y < +\infty$, and obtain

$$f_Y(y) = \frac{dF_Y(y)}{dy} \quad \text{in} \quad a < y < b.$$

However, if $Y = g(X)$ is a continuous function, it is easy to establish a direct procedure to obtain $f_Y(y)$. A continuos function $g(x)$ with $g'(x)$ nonzero at all but a finite number of points, has only a finite number of maxima and minima, and it eventually becomes monotonic as $|x| \to \infty$. Consider a specific y on the y-axis, and a positive increment Δy as shown in Fig. 5.4

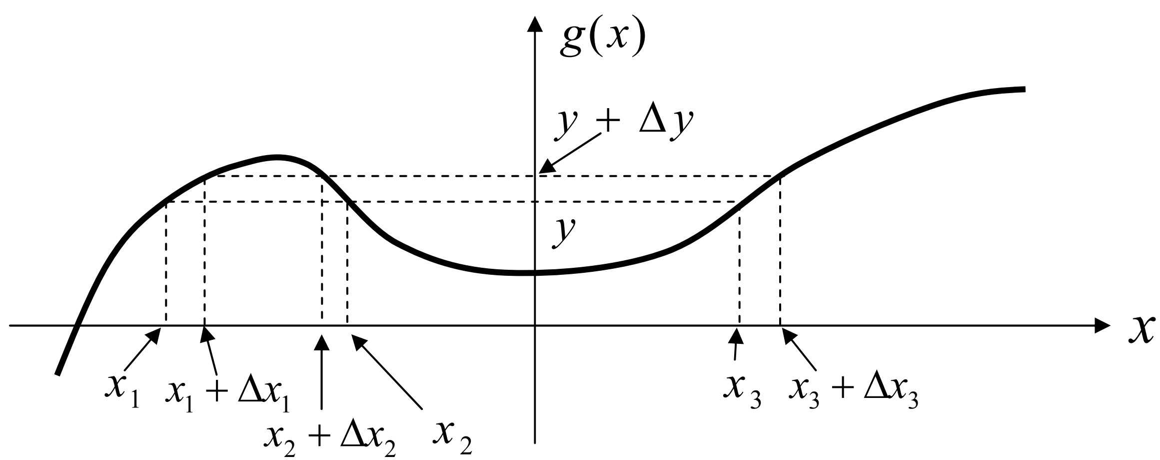


Fig. 5.4

$f_Y(y)$ for $Y = g(X)$, where $g(\cdot)$ is of continuous type.

Using (3-28) we can write

$$P\{y < Y(\xi) \le y + \Delta y\} = \int_y^{y+\Delta y} f_Y(u)du \approx f_Y(y) \cdot \Delta y. \tag{5-26}$$

But the event $\{y < Y(\xi) \le y + \Delta y\}$ can be expressed in terms of $X(\xi)$ as well. To see this, referring back to Fig. 5.4, we notice that the equation $y = g(x)$ has three solutions x_1, x_2, x_3 (for the specific y chosen there). As a result when $\{y < Y(\xi) \le y + \Delta y\}$, the r.v X could be in any one of the three mutually exclusive intervals

$$\{x_1 < X(\xi) \le x_1 + \Delta x_1\}, \ \{x_2 + \Delta x_2 < X(\xi) \le x_2\} \ \text{or} \ \{x_3 < X(\xi) \le x_3 + \Delta x_3\}.$$

Hence the probability of the event in (5-26) is the sum of the probability of the above three events, i.e.,

$$P\{y < Y(\xi) \le y + \Delta y\} = P\{x_1 < X(\xi) \le x_1 + \Delta x_1\}$$
$$+ P\{x_2 + \Delta x_2 < X(\xi) \le x_2\} + P\{x_3 < X(\xi) \le x_3 + \Delta x_3\}. \tag{5-27}$$

For small $\Delta y, \Delta x_i$, making use of the approximation in (5-26), we get

$$f_Y(y)\Delta y = f_X(x_1)\Delta x_1 + f_X(x_2)(-\Delta x_2) + f_X(x_3)\Delta x_3. \tag{5-28}$$

In this case, $\Delta x_1 > 0$, $\Delta x_2 < 0$ and $\Delta x_3 > 0$, so that (5-28) can be rewritten as

$$f_Y(y) = \sum_i f_X(x_i)\frac{|\Delta x_i|}{\Delta y} = \sum_i \frac{1}{|\Delta y / \Delta x_i|} f_X(x_i) \tag{5-29}$$

and as $\Delta y \to 0$, (5-29) can be expressed as

$$f_Y(y) = \sum_i \frac{1}{|dy/dx|_{x_i}} f_X(x_i) = \sum_i \frac{1}{|g'(x_i)|} f_X(x_i). \tag{5-30}$$

The summation index *i* in (5-30) depends on *y*, and for every *y* the equation $y = g(x_i)$ must be solved to obtain the total number of solutions at every *y*, and the actual solutions $x_1, x_2, \cdots$ all in terms of *y*.

For example, if $Y = X^2$, then for all $y > 0$, $x_1 = -\sqrt{y}$ and $x_2 = +\sqrt{y}$ represent the two solutions for each y. Notice that the solutions x_i are all in terms of y so that the right side of (5-30) is only a function of y. Referring back to the example $Y = X^2$ (Example 5.2) here for each $y > 0$, there are two solutions given by $x_1 = -\sqrt{y}$ and $x_2 = +\sqrt{y}$. ($f_Y(y) = 0$ for $y < 0$).
Moreover

$$\frac{dy}{dx} = 2x \quad \text{so that} \quad \left|\frac{dy}{dx}\right|_{x=x_i} = 2\sqrt{y}$$

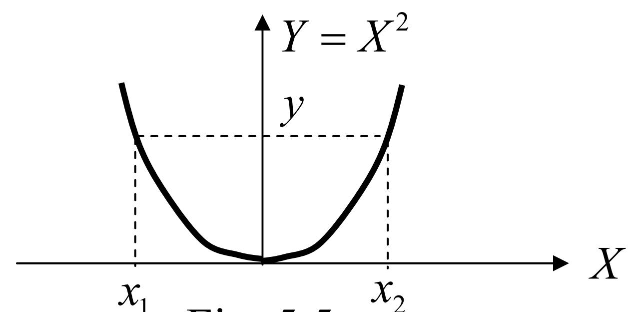


Fig. 5.5

and using (5-30) we get

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}}\left(f_X(\sqrt{y}) + f_X(-\sqrt{y})\right), & y > 0, \\ 0, & \text{otherwise}, \end{cases} \qquad (5\text{-}31)$$

which agrees with (5-14).

Example 5.5: $Y = \dfrac{1}{X}$. Find $f_Y(y)$. (5-32)

Solution: Here for every y, $x_1 = 1/y$ is the only solution, and

$$\frac{dy}{dx} = -\frac{1}{x^2} \quad \text{so that} \quad \left|\frac{dy}{dx}\right|_{x=x_1} = \frac{1}{1/y^2} = y^2,$$

and substituting this into (5-30), we obtain

$$f_Y(y) = \frac{1}{y^2} f_X\left(\frac{1}{y}\right). \tag{5-33}$$

In particular, suppose X is a Cauchy r.v as in (3-39) with parameter α so that

$$f_X(x) = \frac{\alpha/\pi}{\alpha^2 + x^2}, \quad -\infty < x < +\infty. \tag{5-34}$$

In that case from (5-33), $Y = 1/X$ has the p.d.f

$$f_Y(y) = \frac{1}{y^2}\frac{\alpha/\pi}{\alpha^2 + (1/y)^2} = \frac{(1/\alpha)/\pi}{(1/\alpha)^2 + y^2}, \quad -\infty < y < +\infty. \tag{5-35}$$

But (5-35) represents the p.d.f of a Cauchy r.v with parameter $1/\alpha$. Thus if $X \sim C(\alpha)$, then $1/X \sim C(1/\alpha)$.

Example 5.6: Suppose $f_X(x) = 2x/\pi^2, \ \ 0 < x < \pi$, and $Y = \sin X$. Determine $f_Y(y)$.

Solution: Since X has zero probability of falling outside the interval $(0, \pi)$, $y = \sin x$ has zero probability of falling outside the interval $(0,1)$. Clearly $f_Y(y) = 0$ outside this interval. For any $0 < y < 1$, from Fig.5.6(b), the equation $y = \sin x$ has an infinite number of solutions $\cdots, x_1, x_2, x_3, \cdots$, where $x_1 = \sin^{-1} y$ is the principal solution. Moreover, using the symmetry we also get $x_2 = \pi - x_1$ etc. Further,

$$\frac{dy}{dx} = \cos x = \sqrt{1 - \sin^2 x} = \sqrt{1 - y^2}$$

so that

$$\left| \frac{dy}{dx} \right|_{x = x_i} = \sqrt{1 - y^2}.$$

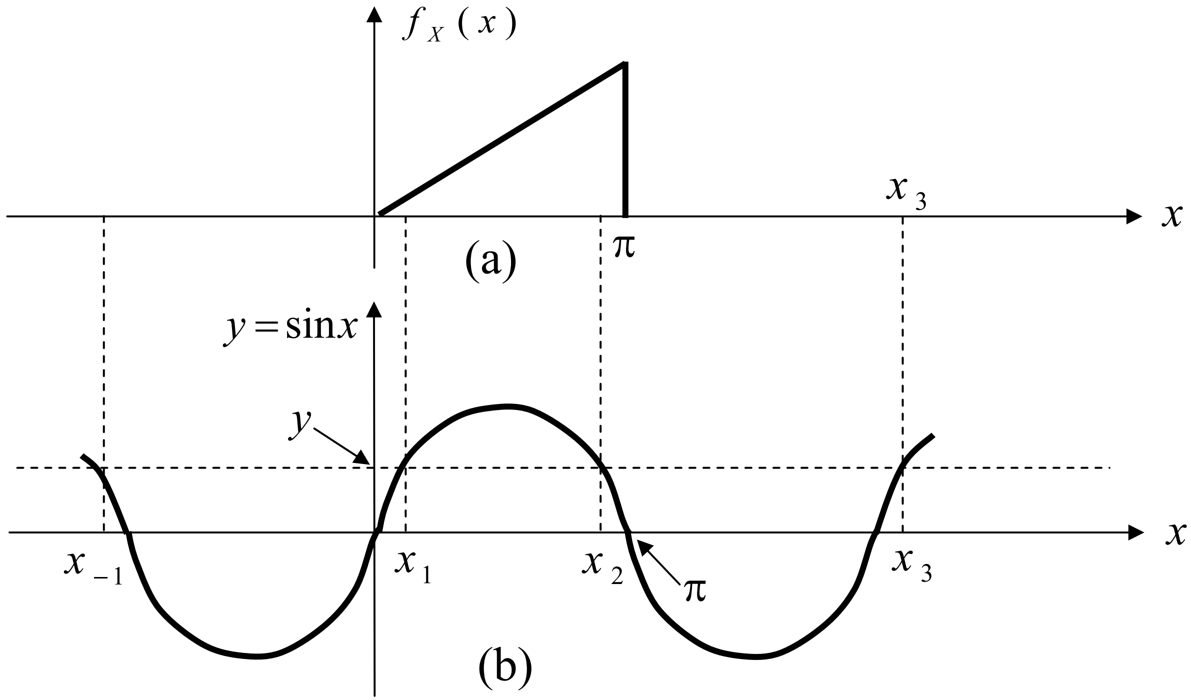


Fig. 5.6

Using this in (5-30), we obtain for $0 < y < 1$,

$$f_Y(y) = \sum_{\substack{i=-\infty \\ i \neq 0}}^{+\infty} \frac{1}{\sqrt{1-y^2}} f_X(x_i). \tag{5-36}$$

But from Fig. 5.6(a), in this case $f_X(x_{-1}) = f_X(x_3) = f_X(x_4) = \cdots = 0$ (Except for $f_X(x_1)$ and $f_X(x_2)$ the rest are all zeros).

Thus (Fig. 5.7)

$$f_Y(y) = \frac{1}{\sqrt{1-y^2}}\left(f_X(x_1) + f_X(x_2)\right) = \frac{1}{\sqrt{1-y^2}}\left(\frac{2x_1}{\pi^2} + \frac{2x_2}{\pi^2}\right)$$

$$= \frac{2(x_1 + \pi - x_1)}{\pi^2\sqrt{1-y^2}} = \begin{cases} \dfrac{2}{\pi\sqrt{1-y^2}}, & 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases} \qquad (5\text{-}37)$$

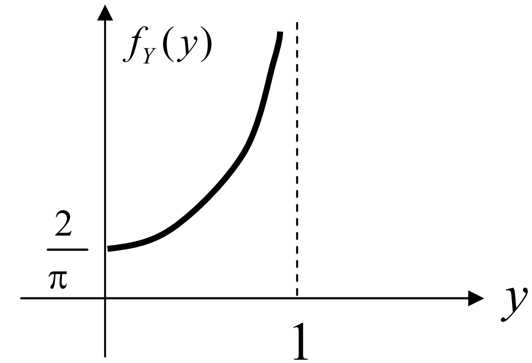


Fig. 5.7

Example 5.7: Let $Y = \tan X$ where $X \sim U(-\pi/2, \pi/2)$. Determine $f_Y(y)$.

Solution: As x moves from $(-\pi/2, \pi/2)$, y moves from $(-\infty, +\infty)$. From Fig.5.8(b), the function $Y = \tan X$ is one-to-one for $-\pi/2 < x < \pi/2$. For any y, $x_1 = \tan^{-1} y$ is the principal solution. Further

$$\frac{dy}{dx} = \frac{d \tan x}{dx} = \sec^2 x = 1 + \tan^2 x = 1 + y^2$$

so that using (5-30)

$$f_Y(y) = \frac{1}{|dy/dx|_{x=x_1}} f_X(x_1) = \frac{1/\pi}{1+y^2}, \quad -\infty < y < +\infty, \tag{5-38}$$

which represents a Cauchy density function with parameter equal to unity (Fig. 5.9).

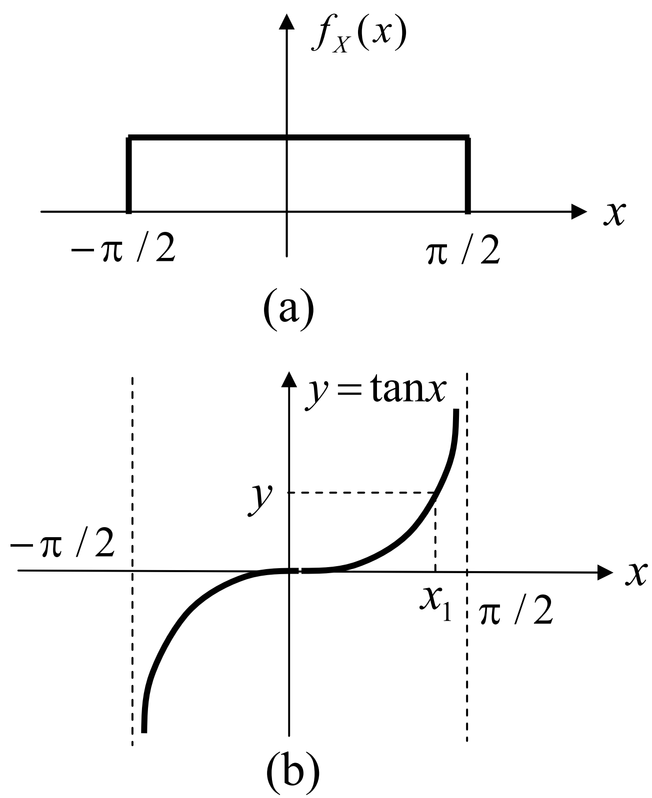


Fig. 5.8

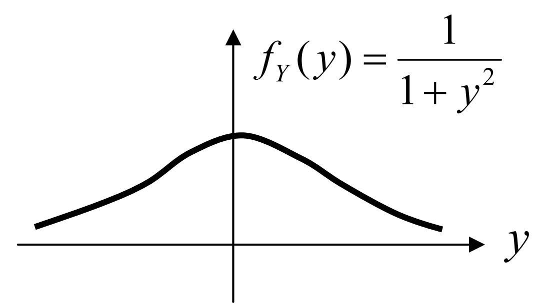


Fig. 5.9

Functions of a discrete-type r.v

Suppose X is a discrete-type r.v with

$$P(X = x_i) = p_i, \quad x = x_1, x_2, \cdots, x_i, \cdots \tag{5-39}$$

and $Y = g(X)$. Clearly Y is also of discrete-type, and when $x = x_i$, $y_i = g(x_i)$, and for those y_i

$$P(Y = y_i) = P(X = x_i) = p_i, \quad y = y_1, y_2, \cdots, y_i, \cdots \tag{5-40}$$

Example 5.8: Suppose $X \sim P(\lambda)$, so that

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0,1,2, \cdots \tag{5-41}$$

Define $Y = X^2 + 1$. Find the p.m.f of Y.
Solution: X takes the values $0,1,2, \cdots, k, \cdots$ so that Y only takes the value $1, 2, 5, \cdots, k^2 + 1, \cdots$ and

$$P(Y = k^2 + 1) = P(X = k)$$

so that for $j = k^2 + 1$

$$P(Y = j) = P\left(X = \sqrt{j-1}\right) = e^{-\lambda} \frac{\lambda^{\sqrt{j-1}}}{(\sqrt{j-1})!}, \quad j = 1, 2, 5, \cdots, k^2 + 1, \cdots. \tag{5-42}$$

6. Mean, Variance, Moments and Characteristic Functions

For a r.v X, its p.d.f $f_X(x)$ represents complete information about it, and for any Borel set B on the x-axis

$$P\big(X(\xi) \in B\big) = \int_B f_X(x)\,dx\,. \tag{6-1}$$

Note that $f_X(x)$ represents very detailed information, and quite often it is desirable to characterize the r.v in terms of its average behavior. In this context, we will introduce two parameters - mean and variance - that are universally used to represent the overall properties of the r.v and its p.d.f.

Mean or the **Expected Value** of a r.v X is defined as

$$\eta_X = \overline{X} = E(X) = \int_{-\infty}^{+\infty} x\, f_X(x)dx. \tag{6-2}$$

If X is a discrete-type r.v, then using (3-25) we get

$$\eta_X = \overline{X} = E(X) = \int x \sum_i p_i \delta(x - x_i)dx = \sum_i x_i p_i \underbrace{\int \delta(x - x_i)dx}_{1}$$

$$= \sum_i x_i p_i = \sum_i x_i\, P(X = x_i). \tag{6-3}$$

Mean represents the average (mean) value of the r.v in a very large number of trials. For example if $X \sim U(a,b)$, then using (3-31) ,

$$E(X) = \int_a^b \frac{x}{b-a}dx = \frac{1}{b-a}\left.\frac{x^2}{2}\right|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2} \tag{6-4}$$

is the midpoint of the interval (a,b).

On the other hand if X is exponential with parameter λ as in (3-32), then

$$E(X) = \int_0^\infty \frac{x}{\lambda} e^{-x/\lambda} dx = \lambda \int_0^\infty y e^{-y} dy = \lambda, \qquad (6\text{-}5)$$

implying that the parameter λ in (3-32) represents the mean value of the exponential r.v.

Similarly if X is Poisson with parameter λ as in (3-45), using (6-3), we get

$$E(X) = \sum_{k=0}^{\infty} kP(X=k) = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!}$$

$$= e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = \lambda e^{-\lambda} e^{\lambda} = \lambda. \qquad (6\text{-}6)$$

Thus the parameter λ in (3-45) also represents the mean of the Poisson r.v.

In a similar manner, if X is binomial as in (3-44), then its mean is given by

$$E(X) = \sum_{k=0}^{n} kP(X=k) = \sum_{k=0}^{n} k\binom{n}{k} p^k q^{n-k} = \sum_{k=1}^{n} k \frac{n!}{(n-k)!\,k!} p^k q^{n-k}$$

$$= \sum_{k=1}^{n} \frac{n!}{(n-k)!\,(k-1)!} p^k q^{n-k} = np \sum_{i=0}^{n-1} \frac{(n-1)!}{(n-i-1)!\,i!} p^i q^{n-i-1} = np(p+q)^{n-1} = np.$$

(6-7)

Thus np represents the mean of the binomial r.v in (3-44).

For the normal r.v in (3-29),

$$E(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} x e^{-(x-\mu)^2/2\sigma^2} dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (y+\mu) e^{-y^2/2\sigma^2} dy$$

$$= \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} y e^{-y^2/2\sigma^2} dy}_{0} + \mu \cdot \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-y^2/2\sigma^2} dy}_{1} = \mu. \qquad (6\text{-}8)$$

Thus the first parameter in $X \sim N(\mu, \sigma^2)$ is infact the mean of the Gaussian r.v X. Given $X \sim f_X(x)$, suppose $Y = g(X)$ defines a new r.v with p.d.f $f_Y(y)$. Then from the previous discussion, the new r.v Y has a mean μ_Y given by (see (6-2))

$$\mu_Y = E(Y) = \int_{-\infty}^{+\infty} y \, f_Y(y) dy. \tag{6-9}$$

From (6-9), it appears that to determine $E(Y)$, we need to determine $f_Y(y)$. However this is not the case if only $E(Y)$ is the quantity of interest. Recall that for any y, $\Delta y > 0$

$$P\left(y < Y \leq y + \Delta y\right) = \sum_i P\left(x_i < X \leq x_i + \Delta x_i\right), \tag{6-10}$$

where x_i represent the multiple solutions of the equation $y = g(x_i)$. But(6-10) can be rewritten as

$$f_Y(y)\Delta y = \sum_i f_X(x_i)\Delta x_i, \tag{6-11}$$

where the $(x_i,\ x_i+\Delta x_i)$ terms form nonoverlapping intervals. Hence

$$y\ f_Y(y)\Delta y=\sum_i y\ f_X(x_i)\Delta x_i=\sum_i g(x_i)f_X(x_i)\Delta x_i, \qquad (6\text{-}12)$$

and hence as Δy covers the entire y-axis, the corresponding Δx's are nonoverlapping, and they cover the entire x-axis. Hence, in the limit as $\Delta y\to 0$, integrating both sides of (6-12), we get the useful formula

$$E(Y)=E\big(g(X)\big)=\int_{-\infty}^{+\infty} y\ f_Y(y)dy=\int_{-\infty}^{+\infty} g(x)f_X(x)dx. \qquad (6\text{-}13)$$

In the discrete case, (6-13) reduces to

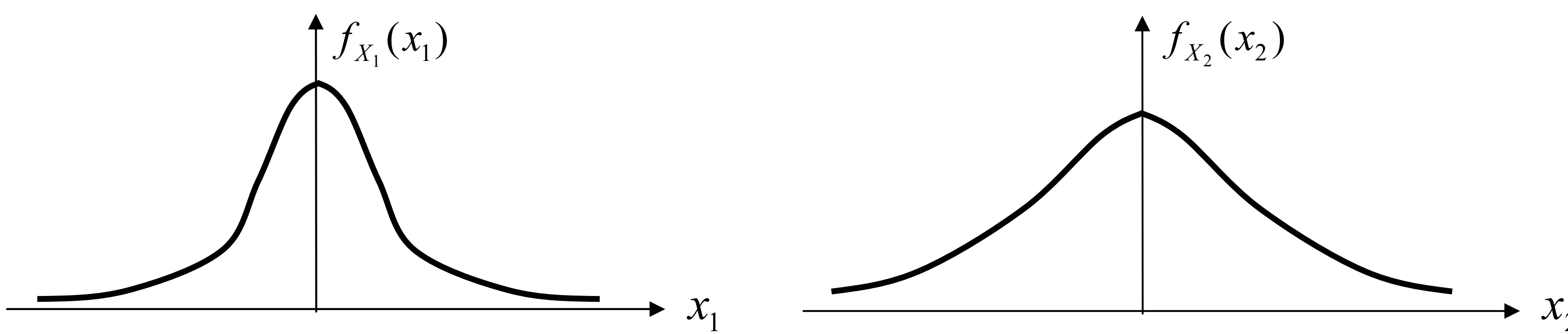
$$E(Y)=\sum_i g(x_i)P(X=x_i). \qquad (6\text{-}14)$$

From (6-13)-(6-14), $f_Y(y)$ is not required to evaluate $E(Y)$ for $Y=g(X)$. We can use (6-14) to determine the mean of $Y=X^2$, where X is a Poisson r.v. Using (3-45)

$$
\begin{aligned}
E\left(X^2\right) &= \sum_{k=0}^{\infty} k^2 P(X=k) = \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{(k-1)!} = e^{-\lambda} \sum_{i=0}^{\infty} (i+1) \frac{\lambda^{i+1}}{i!} \\
&= \lambda e^{-\lambda} \left(\sum_{i=0}^{\infty} i \frac{\lambda^i}{i!} + \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \right) = \lambda e^{-\lambda} \left(\sum_{i=1}^{\infty} i \frac{\lambda^i}{i!} + e^{\lambda} \right) \\
&= \lambda e^{-\lambda} \left(\sum_{i=1}^{\infty} \frac{\lambda^i}{(i-1)!} + e^{\lambda} \right) = \lambda e^{-\lambda} \left(\sum_{m=0}^{\infty} \frac{\lambda^{m+1}}{m!} + e^{\lambda} \right) \\
&= \lambda e^{-\lambda} \left(\lambda e^{\lambda} + e^{\lambda} \right) = \lambda^2 + \lambda. \qquad (6\text{-}15)
\end{aligned}
$$

In general, $E\left(X^k\right)$ is known as the *k*th moment of r.v X. Thus if $X \sim P(\lambda)$, its second moment is given by (6-15).

Mean alone will not be able to truly represent the p.d.f of any r.v. To illustrate this, consider the following scenario: Consider two Gaussian r.vs $X_1 \sim N(0,1)$ and $X_2 \sim N(0,10)$. Both of them have the same mean $\mu = 0.$ However, as Fig. 6.1 shows, their p.d.fs are quite different. One is more concentrated around the mean, whereas the other one (X_2) has a wider spread. Clearly, we need atleast an additional parameter to measure this spread around the mean!



(a) $\sigma^2 = 1$ (b) $\sigma^2 = 10$

Fig.6.1

For a r.v X with mean μ, $X-\mu$ represents the deviation of the r.v from its mean. Since this deviation can be either positive or negative, consider the quantity $(X-\mu)^2$, and its average value $E[(X-\mu)^2]$ represents the average mean square deviation of X around its mean. Define

$$\sigma_X^2 \stackrel{\Delta}{=} E[(X-\mu)^2] > 0. \qquad (6\text{-}16)$$

With $g(X)=(X-\mu)^2$ and using (6-13) we get

$$\sigma_X^2 = \int_{-\infty}^{+\infty}(x-\mu)^2 f_X(x)dx > 0. \qquad (6\text{-}17)$$

σ_X^2 is known as the variance of the r.v X, and its square root $\sigma_X = \sqrt{E(X-\mu)^2}$ is known as the standard deviation of X. Note that the standard deviation represents the root mean square spread of the r.v X around its mean μ.

Expanding (6-17) and using the linearity of the integrals, we get

$$
\begin{aligned}
Var(X) = \sigma_X^2 &= \int_{-\infty}^{+\infty} \left(x^2 - 2x\mu + \mu^2 \right) f_X(x)\,dx \\
&= \int_{-\infty}^{+\infty} x^2 f_X(x)\,dx - 2\mu \int_{-\infty}^{+\infty} x\, f_X(x)\,dx + \mu^2 \\
&= E\left(X^2\right) - \mu^2 = E\left(X^2\right) - \left[E(X)\right]^2 = \overline{X^2} - \overline{X}^2 . \qquad (6\text{-}18)
\end{aligned}
$$

Alternatively, we can use (6-18) to compute σ_X^2.

Thus , for example, returning back to the Poisson r.v in (3-45), using (6-6) and (6-15), we get

$$
\sigma_X^2 = \overline{X^2} - \overline{X}^2 = \left(\lambda^2 + \lambda\right) - \lambda^2 = \lambda . \qquad (6\text{-}19)
$$

Thus for a Poisson r.v, mean and variance are both equal to its parameter λ.

To determine the variance of the normal r.v $N(\mu, \sigma^2)$, we can use (6-16). Thus from (3-29)

$$Var(X) = E[(X-\mu)^2] = \int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx. \qquad (6\text{-}20)$$

To simplify (6-20), we can make use of the identity

$$\int_{-\infty}^{+\infty} f_X(x)dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx = 1$$

for a normal p.d.f. This gives

$$\int_{-\infty}^{+\infty} e^{-(x-\mu)^2/2\sigma^2} dx = \sqrt{2\pi}\,\sigma. \qquad (6\text{-}21)$$

Differentiating both sides of (6-21) with respect to σ, we get

$$\int_{-\infty}^{+\infty} \frac{(x-\mu)^2}{\sigma^3} e^{-(x-\mu)^2/2\sigma^2} dx = \sqrt{2\pi}$$

or

$$\int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx = \sigma^2, \qquad (6\text{-}22)$$

which represents the $Var(X)$ in (6-20). Thus for a normal r.v as in (3-29)

$$Var(X) = \sigma^2 \tag{6-23}$$

and the second parameter in $N(\mu, \sigma^2)$ infact represents the variance of the Gaussian r.v. As Fig. 6.1 shows the larger the σ, the larger the spread of the p.d.f around its mean. Thus as the variance of a r.v tends to zero, it will begin to concentrate more and more around the mean ultimately behaving like a constant.

Moments: As remarked earlier, in general

$$m_n = \overline{X^n} = E(X^n), \quad n \geq 1 \tag{6-24}$$

are known as the moments of the r.v *X*, and

$$\mu_n = E[(X-\mu)^n] \tag{6-25}$$

are known as the central moments of X. Clearly, the mean $\mu = m_1$, and the variance $\sigma^2 = \mu_2$. It is easy to relate m_n and μ_n. Infact

$$\mu_n = E[(X-\mu)^n] = E\left(\sum_{k=0}^{n} \binom{n}{k} X^k (-\mu)^{n-k} \right)$$

$$= \sum_{k=0}^{n} \binom{n}{k} E\left(X^k\right)(-\mu)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} \; m_k \; (-\mu)^{n-k}. \tag{6-26}$$

In general, the quantities

$$E[(X-a)^n] \tag{6-27}$$

are known as the generalized moments of X about a, and

$$E[|X|^n] \tag{6-28}$$

are known as the absolute moments of X.

For example, if $X \sim N(0, \sigma^2)$, then it can be shown that

$$E(X^n) = \begin{cases} 0, & n \text{ odd}, \\ 1 \cdot 3 \cdots (n-1)\sigma^n, & n \text{ even}. \end{cases} \tag{6-29}$$

$$E(|X|^n) = \begin{cases} 1 \cdot 3 \cdots (n-1)\sigma^n, & n \text{ even}, \\ 2^k k! \sigma^{2k+1} \sqrt{2/\pi}, & n = (2k+1), \text{ odd}. \end{cases} \tag{6-30}$$

Direct use of (6-2), (6-13) or (6-14) is often a tedious procedure to compute the mean and variance, and in this context, the notion of the characteristic function can be quite helpful.

Characteristic Function

The characteristic function of a r.v X is defined as

$$\Phi_X(\omega) \stackrel{\Delta}{=} E\left(e^{jX\omega}\right) = \int_{-\infty}^{+\infty} e^{jx\omega} f_X(x)\,dx. \tag{6-31}$$

Thus $\Phi_X(0) = 1,$ and $\left|\Phi_X(\omega)\right| \le 1$ for all $\omega.$

For discrete r.vs the characteristic function reduces to

$$\Phi_X(\omega) = \sum_k e^{jk\omega} P(X = k). \tag{6-32}$$

Thus for example, if $X \sim P(\lambda)$ as in (3-45), then its characteristic function is given by

$$\Phi_X(\omega) = \sum_{k=0}^{\infty} e^{jk\omega} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{j\omega})^k}{k!} = e^{-\lambda} e^{\lambda e^{j\omega}} = e^{\lambda(e^{j\omega}-1)}. \tag{6-33}$$

Similarly, if X is a binomial r.v as in (3-44), its characteristic function is given by

$$\Phi_X(\omega) = \sum_{k=0}^{n} e^{jk\omega} \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (pe^{j\omega})^k q^{n-k} = (pe^{j\omega} + q)^n. \tag{6-34}$$

To illustrate the usefulness of the characteristic function of a r.v in computing its moments, first it is necessary to derive the relationship between them. Towards this, from (6-31)

$$\Phi_X(\omega) = E\left(e^{jX\omega}\right) = E\left[\sum_{k=0}^{\infty} \frac{(j\omega X)^k}{k!}\right] = \sum_{k=0}^{\infty} j^k \frac{E(X^k)}{k!}\omega^k$$

$$= 1 + jE(X)\omega + j^2 \frac{E(X^2)}{2!}\omega^2 + \cdots + j^k \frac{E(X^k)}{k!}\omega^k + \cdots. \qquad (6\text{-}35)$$

Taking the first derivative of (6-35) with respect to ω, and letting it to be equal to zero, we get

$$\left.\frac{\partial \Phi_X(\omega)}{\partial \omega}\right|_{\omega=0} = jE(X) \quad \text{or} \quad E(X) = \frac{1}{j}\left.\frac{\partial \Phi_X(\omega)}{\partial \omega}\right|_{\omega=0}. \qquad (6\text{-}36)$$

Similarly, the second derivative of (6-35) gives

$$E(X^2) = \frac{1}{j^2}\left.\frac{\partial^2 \Phi_X(\omega)}{\partial \omega^2}\right|_{\omega=0}, \qquad (6\text{-}37)$$

and repeating this procedure *k* times, we obtain the *k*th moment of *X* to be

$$E(X^k) = \frac{1}{j^k} \frac{\partial^k \Phi_X(\omega)}{\partial \omega^k}\bigg|_{\omega=0}, \quad k \ge 1. \tag{6-38}$$

We can use (6-36)-(6-38) to compute the mean, variance and other higher order moments of any r.v X. For example, if $X \sim P(\lambda)$, then from (6-33)

$$\frac{\partial \Phi_X(\omega)}{\partial \omega} = e^{-\lambda} e^{\lambda e^{j\omega}} \lambda j e^{j\omega}, \tag{6-39}$$

so that from (6-36)

$$E(X) = \lambda, \tag{6-40}$$

which agrees with (6-6). Differentiating (6-39) one more time, we get

$$\frac{\partial^2 \Phi_X(\omega)}{\partial \omega^2} = e^{-\lambda}\left(e^{\lambda e^{j\omega}}(\lambda j e^{j\omega})^2 + e^{\lambda e^{j\omega}}\lambda j^2 e^{j\omega}\right), \tag{6-41}$$

so that from (6-37)

$$E(X^2) = \lambda^2 + \lambda, \tag{6-42}$$

which again agrees with (6-15). Notice that compared to the tedious calculations in (6-6) and (6-15), the efforts involved in (6-39) and (6-41) are very minimal.

We can use the characteristic function of the binomial r.v $B(n, p)$ in (6-34) to obtain its variance. Direct differentiation of (6-34) gives

$$\frac{\partial \Phi_X(\omega)}{\partial \omega} = jnpe^{j\omega}(pe^{j\omega} + q)^{n-1} \tag{6-43}$$

so that from (6-36), $E(X) = np$ as in (6-7).

One more differentiation of (6-43) yields

$$\frac{\partial^2 \Phi_X(\omega)}{\partial \omega^2} = j^2 np\left(e^{j\omega}(pe^{j\omega}+q)^{n-1} + (n-1)pe^{j2\omega}(pe^{j\omega}+q)^{n-2}\right) \qquad (6\text{-}44)$$

and using (6-37), we obtain the second moment of the binomial r.v to be

$$E(X^2) = np\left(1+(n-1)p\right) = n^2p^2 + npq. \qquad (6\text{-}45)$$

Together with (6-7), (6-18) and (6-45), we obtain the variance of the binomial r.v to be

$$\sigma_X^2 = E(X^2) - \left[E(X)\right]^2 = n^2p^2 + npq - n^2p^2 = npq. \qquad (6\text{-}46)$$

To obtain the characteristic function of the Gaussian r.v, we can make use of (6-31). Thus if $X \sim N(\mu, \sigma^2)$, then

$$\Phi_X(\omega) = \int_{-\infty}^{+\infty} e^{j\omega x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx \quad (\text{Let } x - \mu = y)$$

$$= e^{j\mu\omega} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{j\omega y} e^{-y^2/2\sigma^2} dy = e^{j\mu\omega} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-y/2\sigma^2 (y - j2\sigma^2\omega)} dy$$

$$(\text{Let } y - j\sigma^2\omega = u \text{ so that } y = u + j\sigma^2\omega)$$

$$= e^{j\mu\omega} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-(u + j\sigma^2\omega)(u - j\sigma^2\omega)/2\sigma^2} du$$

$$= e^{j\mu\omega} e^{-\sigma^2\omega^2/2} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-u^2/2\sigma^2} du = e^{(j\mu\omega - \sigma^2\omega^2/2)}. \tag{6-47}$$

Notice that the characteristic function of a Gaussian r.v itself has the "Gaussian" bell shape. Thus if $X \sim N(0, \sigma^2)$, then

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}, \tag{6-48}$$

and

$$\Phi_X(\omega) = e^{-\sigma^2\omega^2/2}. \tag{6-49}$$

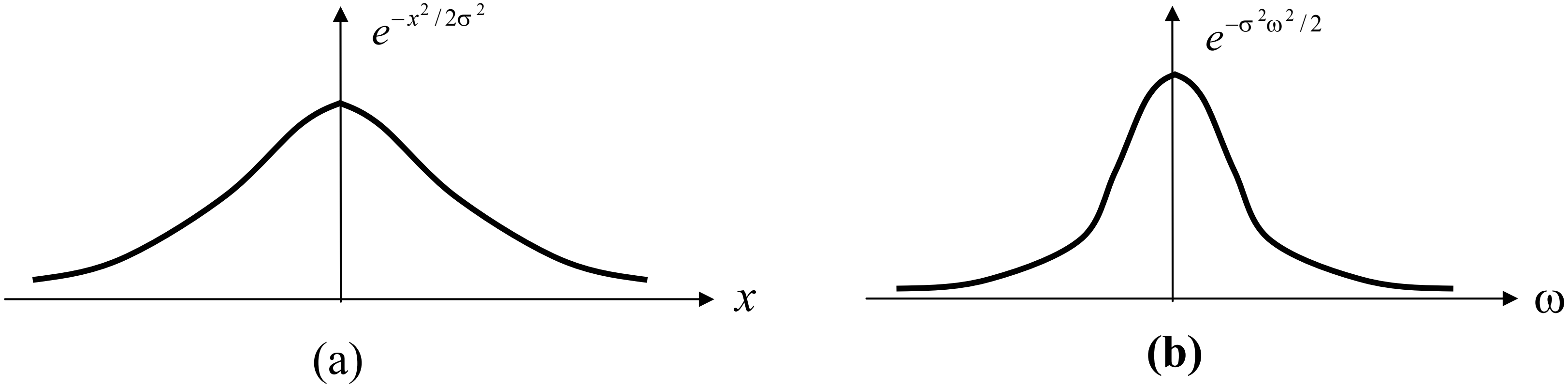


Fig. 6.2

From Fig. 6.2, the reverse roles of σ^2 in $f_X(x)$ and $\Phi_X(\omega)$ are noteworthy $(\sigma^2 \text{ vs } \frac{1}{\sigma^2})$.

In some cases, mean and variance may not exist. For example, consider the Cauchy r.v defined in (3-39). With

$f_X(x) = \frac{(\alpha/\pi)}{\alpha^2 + x^2}$,

$$E(X^2) = \frac{\alpha}{\pi}\int_{-\infty}^{+\infty} \frac{x^2}{\alpha^2 + x^2}\,dx = \frac{\alpha}{\pi}\int_{-\infty}^{+\infty} \left(1 - \frac{\alpha^2}{\alpha^2 + x^2}\right) dx = \infty, \quad (6\text{-}50)$$

clearly diverges to infinity. Similarly

$$E(X)=\frac{\alpha}{\pi}\int_{-\infty}^{+\infty}\frac{x}{\alpha^2+x^2}\,dx. \tag{6-51}$$

To compute (6-51), let us examine its one sided factor $\int_0^{+\infty}\frac{x}{\alpha^2+x^2}\,dx$. With $x=\alpha\tan\theta$

$$\int_0^{+\infty}\frac{x}{\alpha^2+x^2}\,dx=\int_0^{\pi/2}\frac{\alpha\tan\theta}{\alpha^2\sec^2\theta}\,\alpha\sec^2\theta\,d\theta=\int_0^{\pi/2}\frac{\sin\theta}{\cos\theta}d\theta$$

$$=-\int_0^{\pi/2}\frac{d(\cos\theta)}{\cos\theta}=-\log\cos\theta\Big|_0^{\pi/2}=-\log\cos\frac{\pi}{2}=\infty, \tag{6-52}$$

indicating that the double sided integral in (6-51) does not converge and is undefined. From (6-50)-(6-52), the mean and variance of a Cauchy r.v are undefined.

We conclude this section with a bound that estimates the dispersion of the r.v beyond a certain interval centered around its mean. Since σ^2 measures the dispersion of the r.v X around its mean μ, we expect this bound to depend on σ^2 as well.

Chebychev Inequality

Consider an interval of width 2ε symmetrically centered around its mean μ as in Fig. 6.3. What is the probability that X falls outside this interval? We need

$$P\left(|X-\mu| \geq \varepsilon\right) ? \tag{6-53}$$

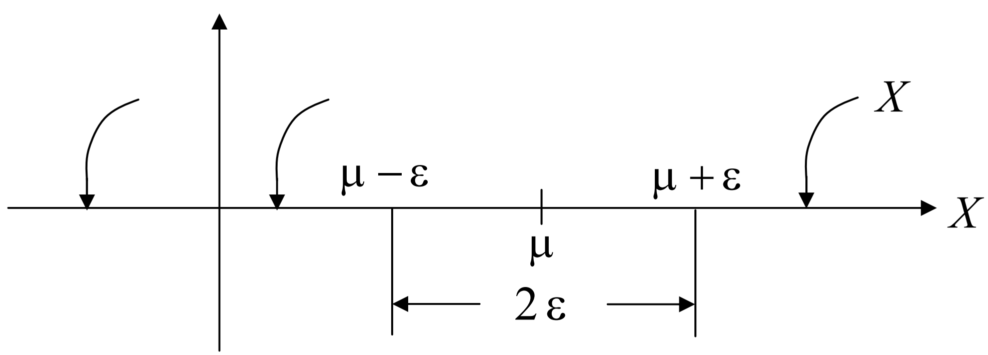


Fig. 6.3

To compute this probability, we can start with the definition of σ^2.

$$\sigma^2 = E\left[(X-\mu)^2\right] = \int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x)dx \geq \int_{|x-\mu|\geq\varepsilon} (x-\mu)^2 f_X(x)dx$$

$$\geq \int_{|x-\mu|\geq\varepsilon} \varepsilon^2 f_X(x)dx \geq \varepsilon^2 \int_{|x-\mu|\geq\varepsilon} f_X(x)dx \geq \varepsilon^2 P\left(|X-\mu| \geq \varepsilon\right). \quad (6\text{-}54)$$

From (6-54), we obtain the desired probability to be

$$P\left(|X-\mu| \geq \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2}, \quad (6\text{-}55)$$

and (6-55) is known as the chebychev inequality. Interestingly, to compute the above probability bound the knowledge of $f_X(x)$ is not necessary. We only need σ^2, the variance of the r.v. In particular with $\varepsilon = k\sigma$ in (6-55) we obtain

$$P\left(|X-\mu| \geq k\sigma\right) \leq \frac{1}{k^2}. \quad (6\text{-}56)$$

Thus with $k=3,$ we get the probability of X being outside the 3σ interval around its mean to be 0.111 for any r.v. Obviously this cannot be a tight bound as it includes all r.vs. For example, in the case of a Gaussian r.v, from Table 4.1 $(\mu = 0, \sigma = 1)$

$$P\left(| X | \geq 3\sigma \right) = 0.0027\,. \tag{6-57}$$

which is much tighter than that given by (6-56). Chebychev inequality always underestimates the exact probability.

Moment Identities :

Suppose X is a discrete random variable that takes only nonnegative integer values. i.e.,

$$P(X = k) = p_k \geq 0, \quad k = 0, 1, 2, \cdots$$

Then

$$\sum_{k=0}^{\infty} P(X > k) = \sum_{k=0}^{\infty} \sum_{i=k+1}^{\infty} P(X = i) = \sum_{i=1}^{\infty} P(X = i) \sum_{k=0}^{i-1} 1$$

$$= \sum_{i=0}^{\infty} i \, P(X = i) = E(X) \qquad \text{(6-58)}$$

similarly

$$\sum_{k=0}^{\infty} k \, P(X > k) = \sum_{i=1}^{\infty} P(X = i) \sum_{k=0}^{i-1} k = \sum_{i=1}^{\infty} \frac{i(i-1)}{2} P(X = i) = \frac{E\{X(X-1)\}}{2}$$

which gives

$$E(X^2) = \sum_{i=1}^{\infty} i^2 P(X=i) = \sum_{k=0}^{\infty} (2k+1) P(X>k). \qquad (6\text{-}59)$$

Equations (6-58) – (6-59) are at times quite useful in simplifying calculations. For example, referring to the Birthday Pairing Problem [Example 2-20., Text], let X represent the minimum number of people in a group for a birthday pair to occur. The probability that "the first n people selected from that group have different birthdays" is given by [$P(B)$ in page 39, Text]

$$p_n = \prod_{k=1}^{n-1} (1 - \frac{k}{N}) \approx e^{-n(n-1)/2N}.$$

But the event the "the first n people selected have

different birthdays" is the same as the event " $X > n$."
Hence

$$P(X > n) \approx e^{-n(n-1)/2N}.$$

Using (6-58), this gives the mean value of X to be

$$E(X) = \sum_{n=0}^{\infty} P(X > n) \approx \sum_{n=0}^{\infty} e^{-n(n-1)/2N} \approx \int_{-1/2}^{\infty} e^{-(x^2-1/4)/2N} dx$$

$$= e^{(1/8N)} \int_{-1/2}^{\infty} e^{-x^2/2N} dx = e^{(1/8N)} \left\{ \frac{1}{2}\sqrt{2\pi N} + \int_0^{1/2} e^{-x^2/2N} dx \right\}$$

$$\approx \sqrt{\pi N/2} + \frac{1}{2} = 24.44. \qquad (6\text{-}60)$$

Similarly using (6-59) we get

$$
\begin{aligned}
E(X^2) &= \sum_{n=0}^{\infty}(2n+1)P(X>n) \\
&= \sum_{n=0}^{\infty}(2n+1)e^{-n(n-1)/2N} = \int_{-1/2}^{\infty} 2\,(x+1)e^{-(x^2-1/4)/2N}dx \\
&= 2e^{(1/8N)}\left\{\int_0^{\infty} xe^{-x^2/2N}dx + \int_0^{1/2} xe^{-x^2/2N}dx\right\} + 2\int_{-1/2}^{\infty} e^{-(x^2-1/4)/2N}dx \\
&= 2\left\{\frac{\sqrt{2\pi\,N}}{2}\sqrt{\frac{2}{\pi}}\sqrt{N} + \frac{1}{8}\right\} + 2E(X) \\
&= 2N + \frac{1}{4} + \sqrt{2\pi\,N} + 1 = 2N + \sqrt{2\pi\,N} + \frac{5}{4} \\
&= 779.139.
\end{aligned}
$$

Thus

$$
Var(X) = E(X^2) - (E(X))^2 = 181.82
$$

which gives

$$\sigma_X \approx 13.48.$$

Since the standard deviation is quite high compared to the mean value, the actual number of people required for a birthday coincidence could be anywhere from 25 to 40.

Identities similar to (6-58)-(6-59) can be derived in the case of continuous random variables as well. For example, if X is a nonnegative random variable with density function $f_X(x)$ and distribution function $F_X(X)$, then

$$\begin{aligned} E\{X\} &= \int_0^\infty x f_X(x)dx = \int_0^\infty \left(\int_0^x dy \right) f_X(x)dx \\ &= \int_0^\infty \left(\int_y^\infty f_X(x)dx \right) dy = \int_0^\infty P(X > y)dy = \int_0^\infty P(X > x)dx \\ &= \int_0^\infty \{1 - F_X(x)\}dx = \int_0^\infty R(x)dx, \end{aligned} \tag{6-61}$$

where

$$R(x) = 1 - F_X(x) \geq 0, \qquad x > 0.$$

Similarly

$$E\{X^2\} = \int_0^\infty x^2 f_X(x)dx = \int_0^\infty \left(\int_0^x 2y dy \right) f_X(x)dx$$

$$= 2\int_0^\infty \left(\int_y^\infty f_X(x)dx \right) y dy$$

$$= 2\int_0^\infty x R(x)dx.$$

A Baseball Trivia (Pete Rose and Dimaggio):

In 1978 Pete Rose set a national league record by hitting a string of 44 games during a 162 game baseball season. How unusual was that event?
As we shall see, that indeed was a rare event. In that context, we will answer the following question: What is the probability that someone in major league baseball will repeat that performance and possibly set a new record in the next 50 year period? The answer will put Pete Rose's accomplishment in the proper perspective.

Solution: As example 5-32 (Text) shows consecutive successes in n trials correspond to a run of length r in n

trials. From (5-133)-(5-134) text, we get the probability of r successive hits in n games to be

$$p_n = 1 - \alpha_{n,r} + p^r \alpha_{n-r,r} \tag{6-62}$$

where

$$\alpha_{n,r} = \sum_{k=0}^{\lfloor n/(r+1) \rfloor} \binom{n-kr}{k} (-1)^k (qp^r)^k \tag{6-63}$$

and p represents the probability of a hit in a game. Pete Rose's batting average is 0.303, and on the average since a batter shows up about four times/game, we get

$$\begin{aligned} p &= P(\text{at least one hit / game}) \\ &= 1 - \text{P(no hit / game)} \\ &= 1 - (1 - 0.303)^4 = 0.76399 \end{aligned} \tag{6-64}$$

Substituting this value for p into the expressions (6-62)-(6-63) with $r = 44$ and $n = 162$, we can compute the desired probability p_n. However since n is quite large compared to r, the above formula is hopelessly time consuming in its implementation, and it is preferable to obtain a good approximation for p_n.

Towards this, notice that the corresponding moment generating function $\phi(z)$ for $q_n = 1 - p_n$ in Eq. (5-130) Text, is rational and hence it can be expanded in partial fraction as

$$\phi(z) = \frac{1 - p^r z^r}{1 - z + qp^r z^{r+1}} = \sum_{k=1}^{r} \frac{a_k}{z - z_k}, \qquad (6\text{-}65)$$

where only r roots (out of $r+1$) are accounted for, since the root $z = 1/p$ is common to both the numerator and the denominator of $\phi(z)$. Here

$$a_k = \lim_{z \to z_k} \frac{(1 - p^r z^r)(z - z_k)}{1 - z + qp^r z^{r+1}}$$

$$= \lim_{z \to z_k} \frac{(1 - p^r z^r) - rp^r z^{r-1}(z - z_k)}{-1 + (r+1)qp^r z^r}$$

or

$$a_k = \frac{p^r z_k{}^r - 1}{1 - (r+1)qp^r z_k^r}, \qquad k = 1, 2, \cdots, r \qquad \text{(6-66)}$$

From (6-65) – (6-66)

$$\phi(z) = \sum_{k=1}^{r} \frac{a_k}{(-z_k)} \, \frac{1}{1 - z / z_k} = \sum_{n=0}^{\infty} \underbrace{(\sum_{k=1}^{r} A_k z_k^{-(n+1)})}_{q_n} z^n \;\stackrel{\Delta}{=}\; \sum_{n=0}^{\infty} q_n z^n \qquad \text{(6-67)}$$

where

$$A_k = -a_k = \frac{1 - p^r z_k^r}{1 - (r+1)qp^r z_k^r}$$

and

$$q_n = 1 - p_n = \sum_{k=1}^{r} A_k \; z_k^{-(n+1)}. \tag{6-68}$$

However (fortunately), the roots $z_k, \;\; k = 1, 2, \cdots, r$ in (6-65)-(6-67) are all not of the same importance (in terms of their relative magnitude with respect to unity). Notice that since for large *n*, $z_k^{-(n+1)} \to 0$ for $|z_k| > 1,$ only the roots nearest to unity contribute to (6-68) as n becomes larger.

To examine the nature of the roots of the denominator

$$A(z) = z - 1 - qp^r z^{r+1}$$

in (6-65), note that (refer to Fig 6.1) $A(0) = -1 < 0,$ $A(1) = -qp^r > A(0), \; A(1/p) = 0, \; A(\infty) < 0$ implying that for $z \geq 0, \;\; A(z)$ increases from –1 and reaches a positive maximum at z_0 given by

$$\left.\frac{dA(z)}{dz}\right|_{z-z_0} = 1 - qp^r(r+1)z_0^r = 0,$$

which gives

$$z_0^r = \frac{1}{qp^r(r+1)}. \tag{6-69}$$

There onwards $A(z)$ decreases to $-\infty$. Thus there are two positive roots for the equation $A(z) = 0$ given by $z_1 < z_0$ and $z_2 = 1/p > 1$. Since $A(1) = -qp^r \approx 0$ but negative, by continuity z_1 has the form $z_1 = 1 + \varepsilon, \ \ \varepsilon > 0$. (see Fig 6.1)

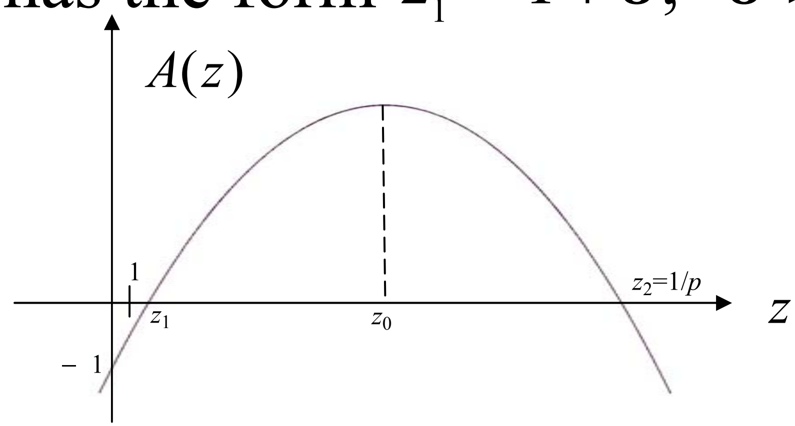


Fig 6.1 $A(z)$ for r odd

It is possible to obtain a bound for z_0 in (6-69). When P varies from 0 to 1, the maximum of $qp^r = (1-p)p^r$ is attained for $p = r/(r+1)$ and it equals $r^r/(r+1)^{r+1}$. Thus

$$qp^r \leq \frac{r^r}{(r+1)^{r+1}} \tag{6-70}$$

and hence substituting this into (6-69), we get

$$z_0 \geq \frac{r+1}{r} = 1 + \frac{1}{r}. \tag{6-71}$$

Hence it follows that the two positive roots of $A(z)$ satisfy

$$1 < z_1 < 1 + \frac{1}{r} < z_2 = \frac{1}{p} > 1. \tag{6-72}$$

Clearly, the remaining roots of $A(z)$ are complex if r is odd , and there is one negative root $-\alpha$ if r is even (see Fig 6.2). It is easy to show that the absolute value of *every* such complex or negative root is greater than $1/p > 1$.

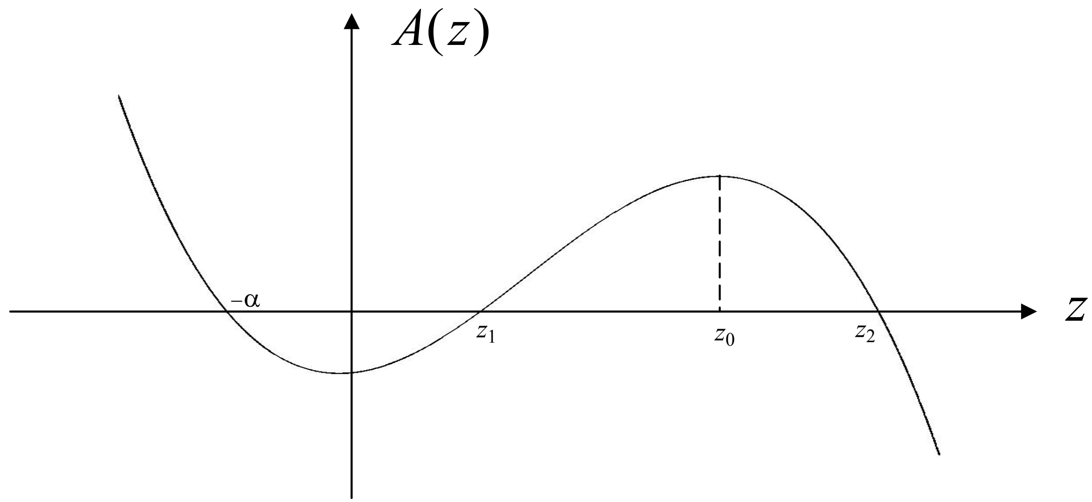


Fig 6.2 $A(z)$ for r even

To show this when r is even, suppose $-\alpha$ represents the negative root. Then

$$A(-\alpha) = -(\alpha + 1 - qp^r \alpha^{r+1}) = 0$$

so that the function

$$B(x) = x + 1 - qp^r x^{r+1} = A(x) + 2 \qquad (6\text{-}73)$$

starts positive, for $x > 0$ and increases till it reaches once again maximum at $z_0 \geq 1 + 1/r$ and then decreases to $-\infty$ through the root $x = \alpha > z_0 > 1$. Since $B(1/p) = 2$, we get $\alpha > 1/p > 1$, which proves our claim.

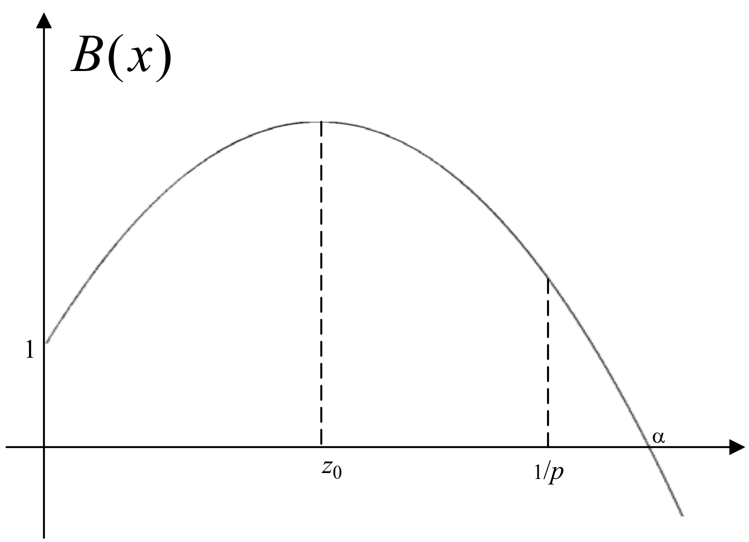


Fig 6.3 Negative root $B(\alpha) = 0$

Finally if $z = \rho\, e^{j\theta}$ is a complex root of $A(z)$, then

$$A(\rho\, e^{j\theta}) = \rho\, e^{j\theta} - 1 - qp^r \rho^{r+1} e^{j(r+1)\theta} = 0 \qquad (6\text{-}74)$$

so that

$$\rho = | 1 + qp^r \rho^{r+1} e^{j(r+1)\theta} | \le 1 + qp^r \rho^{r+1}$$

or

$$A(\rho) = \rho - 1 - qp^r \rho^{r+1} < 0.$$

Thus from (6-72), ρ belongs to either the interval $(0,\ z_1)$ or the interval $(\frac{1}{p}, \infty)$ in Fig 6.1. Moreover , by equating the imaginary parts in (6-74) we get

$$qp^r \rho^r \frac{\sin(r+1)\theta}{\sin\theta} = 1. \qquad (6\text{-}75)$$

But

$$\left|\frac{\sin(r+1)\theta}{\sin\theta}\right| \le r+1, \qquad \text{(6-76)}$$

equality being excluded if $\theta \neq 0$. Hence from (6-75)-(6-76) and (6-70)

$$(r+1)qp^r\rho^r > 1 \;\;\Rightarrow\;\; \rho^r > \frac{1}{(r+1)qp^r} = z_0^r > \left(\frac{r+1}{r}\right)^r$$

or

$$\rho > z_0 \ge 1 + \frac{1}{r}.$$

But $z_1 < z_0$. As a result ρ lies in the interval $(\frac{1}{p}, \infty)$ only. Thus

$$\rho > \frac{1}{p} > 1. \qquad \text{(6-77)}$$

To summarize the two real roots of the polynomial $A(z)$ are given by

$$z_1 = 1+\varepsilon, \quad \varepsilon > 0; \quad z_2 = \frac{1}{p} > 1, \tag{6-78}$$

and all other roots are (negative or complex) of the form

$$z_k = \rho\, e^{j\theta} \quad \text{where } \rho > \frac{1}{p} > 1. \tag{6-79}$$

Hence except for the first root z_1 (which is very close to unity), for all other roots

$$z_k^{-(n+1)} \;\rightarrow\; 0 \;\text{ rapidly for all } k.$$

As a result, the most dominant term in (6-68) is the first term, and the contributions from all other terms to q_n in (6-68) can be bounded by

$$\left|\sum_{k=2}^{r} A_k z_k^{-(n+1)}\right| \le \sum_{k=2}^{r} |A_k|\,|z_k|^{-(n+1)}$$

$$\le \sum_{k=2}^{r} \frac{1-(p\,|z_k|)^r}{1-(r+1)q(p\,|z_k|)^r}\, p^{n+1}$$

$$\le \sum_{k=2}^{r} \frac{(p\,|z_k|)^r}{(r+1)q(p\,|z_k|)^r}\, p^{n+1}$$

$$= \frac{r-1}{r+1}\frac{p^{n+1}}{q} \le \frac{p^{n+1}}{q} \;\to\; 0. \qquad (6\text{-}80)$$

Thus from (6-68), to an excellent approximation

$$q_n = A_1 z_1^{-(n+1)}. \qquad (6\text{-}81)$$

This gives the desired probability to be

$$p_n = 1 - q_n = 1 - \left(\frac{1-(pz_1)^r}{1-(r+1)q(pz_1)^r} \right) z_1^{-(n+1)}. \qquad (6\text{-}82)$$

Notice that since the dominant root z_1 is very close to unity, an excellent closed form approximation for z_1 can be obtained by considering the first order Taylor series expansion for $A(z)$. In the immediate neighborhood of $z=1$ we get

$$A(1+\varepsilon) = A(1) + A'(1)\varepsilon = -qp^r + (1-(r+1)qp^r)\varepsilon$$

so that $A(z_1) = A(1+\varepsilon) = 0$ gives

$$\varepsilon = \frac{qp^r}{1-(r+1)qp^r},$$

or

$$z_1 \approx 1 + \frac{qp^r}{1-(r+1)qp^r}. \qquad (6\text{-}83)$$

Returning back to Pete Rose's case, $p = 0.763989$, $r = 44$ gives the smallest positive root of the denominator polynomial

$$A(z) = z - 1 - qp^{44}z^{45}$$

to be

$$z_1 = 1.00000169360549.$$

(The approximation (6-83) gives $z_1 = 1.00000169360548$).
Thus with n = 162 in (6-82) we get

$$p_{162} = 0.0002069970 \qquad (6\text{-}84)$$

to be the probability for scoring 44 or more consecutive hits in 162 games for a player of Pete Rose's caliber – a very small probability indeed! In that sense it is a very rare event.

Assuming that during any baseball season there are on the average about $2 \times 25 = 50$ (?) such players over all major league baseball teams, we obtain [use Lecture #2, Eqs.(2-3)-(2-6) for the independence of 50 players]

$$P_1 = 1 - (1 - p_{162})^{50} = 0.0102975349$$

to be the probability that one of those players will hit the desired event. If we consider a period of 50 years, then the probability of *some* player hitting 44 or more consecutive games during one of these game seasons turns out to be

$$1 - (1 - P_1)^{50} = 0.40401874. \qquad (6\text{-}85)$$

(We have once again used the independence of the 50 seasons.)

Thus Pete Rose's 44 hit performance has a 60-40 chance of survival for about 50 years.From (6-85), rare events do indeed occur. In other words, *some* unlikely event is likely to happen.
However, as (6-84) shows a *particular* unlikely event – such as Pete Rose hitting 44 games in a sequence – is indeed rare.

Table 6.1 lists p_{162} for various values of r. From there, every reasonable batter should be able to hit at least 10 to 12 consecutive games during every season!

r	$p_n \quad ; \; n = 162$
44	0.000207
25	0.03928
20	0.14937
15	0.48933
10	0.95257

Table 6.1 Probability of r runs in n trials for p=0.76399.

As baseball fans well know, Dimaggio holds the record of consecutive game hitting streak at 56 games (1941). With a lifetime batting average of 0.325 for Dimaggio, the above calculations yield [use (6-64), (6-82)-(6-83)] the probability for that event to be

$$p_n = 0.0000504532. \tag{6-86}$$

Even over a 100 year period, with an average of 50 excellent hitters / season, the probability is only

$$1-(1-P_0)^{100} = 0.2229669 \tag{6-87}$$

(where $P_0 = 1-(1-p_n)^{50} = 0.00251954$) that someone will repeat or outdo Dimaggio's performance.Remember, 60 years have already passed by, and no one has done it yet!

7. Two Random Variables

In many experiments, the observations are expressible not as a single quantity, but as a family of quantities. For example to record the height and weight of each person in a community or the number of people and the total income in a family, we need two numbers.

Let X and Y denote two random variables (r.v) based on a probability model (Ω, F, P). Then

$$P\left(x_1 < X(\xi) \leq x_2\right) = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)\,dx,$$

and

$$P\left(y_1 < Y(\xi) \leq y_2\right) = F_Y(y_2) - F_Y(y_1) = \int_{y_1}^{y_2} f_Y(y)\,dy.$$

What about the probability that the pair of r.vs (X,Y) belongs to an arbitrary region D? In other words, how does one estimate, for example, $P[(x_1 < X(\xi) \le x_2) \cap (y_1 < Y(\xi) \le y_2)] = ?$ Towards this, we define the joint probability distribution function of X and Y to be

$$\begin{aligned} F_{XY}(x, y) &= P[(X(\xi) \le x) \cap (Y(\xi) \le y)] \\ &= P(X \le x, Y \le y) \ge 0, \end{aligned} \qquad (7\text{-}1)$$

where x and y are arbitrary real numbers.

Properties

(i) $F_{XY}(-\infty, y) = F_{XY}(x, -\infty) = 0, \quad F_{XY}(+\infty, +\infty) = 1.$ (7-2)

since $(X(\xi) \le -\infty, Y(\xi) \le y) \subset (X(\xi) \le -\infty),$ we get

$F_{XY}(-\infty, y) \le P\big(X(\xi) \le -\infty\big) = 0.$ Similarly $\big(X(\xi) \le +\infty, Y(\xi) \le +\infty\big) = \Omega,$

we get $F_{XY}(\infty, \infty) = P(\Omega) = 1.$

(ii) $P\big(x_1 < X(\xi) \le x_2, Y(\xi) \le y\big) = F_{XY}(x_2, y) - F_{XY}(x_1, y).$ (7-3)

$P\big(X(\xi) \le x, y_1 < Y(\xi) \le y_2\big) = F_{XY}(x, y_2) - F_{XY}(x, y_1).$ (7-4)

To prove (7-3), we note that for $x_2 > x_1,$

$$\big(X(\xi) \le x_2, Y(\xi) \le y\big) = \big(X(\xi) \le x_1, Y(\xi) \le y\big) \cup \big(x_1 < X(\xi) \le x_2, Y(\xi) \le y\big)$$

and the mutually exclusive property of the events on the right side gives

$$P\big(X(\xi) \le x_2, Y(\xi) \le y\big) = P\big(X(\xi) \le x_1, Y(\xi) \le y\big) + P\big(x_1 < X(\xi) \le x_2, Y(\xi) \le y\big)$$

which proves (7-3). Similarly (7-4) follows.

(iii) $P\left(x_1 < X(\xi) \le x_2, y_1 < Y(\xi) \le y_2\right) = F_{XY}(x_2, y_2) - F_{XY}(x_2, y_1)$
$- F_{XY}(x_1, y_2) + F_{XY}(x_1, y_1).$ (7-5)

This is the probability that (X,Y) belongs to the rectangle R_0 in Fig. 7.1. To prove (7-5), we can make use of the following identity involving mutually exclusive events on the right side.

$$\left(x_1 < X(\xi) \le x_2, Y(\xi) \le y_2\right) = \left(x_1 < X(\xi) \le x_2, Y(\xi) \le y_1\right) \cup \left(x_1 < X(\xi) \le x_2, y_1 < Y(\xi) \le y_2\right).$$

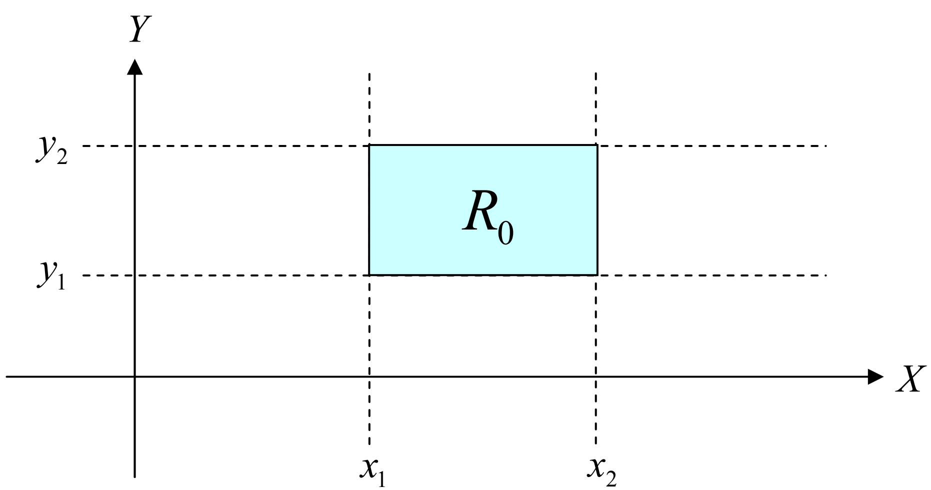


Fig. 7.1

This gives

$$P\left(x_1 < X(\xi) \le x_2, Y(\xi) \le y_2\right) = P\left(x_1 < X(\xi) \le x_2, Y(\xi) \le y_1\right) + P\left(x_1 < X(\xi) \le x_2, y_1 < Y(\xi) \le y_2\right)$$

and the desired result in (7-5) follows by making use of (7-3) with $y = y_2$ and y_1 respectively.

Joint probability density function (Joint p.d.f)

By definition, the joint p.d.f of X and Y is given by

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x\, \partial y}. \qquad (7\text{-}6)$$

and hence we obtain the useful formula

$$F_{XY}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(u, v)\, du dv\ . \qquad (7\text{-}7)$$

Using (7-2), we also get

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y)\, dx dy \ = 1. \qquad (7\text{-}8)$$

To find the probability that (X,Y) belongs to an arbitrary region D, we can make use of (7-5) and (7-7). From (7-5) and (7-7)

$$
\begin{aligned}
P\big(x < X(\xi) \le x+\Delta x,\, y < Y(\xi) \le y+\Delta y\big) &= F_{XY}(x+\Delta x, y+\Delta y) \\
&\quad - F_{XY}(x, y+\Delta y) - F_{XY}(x+\Delta x, y) + F_{XY}(x,y) \\
&= \int_x^{x+\Delta x}\int_y^{y+\Delta y} f_{XY}(u,v)\,du\,dv = f_{XY}(x,y)\Delta x \Delta y.
\end{aligned}
\tag{7-9}
$$

Thus the probability that (X,Y) belongs to a differential rectangle $\Delta x\ \Delta y$ equals $f_{XY}(x,y)\cdot\Delta x\Delta y$, and repeating this procedure over the union of no overlapping differential rectangles in D, we get the useful result

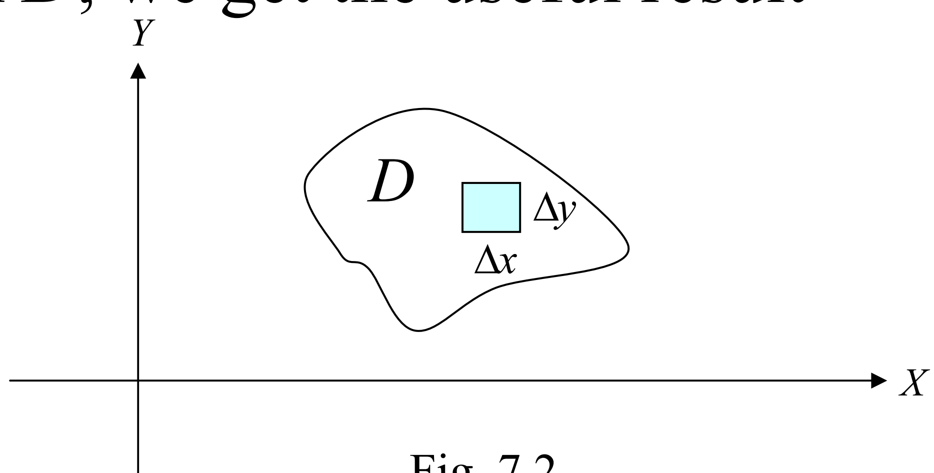


Fig. 7.2

$$P\big((X,Y)\in D\big)=\int\int_{(x,y)\in D} f_{XY}(x,y)\,dx\,dy\ . \tag{7-10}$$

(iv) Marginal Statistics

In the context of several r.vs, the statistics of each individual ones are called marginal statistics. Thus $F_X(x)$ is the marginal probability distribution function of X, and $f_X(x)$ is the marginal p.d.f of X. It is interesting to note that all marginals can be obtained from the joint p.d.f. In fact

$$F_X(x)=F_{XY}(x,+\infty),\quad F_Y(y)=F_{XY}(+\infty,y). \tag{7-11}$$

Also

$$f_X(x)=\int_{-\infty}^{+\infty} f_{XY}(x,y)\,dy,\quad f_Y(y)=\int_{-\infty}^{+\infty} f_{XY}(x,y)\,dx. \tag{7-12}$$

To prove (7-11), we can make use of the identity

$$(X\le x)=(X\le x)\cap(Y\le+\infty)$$

so that $F_X(x) = P(X \le x) = P(X \le x, Y \le \infty) = F_{XY}(x, +\infty).$

To prove (7-12), we can make use of (7-7) and (7-11), which gives

$$F_X(x) = F_{XY}(x, +\infty) = \int_{-\infty}^{x} \int_{-\infty}^{+\infty} f_{XY}(u, y)\, du dy \tag{7-13}$$

and taking derivative with respect to x in (7-13), we get

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy. \tag{7-14}$$

At this point, it is useful to know the formula for differentiation under integrals. Let

$$H(x) = \int_{a(x)}^{b(x)} h(x, y) dy. \tag{7-15}$$

Then its derivative with respect to x is given by

$$\frac{dH(x)}{dx} = \frac{db(x)}{dx} h(x, b) - \frac{da(x)}{dx} h(x, a) + \int_{a(x)}^{b(x)} \frac{\partial h(x, y)}{\partial x} dy. \tag{7-16}$$

Obvious use of (7-16) in (7-13) gives (7-14).

If X and Y are discrete r.vs, then $p_{ij} \stackrel{\Delta}{=} P(X = x_i, Y = y_j)$ represents their joint p.d.f, and their respective marginal p.d.fs are given by

$$P(X = x_i) = \sum_j P(X = x_i, Y = y_j) = \sum_j p_{ij} \tag{7-17}$$

and

$$P(Y = y_j) = \sum_i P(X = x_i, Y = y_j) = \sum_i p_{ij} \tag{7-18}$$

Assuming that $P(X = x_i, Y = y_j)$ is written out in the form of a rectangular array, to obtain $P(X = x_i)$, from (7-17), one need to add up all entries in the i-th row.

It used to be a practice for insurance companies routinely to scribble out these sum values in the left and top margins, thus suggesting the name marginal densities! (Fig 7.3).

$$\begin{array}{c|cccccc}
 & & & \sum_i p_{ij} & & & \\
\hline
 & p_{11} & p_{12} & \cdots & p_{1j} & \cdots & p_{1n} \\
 & p_{21} & p_{22} & \cdots & p_{2j} & \cdots & p_{2n} \\
 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\sum_j p_{ij} & p_{i1} & p_{i2} & \cdots & p_{ij} & \cdots & p_{in} \\
 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 & p_{m1} & p_{m2} & \cdots & p_{mj} & \cdots & p_{mn}
\end{array}$$

Fig. 7.3

From (7-11) and (7-12), the joint P.D.F and/or the joint p.d.f represent complete information about the r.vs, and their marginal p.d.fs can be evaluated from the joint p.d.f. However, given marginals, (most often) it will not be possible to compute the joint p.d.f. Consider the following example:

Example 7.1: Given

$$f_{XY}(x,y) = \begin{cases} \text{constant}, & 0 < x < y < 1, \\ 0, & \text{otherwise} \end{cases}. \tag{7-19}$$

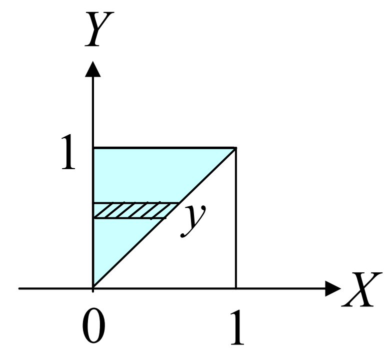


Fig. 7.4

Obtain the marginal p.d.fs $f_X(x)$ and $f_Y(y)$.

Solution: It is given that the joint p.d.f $f_{XY}(x,y)$ is a constant in the shaded region in Fig. 7.4. We can use (7-8) to determine that constant c. From (7-8)

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}(x,y)\,dxdy = \int_{y=0}^{1}\left(\int_{x=0}^{y} c\cdot dx\right)dy = \int_{y=0}^{1} cy\,dy = \frac{cy^2}{2}\Bigg|_0^1 = \frac{c}{2} = 1. \tag{7-20}$$

Thus $c = 2$. Moreover from (7-14)

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y)dy = \int_{y=x}^{1} 2dy = 2(1-x), \quad 0 < x < 1, \tag{7-21}$$

and similarly

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y)dx = \int_{x=0}^{y} 2dx = 2y, \quad 0 < y < 1. \tag{7-22}$$

Clearly, in this case given $f_X(x)$ and $f_Y(y)$ as in (7-21)-(7-22), it will not be possible to obtain the original joint p.d.f in (7-19).

Example 7.2: X and Y are said to be jointly normal (Gaussian) distributed, if their joint p.d.f has the following form:

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)}, \tag{7-23}$$

$$-\infty < x < +\infty, \; -\infty < y < +\infty, \; |\rho| < 1.$$

By direct integration, using (7-14) and completing the square in (7-23), it can be shown that

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dy = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-(x-\mu_X)^2/2\sigma_X^2} \sim N(\mu_X, \sigma_X^2), \tag{7-24}$$

and similarly

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dx = \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-(y-\mu_Y)^2/2\sigma_Y^2} \sim N(\mu_Y, \sigma_Y^2), \tag{7-25}$$

Following the above notation, we will denote (7-23) as $N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$. Once again, knowing the marginals in (7-24) and (7-25) alone doesn't tell us everything about the joint p.d.f in (7-23).

As we show below, the only situation where the marginal p.d.fs can be used to recover the joint p.d.f is when the random variables are statistically independent.

Independence of r.vs

Definition: The random variables X and Y are said to be statistically independent if the events $\{X(\xi) \in A\}$ and $\{Y(\xi) \in B\}$ are independent events for any two Borel sets A and B in x and y axes respectively. Applying the above definition to the events $\{X(\xi) \leq x\}$ and $\{Y(\xi) \leq y\}$, we conclude that, if the r.vs X and Y are independent, then

$$P\big((X(\xi) \leq x) \cap (Y(\xi) \leq y)\big) = P(X(\xi) \leq x)P(Y(\xi) \leq y) \tag{7-26}$$

i.e.,

$$F_{XY}(x, y) = F_X(x)F_Y(y) \tag{7-27}$$

or equivalently, if X and Y are independent, then we must have

$$f_{XY}(x, y) = f_X(x)f_Y(y). \tag{7-28}$$

If X and Y are discrete-type r.vs then their independence implies

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j) \quad \text{for all } \ i, j. \tag{7-29}$$

Equations (7-26)-(7-29) give us the procedure to test for independence. Given $f_{XY}(x,y)$, obtain the marginal p.d.fs $f_X(x)$ and $f_Y(y)$ and examine whether (7-28) or (7-29) is valid. If so, the r.vs are independent, otherwise they are dependent. Returning back to Example 7.1, from (7-19)-(7-22), we observe by direct verification that $f_{XY}(x,y) \neq f_X(x) f_Y(y)$. Hence X and Y are dependent r.vs in that case. It is easy to see that such is the case in the case of Example 7.2 also, unless $\rho = 0.$ In other words, two jointly Gaussian r.vs as in (7-23) are independent if and only if the fifth parameter $\rho = 0.$

Example 7.3: Given

$$f_{XY}(x,y)=\begin{cases} xy^2e^{-y}, & 0<y<\infty, \quad 0<x<1, \\ 0, & \text{otherwise.} \end{cases} \tag{7-30}$$

Determine whether X and Y are independent.

Solution:

$$\begin{aligned} f_X(x) &= \int_0^{+\infty} f_{XY}(x,y)dy = x\int_0^{\infty} y^2e^{-y}dy \\ &= x\left(-2ye^{-y}\Big|_0^{\infty} + 2\int_0^{\infty} ye^{-y}dy\right) = 2x, \quad 0<x<1. \end{aligned} \tag{7-31}$$

Similarly

$$f_Y(y) = \int_0^1 f_{XY}(x,y)dx = \frac{y^2}{2}e^{-y}, \quad 0<y<\infty. \tag{7-32}$$

In this case

$$f_{XY}(x,y) = f_X(x)f_Y(y),$$

and hence X and Y are independent random variables.

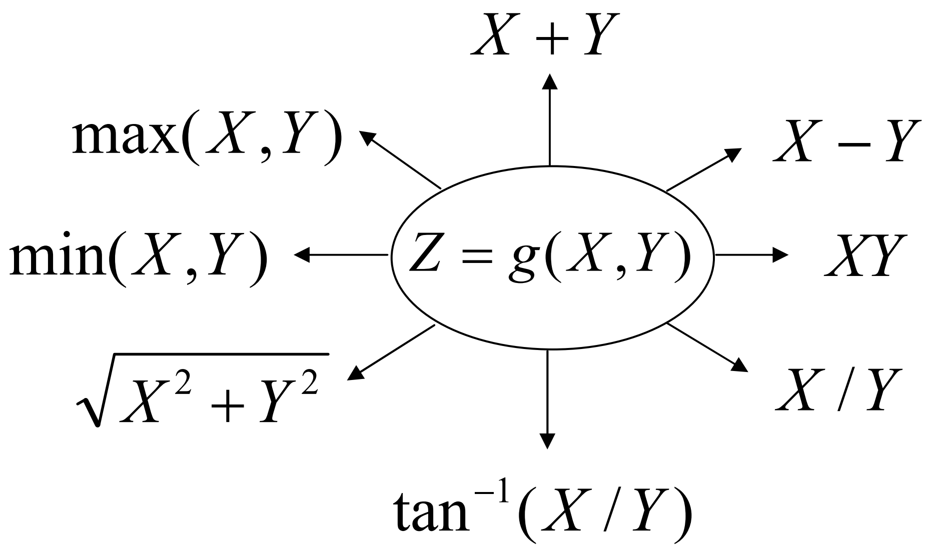
8. One Function of Two Random Variables

Given two random variables X and Y and a function $g(x,y)$, we form a new random variable Z as

$$Z = g(X, Y). \tag{8-1}$$

Given the joint p.d.f $f_{XY}(x, y)$, how does one obtain $f_Z(z)$, the p.d.f of Z ? Problems of this type are of interest from a practical standpoint. For example, a receiver output signal usually consists of the desired signal buried in noise, and the above formulation in that case reduces to $Z = X + Y$.

It is important to know the statistics of the incoming signal for proper receiver design. In this context, we shall analyze problems of the following type:



(8-2)

Referring back to (8-1), to start with

$$F_Z(z) = P\big(Z(\xi) \le z\big) = P\big(g(X,Y) \le z\big) = P\big[(X,Y) \in D_z\big]$$

$$= \int\int_{x,y\in D_z} f_{XY}(x,y)\,dxdy\,, \qquad (8\text{-}3)$$

where D_z in the XY plane represents the region such that $g(x, y) \le z$ is satisfied. Note that D_z need not be simply connected (Fig. 8.1). From (8-3), to determine $F_Z(z)$ it is enough to find the region D_z for every z, and then evaluate the integral there.

We shall illustrate this method through various examples.

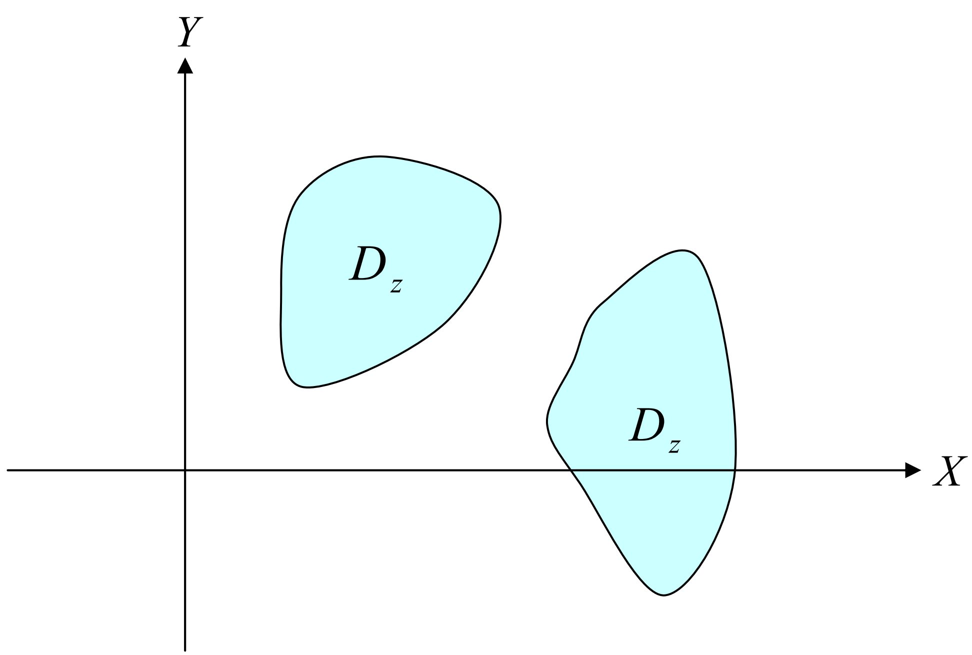


Fig. 8.1

Example 8.1: $Z = X + Y$. Find $f_Z(z)$.

Solution:

$$F_Z(z) = P\left(X + Y \le z\right) = \int_{y=-\infty}^{+\infty} \int_{x=-\infty}^{z-y} f_{XY}(x, y)\,dx\,dy, \tag{8-4}$$

since the region D_z of the xy plane where $x + y \le z$ is the shaded area in Fig. 8.2 to the left of the line $x + y = z$. Integrating over the horizontal strip along the x-axis first (inner integral) followed by sliding that strip along the y-axis from $-\infty$ to $+\infty$ (outer integral) we cover the entire shaded area.

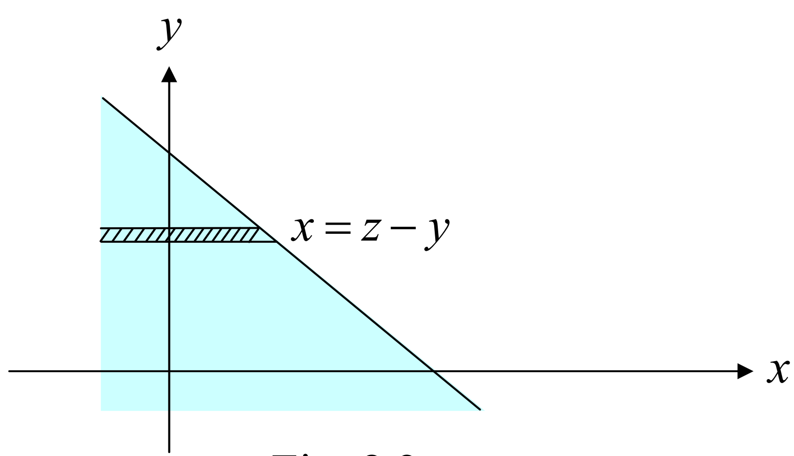


Fig. 8.2

We can find $f_Z(z)$ by differentiating $F_Z(z)$ directly. In this context, it is useful to recall the differentiation rule in (7-15) - (7-16) due to Leibnitz. Suppose

$$H(z) = \int_{a(z)}^{b(z)} h(x, z)\,dx. \tag{8-5}$$

Then

$$\frac{dH(z)}{dz} = \frac{db(z)}{dz} h\big(b(z), z\big) - \frac{da(z)}{dz} h\big(a(z), z\big) + \int_{a(z)}^{b(z)} \frac{\partial h(x, z)}{\partial z}\,dx. \tag{8-6}$$

Using (8-6) in (8-4) we get

$$f_Z(z) = \int_{-\infty}^{+\infty} \left(\frac{\partial}{\partial z} \int_{-\infty}^{z-y} f_{XY}(x, y)\,dx \right) dy = \int_{-\infty}^{+\infty} \left(f_{XY}(z-y, y) - 0 + \int_{-\infty}^{z-y} \frac{\partial f_{XY}(x, y)}{\partial z} \right) dy$$

$$= \int_{-\infty}^{+\infty} f_{XY}(z-y, y)\,dy. \tag{8-7}$$

Alternatively, the integration in (8-4) can be carried out first along the *y*-axis followed by the *x*-axis as in Fig. 8.3.

In that case

$$F_Z(z) = \int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{z-x} f_{XY}(x,y)\,dx\,dy, \qquad \text{(8-8)}$$

and differentiation of (8-8) gives

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \int_{x=-\infty}^{+\infty} \left(\frac{\partial}{\partial z} \int_{y=-\infty}^{z-x} f_{XY}(x,y)\,dy \right) dx$$

$$= \int_{x=-\infty}^{+\infty} f_{XY}(x, z-x)\,dx. \qquad \text{(8-9)}$$

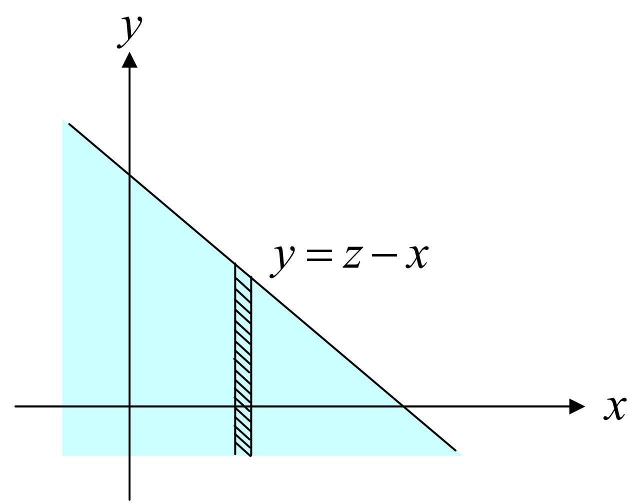


Fig. 8.3

If X and Y are independent, then

$$f_{XY}(x,y) = f_X(x) f_Y(y) \qquad \text{(8-10)}$$

and inserting (8-10) into (8-8) and (8-9), we get

$$f_Z(z) = \int_{y=-\infty}^{+\infty} f_X(z-y) f_Y(y)\,dy = \int_{x=-\infty}^{+\infty} f_X(x) f_Y(z-x)\,dx. \qquad \text{(8-11)}$$

The above integral is the standard convolution of the functions $f_X(z)$ and $f_Y(z)$ expressed two different ways. We thus reach the following conclusion: If two r.vs are independent, then the density of their sum equals the convolution of their density functions.

As a special case, suppose that $f_X(x) = 0$ for $x < 0$ and $f_Y(y) = 0$ for $y < 0,$ then we can make use of Fig. 8.4 to determine the new limits for D_z.

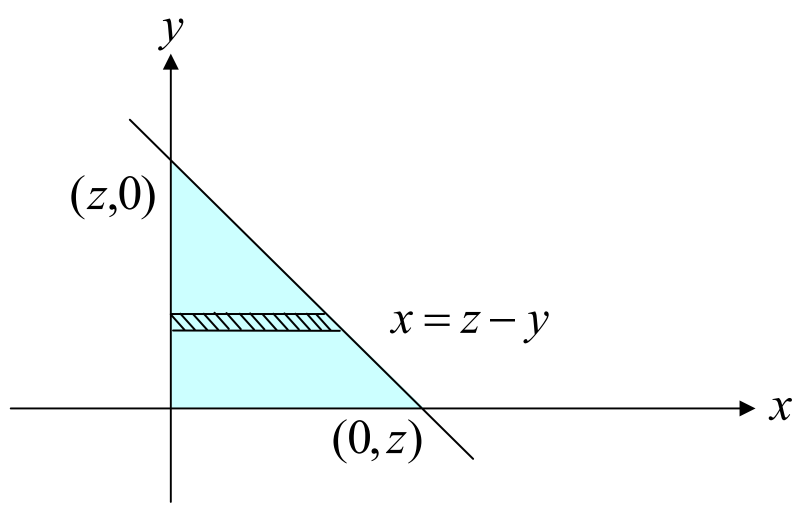


Fig. 8.4

In that case

$$F_Z(z) = \int_{y=0}^{z} \int_{x=0}^{z-y} f_{XY}(x,y)\,dx\,dy$$

or

$$f_Z(z) = \int_{y=0}^{z} \left(\frac{\partial}{\partial z} \int_{x=0}^{z-y} f_{XY}(x,y)\,dx \right) dy = \begin{cases} \int_0^z f_{XY}(z-y,y)\,dy, & z>0, \\ 0, & z \le 0. \end{cases} \qquad \text{(8-12)}$$

On the other hand, by considering vertical strips first in Fig. 8.4, we get

$$F_Z(z) = \int_{x=0}^{z} \int_{y=0}^{z-x} f_{XY}(x,y)\,dy\,dx$$

or

$$f_Z(z) = \int_{x=0}^{z} f_{XY}(x, z-x)\,dx = \begin{cases} \int_{y=0}^{z} f_X(x) f_Y(z-x)\,dx, & z>0, \\ 0, & z \le 0, \end{cases} \qquad \text{(8-13)}$$

if X and Y are independent random variables.

Example 8.2: Suppose X and Y are independent exponential r.vs with common parameter λ, and let $Z = X + Y$. Determine $f_Z(z)$.

Solution: We have $f_X(x) = \lambda e^{-\lambda x} U(x), \quad f_Y(y) = \lambda e^{-\lambda y} U(y),$ (8-14)
and we can make use of (13) to obtain the p.d.f of $Z = X + Y$.

$$f_Z(z) = \int_0^z \lambda^2 e^{-\lambda x} e^{-\lambda(z-x)} dx = \lambda^2 e^{-\lambda z} \int_0^z dx = z\lambda^2 e^{-\lambda z} U(z). \qquad (8\text{-}15)$$

As the next example shows, care should be taken in using the convolution formula for r.vs with finite range.

Example 8.3: X and Y are independent uniform r.vs in the common interval (0,1). Determine $f_Z(z)$, where $Z = X + Y$.
Solution: Clearly, $Z = X + Y \Rightarrow 0 < z < 2$ here, and as Fig. 8.5 shows there are two cases of z for which the shaded areas are quite different in shape and they should be considered separately.

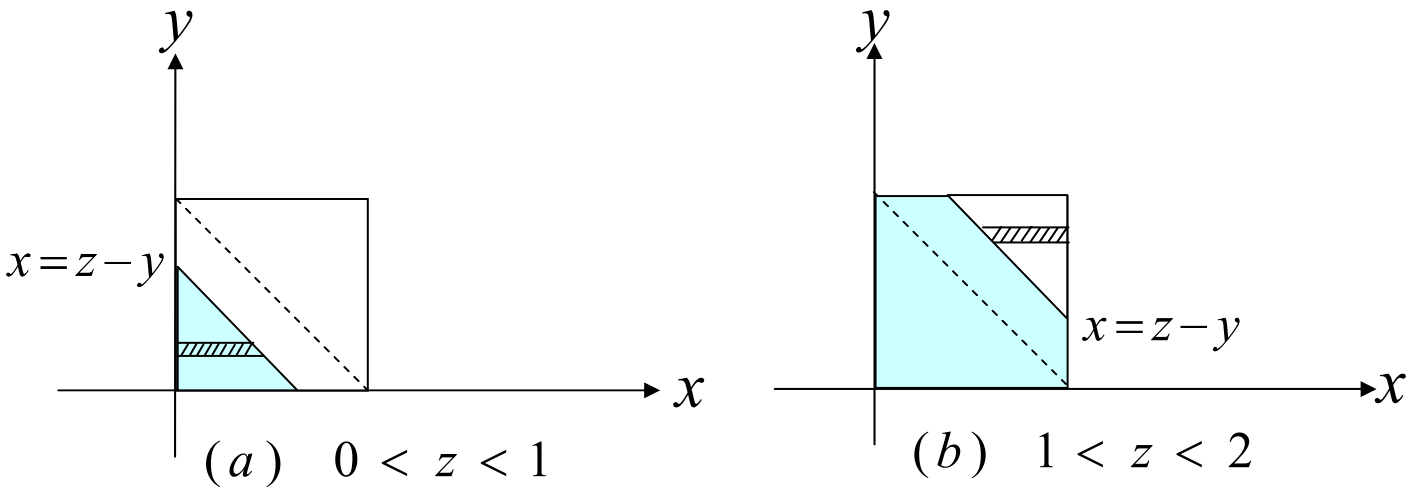


Fig. 8.5

For $0 \le z < 1,$

$$F_Z(z) = \int_{y=0}^{z} \int_{x=0}^{z-y} 1 \; dxdy = \int_{y=0}^{z} (z-y)dy = \frac{z^2}{2}, \quad 0 \le z < 1. \qquad (8\text{-}16)$$

For $1 \le z < 2,$ notice that it is easy to deal with the unshaded region. In that case

$$F_Z(z) = 1 - P\big(Z > z\big) = 1 - \int_{y=z-1}^{1} \int_{x=z-y}^{1} 1 \; dxdy$$

$$= 1 - \int_{y=z-1}^{1} (1 - z + y)dy = 1 - \frac{(2-z)^2}{2}, \quad 1 \le z < 2. \qquad (8\text{-}17)$$

Using (8-16) - (8-17), we obtain

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \begin{cases} z & 0 \le z < 1, \\ 2-z, & 1 \le z < 2. \end{cases} \tag{8-18}$$

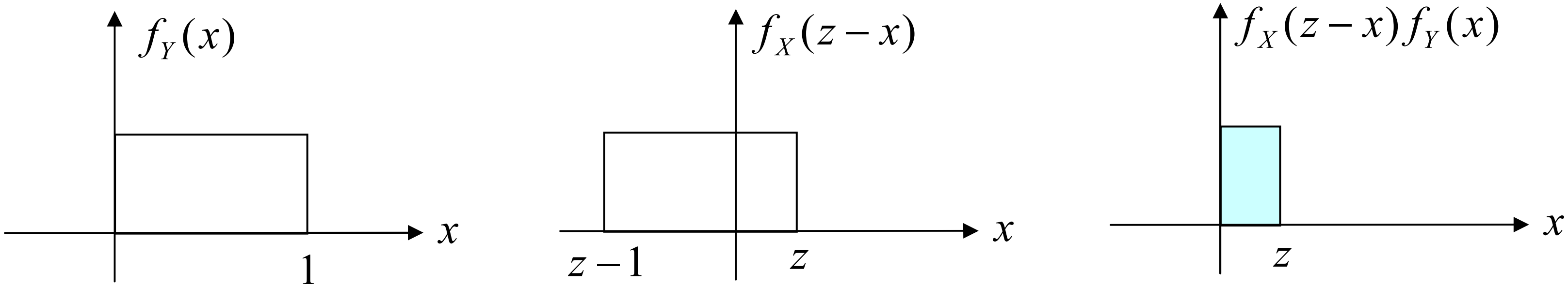
By direct convolution of $f_X(x)$ and $f_Y(y)$, we obtain the same result as above. In fact, for $0 \le z < 1$ (Fig. 8.6(a))

$$f_Z(z) = \int f_X(z-x) f_Y(x) dx = \int_0^z 1 \; dx = z. \tag{8-19}$$

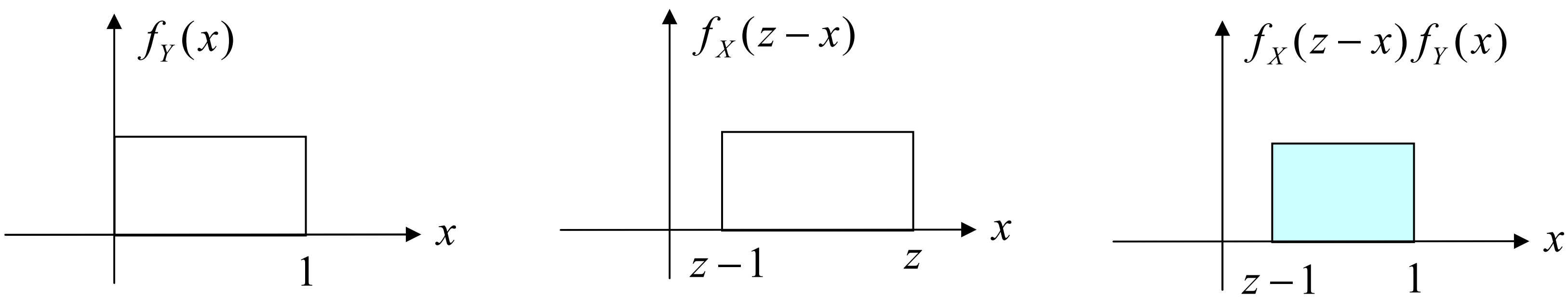
and for $1 \le z < 2$ (Fig. 8.6(b))

$$f_Z(z) = \int_{z-1}^{1} 1 \; dx = 2 - z. \tag{8-20}$$

Fig 8.6 (c) shows $f_Z(z)$ which agrees with the convolution of two rectangular waveforms as well.



$(a)\quad 0 \le z < 1$



$(b)\quad 1 \le z < 2$

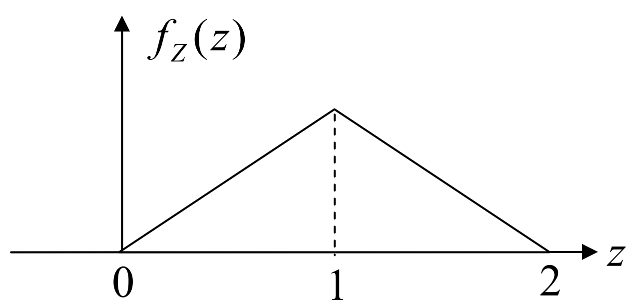


Fig. 8.6 (c)

Example 8.3: Let $Z = X - Y$. Determine its p.d.f $f_Z(z)$.

Solution: From (8-3) and Fig. 8.7

$$F_Z(z) = P(X - Y \le z) = \int_{y=-\infty}^{+\infty} \int_{x=-\infty}^{z+y} f_{XY}(x,y)\,dx\,dy$$

and hence

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \int_{y=-\infty}^{+\infty} \left(\frac{\partial}{\partial z} \int_{x=-\infty}^{z+y} f_{XY}(x,y)\,dx \right) dy = \int_{-\infty}^{+\infty} f_{XY}(y+z,y)\,dy. \quad (8\text{-}21)$$

If X and Y are independent, then the above formula reduces to

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(z+y) f_Y(y)\,dy = f_X(-z) \otimes f_Y(y), \quad (8\text{-}22)$$

which represents the convolution of $f_X(-z)$ with $f_Y(z)$.

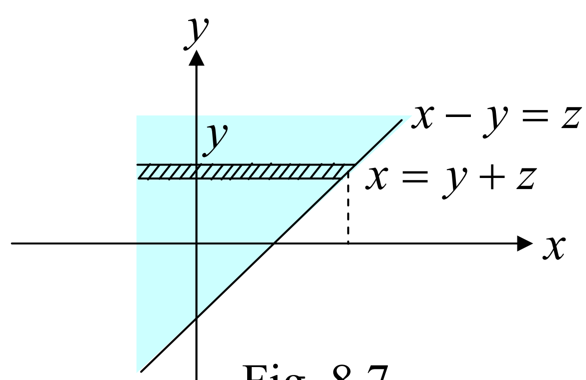


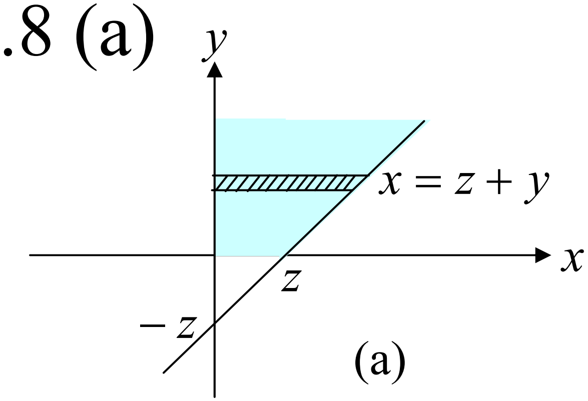
Fig. 8.7

As a special case, suppose

$$f_X(x)=0,\ x<0,\ \text{ and }\ f_Y(y)=0,\ y<0.$$

In this case, Z can be negative as well as positive, and that gives rise to two situations that should be analyzed separately, since the region of integration for $z \ge 0$ and $z<0$ are quite different. For $z \ge 0$, from Fig. 8.8 (a)

$$F_Z(z)=\int_{y=0}^{+\infty}\int_{x=0}^{z+y} f_{XY}(x,y)dxdy$$



(a)

and for $z<0$, from Fig 8.8 (b)

$$F_Z(z)=\int_{y=-z}^{+\infty}\int_{x=0}^{z+y} f_{XY}(x,y)dxdy$$

After differentiation, this gives

$$f_Z(z)=\begin{cases}\int_0^{+\infty} f_{XY}(z+y,y)dy, & z\ge 0,\\ \int_{-z}^{+\infty} f_{XY}(z+y,y)dy, & z<0.\end{cases} \qquad (8\text{-}23)$$

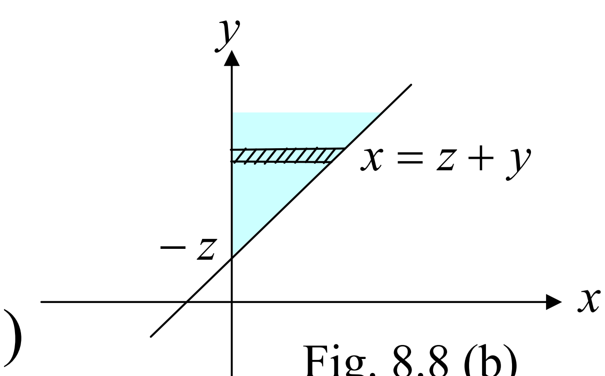


Fig. 8.8 (b)

Example 8.4: Given $Z = X / Y$, obtain its density function.

Solution: We have $F_Z(z) = P(X / Y \le z).$ (8-24)

The inequality $X / Y \le z$ can be rewritten as $X \le Yz$ if $Y > 0$, and $X \ge Yz$ if $Y < 0$. Hence the event $(X / Y \le z)$ in (8-24) need to be conditioned by the event $A = (Y > 0)$ and its compliment $\overline{A}$. Since $A \cup \overline{A} = \Omega$, by the partition theorem, we have

$$\{X / Y \le z\} = \{(X / Y \le z) \cap (A \cup \overline{A})\} = \{(X / Y \le z) \cap A\} \cup \{(X / Y \le z) \cap \overline{A}\}$$

and hence by the mutually exclusive property of the later two events

$$\begin{aligned} P(X / Y \le z) &= P(X / Y \le z, Y > 0) + P(X / Y \le z, Y < 0) \\ &= P(X \le Yz, Y > 0) + P(X \ge Yz, Y < 0). \end{aligned} \tag{8-25}$$

Fig. 8.9(a) shows the area corresponding to the first term, and Fig. 8.9(b) shows that corresponding to the second term in (8-25).

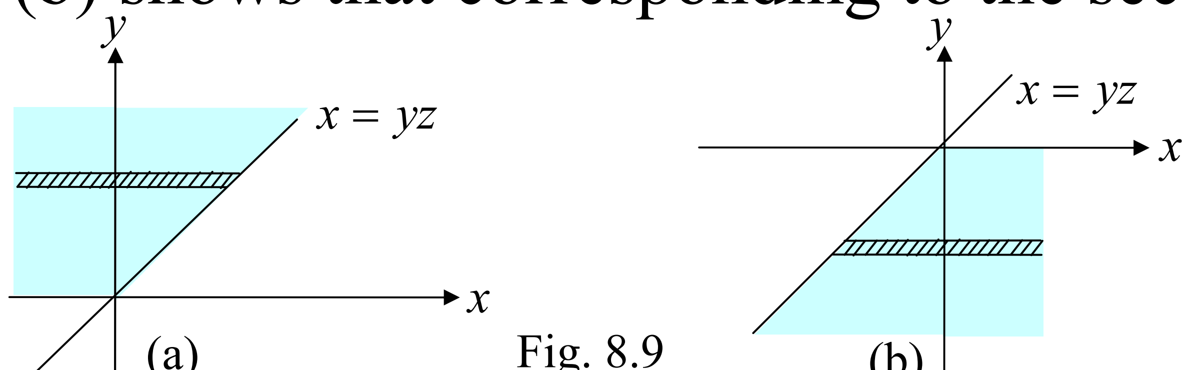


Fig. 8.9

Integrating over these two regions, we get

$$F_Z(z) = \int_{y=0}^{+\infty} \int_{x=-\infty}^{yz} f_{XY}(x,y)\,dx\,dy + \int_{y=-\infty}^{0} \int_{x=yz}^{\infty} f_{XY}(x,y)\,dx\,dy. \qquad (8\text{-}26)$$

Differentiation with respect to z gives

$$f_Z(z) = \int_0^{+\infty} y f_{XY}(yz,y)\,dy + \int_{-\infty}^{0} (-y) f_{XY}(yz,y)\,dy$$

$$= \int_{-\infty}^{+\infty} |\,y\,|\, f_{XY}(yz,y)\,dy, \qquad -\infty < z < +\infty. \qquad (8\text{-}27)$$

Note that if X and Y are nonnegative random variables, then the area of integration reduces to that shown in Fig. 8.10.

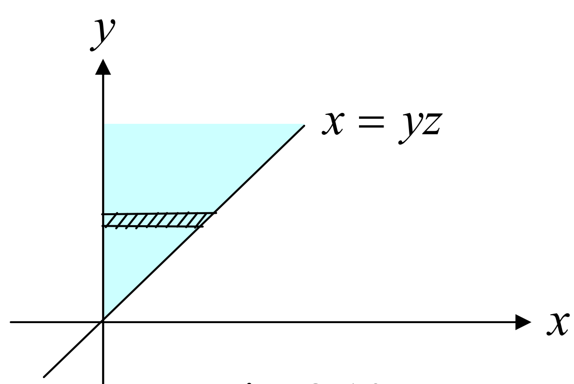


Fig. 8.10

This gives

$$F_Z(z) = \int_{y=0}^{\infty} \int_{x=0}^{yz} f_{XY}(x, y)\,dx\,dy$$

or

$$f_Z(z) = \begin{cases} \int_0^{+\infty} y\, f_{XY}(yz, y)\,dy, & z > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{8-28}$$

Example 8.5: X and Y are jointly normal random variables with zero mean so that

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}\left(\frac{x^2}{\sigma_1^2} - \frac{2rxy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right)}. \tag{8-29}$$

Show that the ratio $Z = X/Y$ has a Cauchy density function centered at $r\sigma_1/\sigma_2$.

Solution: Inserting (8-29) into (8-27) and using the fact that $f_{XY}(-x,-y) = f_{XY}(x, y)$, we obtain

$$f_Z(z) = \frac{2}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \int_0^{\infty} y e^{-y^2/2\sigma_0^2}\,dy = \frac{\sigma_0^2(z)}{\pi\sigma_1\sigma_2\sqrt{1-r^2}},$$

where

$$\sigma_0^2(z) = \frac{1-r^2}{\dfrac{z^2}{\sigma_1^2} - \dfrac{2rz}{\sigma_1\sigma_2} + \dfrac{1}{\sigma_2^2}}.$$

Thus

$$f_Z(z) = \frac{\sigma_1\sigma_2\sqrt{1-r^2}/\pi}{\sigma_2^2(z - r\sigma_1/\sigma_2)^2 + \sigma_1^2(1-r^2)}, \tag{8-30}$$

which represents a Cauchy r.v centered at $r\sigma_1/\sigma_2.$ Integrating (8-30) from $-\infty$ to z, we obtain the corresponding distribution function to be

$$F_Z(z) = \frac{1}{2} + \frac{1}{\pi}\arctan\frac{\sigma_2 z - r\sigma_1}{\sigma_1\sqrt{1-r^2}}. \tag{8-31}$$

Example 8.6: $Z = X^2 + Y^2.$ Obtain $f_Z(z).$

Solution: We have

$$F_Z(z) = P\left(X^2 + Y^2 \le z\right) = \int\int_{X^2+Y^2\le z} f_{XY}(x,y)\,dxdy\,. \tag{8-32}$$

But, $X^2 + Y^2 \le z$ represents the area of a circle with radius $\sqrt{z}$, and hence from Fig. 8.11,

$$F_Z(z) = \int_{y=-\sqrt{z}}^{\sqrt{z}} \int_{x=-\sqrt{z-y^2}}^{\sqrt{z-y^2}} f_{XY}(x, y)\,dx\,dy. \tag{8-33}$$

This gives after repeated differentiation

$$f_Z(z) = \int_{y=-\sqrt{z}}^{\sqrt{z}} \frac{1}{2\sqrt{z-y^2}} \left(f_{XY}(\sqrt{z-y^2}, y) + f_{XY}(-\sqrt{z-y^2}, y) \right) dy. \tag{8-34}$$

As an illustration, consider the next example.

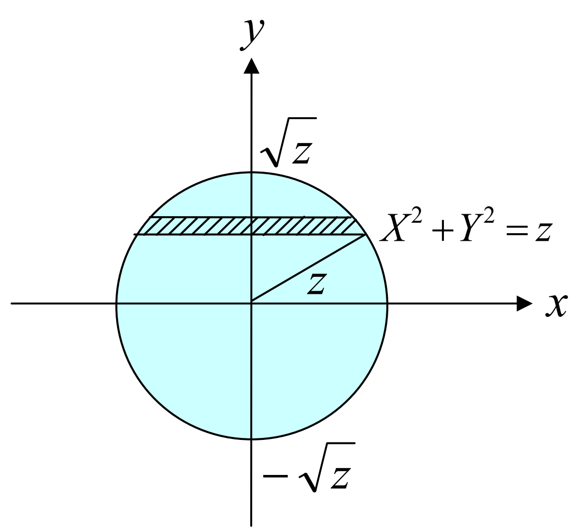


Fig. 8.11

Example 8.7 : X and Y are independent normal r.vs with zero Mean and common variance σ^2. Determine $f_Z(z)$ for $Z = X^2 + Y^2$.
Solution: Direct substitution of (8-29) with $r = 0, \quad \sigma_1 = \sigma_2 = \sigma$ Into (8-34) gives

$$f_Z(z) = \int_{y=-\sqrt{z}}^{\sqrt{z}} \frac{1}{2\sqrt{z - y^2}} \left(2 \cdot \frac{1}{2\pi\sigma^2} e^{-(z-y^2+y^2)/2\sigma^2} \right) dy = \frac{e^{-z/2\sigma^2}}{\pi\sigma^2} \int_0^{\sqrt{z}} \frac{1}{\sqrt{z - y^2}} \, dy$$

$$= \frac{e^{-z/2\sigma^2}}{\pi\sigma^2} \int_0^{\pi/2} \frac{\sqrt{z}\cos\theta}{\sqrt{z}\cos\theta} \, d\theta = \frac{1}{2\sigma^2} e^{-z/2\sigma^2} U(z), \qquad (8\text{-}35)$$

where we have used the substitution $y = \sqrt{z}\sin\theta$. From (8-35) we have the following result: If X and Y are independent zero mean Gaussian r.vs with common variance σ^2, then $X^2 + Y^2$ is an exponential r.vs with parameter $2\sigma^2$.

Example 8.8 : Let $Z = \sqrt{X^2 + Y^2}$. Find $f_Z(z)$.
Solution: From Fig. 8.11, the present case corresponds to a circle with radius z^2. Thus

$$F_Z(z) = \int_{y=-z}^{z} \int_{x=-\sqrt{z^2-y^2}}^{\sqrt{z^2-y^2}} f_{XY}(x, y)\,dx\,dy \ .$$

And by repeated differentiation, we obtain

$$f_Z(z) = \int_{-z}^{z} \frac{z}{\sqrt{z^2 - y^2}} \left(f_{XY}(\sqrt{z^2 - y^2}, y) + f_{XY}(-\sqrt{z^2 - y^2}, y) \right) dy \ . \quad (8\text{-}36)$$

Now suppose X and Y are independent Gaussian as in Example 8.7. In that case, (8-36) simplifies to

$$f_Z(z) = 2\int_0^z \frac{z}{\sqrt{z^2 - y^2}} \frac{1}{2\pi\sigma^2} e^{(z^2 - y^2 + y^2)/2\sigma^2} dy = \frac{2z}{\pi\sigma^2} e^{-z^2/2\sigma^2} \int_0^z \frac{1}{\sqrt{z^2 - y^2}} dy$$

$$= \frac{2z}{\pi\sigma^2} e^{-z^2/2\sigma^2} \int_0^{\pi/2} \frac{z\cos\theta}{z\cos\theta} d\theta = \frac{z}{\sigma^2} e^{-z^2/2\sigma^2} U(z), \quad (8\text{-}37)$$

which represents a Rayleigh distribution. Thus, if $W = X + iY$, where X and Y are real, independent normal r.vs with zero mean and equal variance, then the r.v $|W| = \sqrt{X^2 + Y^2}$ has a Rayleigh density. W is said to be a complex Gaussian r.v with zero mean, whose real and imaginary parts are independent r.vs. From (8-37), we have seen that its magnitude has Rayleigh distribution.

What about its phase

$$\theta = \tan^{-1}\left(\frac{X}{Y}\right)? \tag{8-38}$$

Clearly, the principal value of θ lies in the interval $(-\pi/2, \pi/2)$. If we let $U = \tan\theta = X/Y$, then from example 8.5, U has a Cauchy distribution with (see (8-30) with $\sigma_1 = \sigma_2,\ r = 0$)

$$f_U(u) = \frac{1/\pi}{u^2+1}, \quad -\infty < u < \infty. \tag{8-39}$$

As a result

$$f_\theta(\theta) = \frac{1}{|d\theta/du|} f_U(\tan\theta) = \frac{1}{(1/\sec^2\theta)} \frac{1/\pi}{\tan^2\theta + 1} = \begin{cases} 1/\pi, & -\pi/2 < \theta < \pi/2, \\ 0, & \text{otherwise}. \end{cases} \tag{8-40}$$

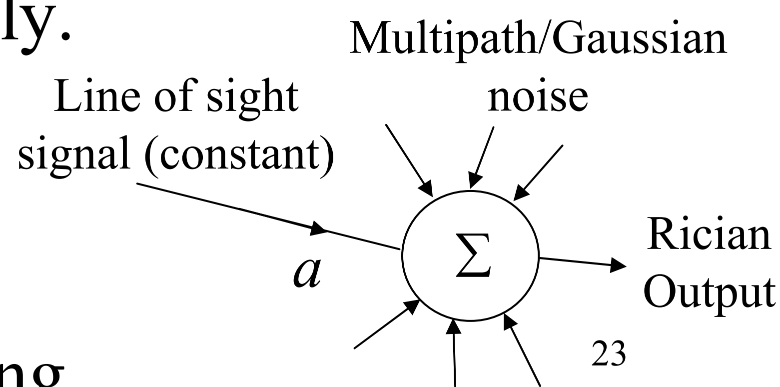
To summarize, the magnitude and phase of a zero mean complex Gaussian r.v has Rayleigh and uniform distributions respectively. Interestingly, as we will show later, these two derived r.vs are also independent of each other!

Let us reconsider example 8.8 where X and Y have nonzero means μ_X and μ_Y respectively. Then $Z = \sqrt{X^2 + Y^2}$ is said to be a Rician r.v. Such a scene arises in fading multipath situation where there is a dominant constant component (mean) in addition to a zero mean Gaussian r.v. The constant component may be the line of sight signal and the zero mean Gaussian r.v part could be due to random multipath components adding up incoherently (see diagram below). The envelope of such a signal is said to have a Rician p.d.f.

Example 8.9: Redo example 8.8, where X and Y have nonzero means μ_X and μ_Y respectively.

Solution: Since

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma^2} e^{-[(x-\mu_X)^2 + (y-\mu_Y)^2]/2\sigma^2},$$



substituting this into (8-36) and letting

$$x = z\cos\theta,\; y = z\sin\theta,\;\; \mu = \sqrt{\mu_X^2 + \mu_Y^2},\;\; \mu_X = \mu\cos\phi,\;\; \mu_Y = \mu\sin\phi,$$

we get the Rician probability density function to be

$$\begin{aligned} f_Z(z) &= \frac{z e^{-(z^2+\mu^2)/2\sigma^2}}{2\pi\sigma^2} \int_{-\pi/2}^{\pi/2} \left(e^{z\mu\cos(\theta-\phi)/\sigma^2} + e^{-z\mu\cos(\theta+\phi)/\sigma^2} \right) d\theta \\ &= \frac{z e^{-(z^2+\mu^2)/2\sigma^2}}{2\pi\sigma^2} \left(\int_{-\pi/2}^{\pi/2} e^{z\mu\cos(\theta-\phi)/\sigma^2}\, d\theta + \int_{\pi/2}^{3\pi/2} e^{z\mu\cos(\theta-\phi)/\sigma^2}\, d\theta \right) \\ &= \frac{z e^{-(z^2+\mu^2)/2\sigma^2}}{2\pi\sigma^2} I_0\left(\frac{z\mu}{\sigma^2} \right), \end{aligned} \tag{8-41}$$

where

$$I_0(\eta) \triangleq \frac{1}{2\pi} \int_0^{2\pi} e^{\eta\cos(\theta-\phi)} d\theta = \frac{1}{\pi} \int_0^{\pi} e^{\eta\cos\theta} d\theta \tag{8-42}$$

is the modified Bessel function of the first kind and zeroth order.

Example 8.10: $Z = \max(X, Y), \quad W = \min(X, Y).$ Determine $f_Z(z).$

Solution: The functions *max* and *min* are nonlinear operators and represent special cases of the more general order statistics. In general, given any n-tuple $X_1, X_2, \cdots, X_n$, we can arrange them in an increasing order of magnitude such that

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}, \tag{8-43}$$

where $X_{(1)} = \min(X_1, X_2, \cdots, X_n)$, and $X_{(2)}$ is the second smallest value among $X_1, X_2, \cdots, X_n$, and finally $X_{(n)} = \max(X_1, X_2, \cdots, X_n)$. If $X_1, X_2, \cdots, X_n$ represent r.vs, the function $X_{(k)}$ that takes on the value $x_{(k)}$ in each possible sequence $(x_1, x_2, \cdots, x_n)$ is known as the k-th order statistic. $(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$ represent the set of order statistics among n random variables. In this context

$$R = X_{(n)} - X_{(1)} \tag{8-44}$$

represents the range, and when $n = 2$, we have the *max* and *min* statistics.

Returning back to that problem, since

$$Z = \max(X, Y) = \begin{cases} X, & X > Y, \\ Y, & X \le Y, \end{cases} \tag{8-45}$$

we have (see also (8-25))

$$\begin{aligned} F_Z(z) &= P\big(\max(X,Y) \le z\big) = P\big[(X \le z, X > Y) \cup (Y \le z, X \le Y)\big] \\ &= P\big(X \le z, X > Y\big) + P\big(Y \le z, X \le Y\big), \end{aligned}$$

since $(X > Y)$ and $(X \le Y)$ are mutually exclusive sets that form a partition. Figs 8.12 (a)-(b) show the regions satisfying the corresponding inequalities in each term above.

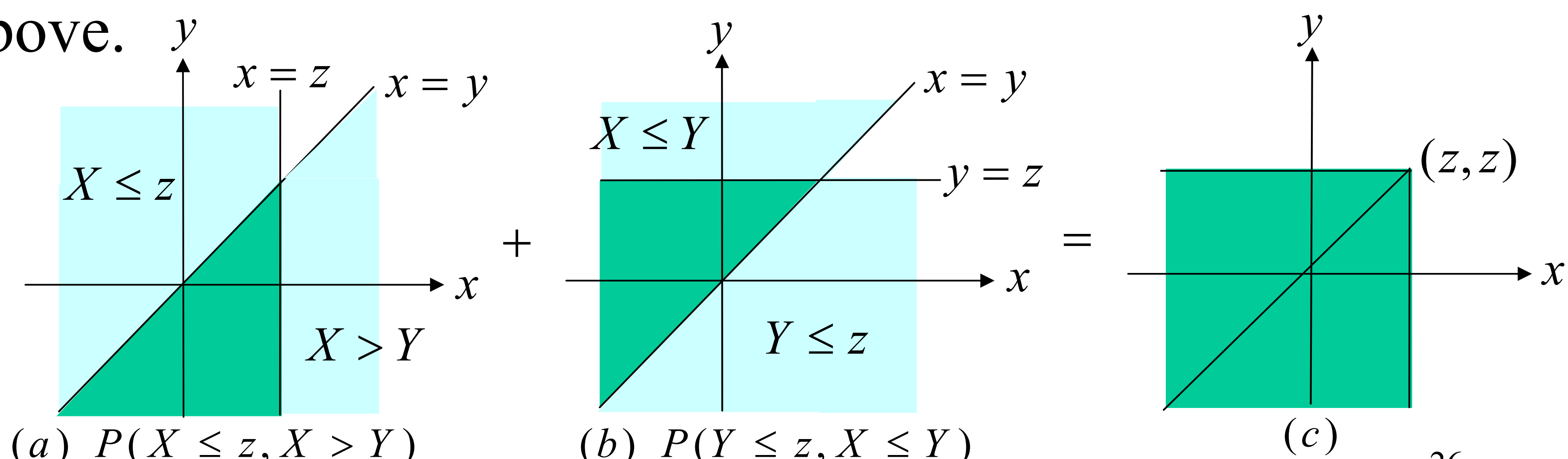


Fig. 8.12

Fig. 8.12 (c) represents the total region, and from there

$$F_Z(z) = P\left(X \le z, Y \le z\right) = F_{XY}(z, z). \tag{8-46}$$

If X and Y are independent, then

$$F_Z(z) = F_X(x) F_Y(y)$$

and hence

$$f_Z(z) = F_X(z) f_Y(z) + f_X(z) F_Y(z). \tag{8-47}$$

Similarly

$$W = \min(X, Y) = \begin{cases} Y, & X > Y, \\ X, & X \le Y. \end{cases} \tag{8-48}$$

Thus

$$F_W(w) = P\left(\min(X, Y) \le w\right) = P\left[\left(Y \le w, X > Y\right) \cup \left(X \le w, X \le Y\right)\right].$$

Once again, the shaded areas in Fig. 8.13 (a)-(b) show the regions satisfying the above inequalities and Fig 8.13 (c) shows the overall region.

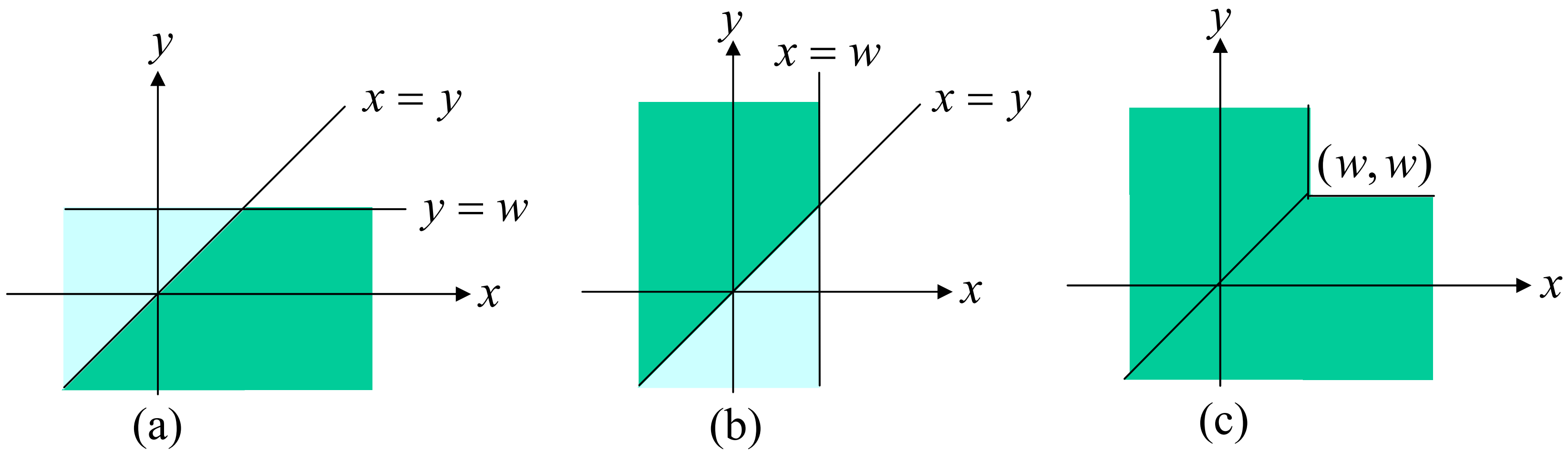


Fig. 8.13

From Fig. 8.13 (c),

$$F_W(w) = 1 - P(W > w) = 1 - P(X > w, Y > w)$$

$$= F_X(w) + F_Y(w) - F_{XY}(w, w), \qquad (8\text{-}49)$$

where we have made use of (7-5) and (7-12) with $x_2 = y_2 = +\infty$, and $x_1 = y_1 = w$.

Example 8.11: Let X and Y be independent exponential r.vs with common parameter λ. Define $W = \min(X, Y)$. Find $f_W(w)$?
Solution: From (8-49)

$$F_W(w) = F_X(w) + F_Y(w) - F_X(w)F_Y(w)$$

and hence

$$f_W(w) = f_X(w) + f_Y(w) - f_X(w)F_Y(w) - F_X(w)f_Y(w).$$

But $f_X(w) = f_Y(w) = \lambda e^{-\lambda w}$, and $F_X(w) = F_Y(w) = 1 - e^{-\lambda w}$, so that

$$f_W(w) = 2\lambda e^{\lambda w} - 2(1 - e^{-\lambda w})\lambda e^{-\lambda w} = 2\lambda e^{-2\lambda w}U(w). \qquad (8\text{-}50)$$

Thus min (X, Y) is also exponential with parameter 2λ.

Example 8.12: Suppose X and Y are as give in the above example. Define $Z = \left[\min(X, Y) / \max(X, Y)\right]$. Determine $f_Z(z)$.

Solution: Although $\min(\cdot)/\max(\cdot)$ represents a complicated function, by partitioning the whole space as before, it is possible to simplify this function. In fact

$$Z = \begin{cases} X/Y, & X \le Y, \\ Y/X, & X > Y. \end{cases} \tag{8-51}$$

As before, this gives

$$F_Z(z) = P(Z \le z) = P(X/Y \le z, X \le Y) + P(Y/X \le z, X > Y)$$
$$= P(X \le Yz, X \le Y) + P(Y \le Xz, X > Y) \; . \tag{8-52}$$

Since X and Y are both positive random variables in this case, we have $0 < z < 1$. The shaded regions in Figs 8.14 (a)-(b) represent the two terms in the above sum.

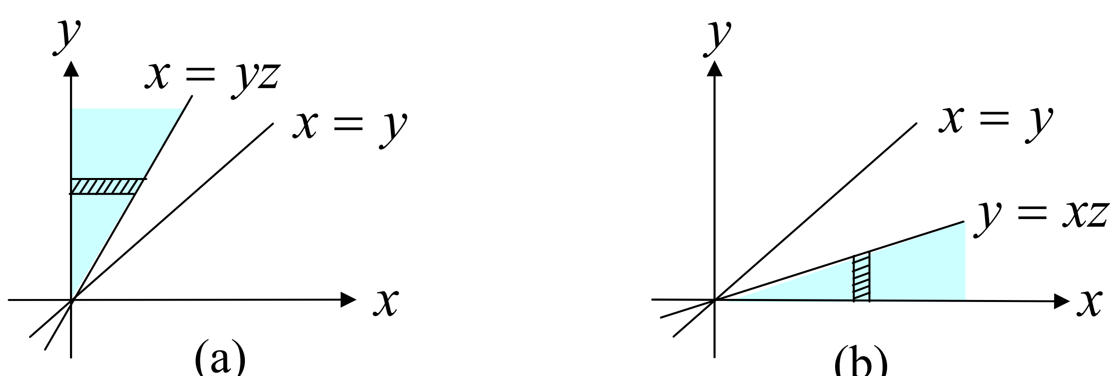


Fig. 8.14

From Fig. 8.14

$$F_Z(z) = \int_0^\infty \int_{x=0}^{yz} f_{XY}(x,y)dxdy + \int_0^\infty \int_{y=0}^{xz} f_{XY}(x,y)dydx. \tag{8-53}$$

Hence

$$f_Z(z) = \int_0^\infty y\, f_{XY}(yz,y)dy + \int_0^\infty x\, f_{XY}(x,xz)dx = \int_0^\infty y\{f_{XY}(yz,y) + f_{XY}(y,yz)\}dy$$

$$= \int_0^\infty y\lambda^2\left\{e^{-\lambda(yz+y)} + e^{-\lambda(y+yz)}\right\}dy = 2\lambda^2\int_0^\infty ye^{-\lambda(1+z)y}dy = \frac{2}{(1+z)^2}\int_0^\infty ue^{-u}dy$$

$$= \begin{cases} \dfrac{2}{(1+z)^2}, & 0 < z < 1, \\ 0, & \text{otherwise.} \end{cases} \tag{8-54}$$

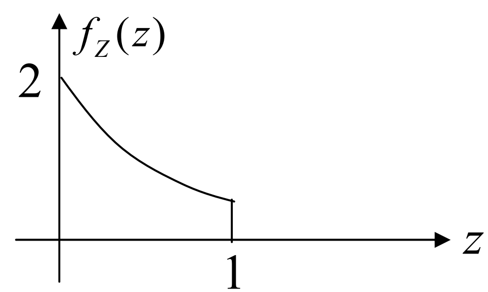


Fig. 8.15

Example 8.13 (Discrete Case): Let X and Y be independent Poisson random variables with parameters λ_1 and λ_2 respectively. Let $Z = X + Y$. Determine the p.m.f of Z.

Solution: Since X and Y both take integer values $\{0, 1, 2, \cdots\}$, the same is true for Z. For any $n = 0, 1, 2, \cdots, \quad X + Y = n$ gives only a finite number of options for X and Y. In fact, if $X = 0$, then Y must be n; if $X = 1$, then Y must be n-1, etc. Thus the event $\{X + Y = n\}$ is the union of $(n + 1)$ mutually exclusive events A_k given by

$$A_k = \{X = k, \ \ Y = n - k\}, \qquad k = 0, 1, 2, \cdots, n. \tag{8-55}$$

As a result

$$P(Z = n) = P(X + Y = n) = P\left(\bigcup_{k=0}^{n} (X = k, \ Y = n - k)\right)$$

$$= \sum_{k=0}^{n} P(X = k, \ Y = n - k). \tag{8-56}$$

If X and Y are also independent, then

$$P\left(X = k, \ Y = n - k\right) = P(X = k) P(Y = n - k)$$

and hence

$$P(Z=n)=\sum_{k=0}^{n} P(X=k,\ Y=n-k)$$
$$=\sum_{k=0}^{n} e^{-\lambda_1}\frac{\lambda_1^k}{k!}e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k)!}=\frac{e^{-(\lambda_1+\lambda_2)}}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_1^k\lambda_2^{n-k}$$
$$=e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^n}{n!}, \qquad n=0,1,2,\cdots,\infty. \tag{8-57}$$

Thus Z represents a Poisson random variable with parameter $\lambda_1+\lambda_2$, indicating that sum of independent Poisson random variables is also a Poisson random variable whose parameter is the sum of the parameters of the original random variables.

As the last example illustrates, the above procedure for determining the p.m.f of functions of discrete random variables is somewhat tedious. As we shall see in Lecture 10, the joint characteristic function can be used in this context to solve problems of this type in an easier fashion.

9. Two Functions of Two Random Variables

In the spirit of the previous lecture, let us look at an immediate generalization: Suppose X and Y are two random variables with joint p.d.f $f_{XY}(x,y)$. Given two functions $g(x,y)$ and $h(x,y)$, define the new random variables

$$Z = g(X,Y) \tag{9-1}$$

$$W = h(X,Y). \tag{9-2}$$

How does one determine their joint p.d.f $f_{ZW}(z,w)$? Obviously with $f_{ZW}(z,w)$ in hand, the marginal p.d.fs $f_Z(z)$ and $f_W(w)$ can be easily determined.

The procedure is the same as that in (8-3). In fact for given z and w,

$$F_{ZW}(z,w) = P\big(Z(\xi) \le z, W(\xi) \le w\big) = P\big(g(X,Y) \le z, h(X,Y) \le w\big)$$
$$= P\big((X,Y) \in D_{z,w}\big) = \int\int_{(x,y)\in D_{z,w}} f_{XY}(x,y)\,dx\,dy, \tag{9-3}$$

where $D_{z,w}$ is the region in the xy plane such that the inequalities $g(x,y) \le z$ and $h(x,y) \le w$ are simultaneously satisfied.

We illustrate this technique in the next example.

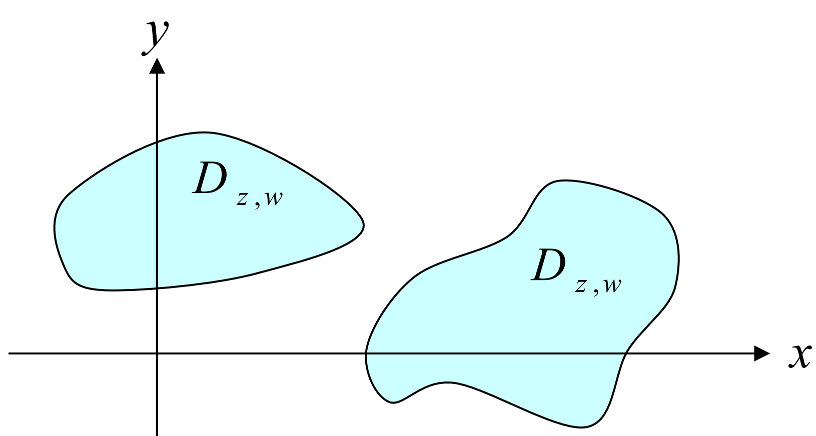


Fig. 9.1

Example 9.1: Suppose X and Y are independent uniformly distributed random variables in the interval $(0, \theta)$.
Define $Z = \min(X, Y), \; W = \max(X, Y)$. Determine $f_{ZW}(z, w)$.
Solution: Obviously both w and z vary in the interval $(0, \theta)$.
Thus $F_{ZW}(z, w) = 0, \quad \text{if} \;\; z < 0 \quad \text{or} \quad w < 0.$ (9-4)

$$F_{ZW}(z, w) = P(Z \le z, W \le w) = P(\min(X, Y) \le z, \; \max(X, Y) \le w). \tag{9-5}$$

We must consider two cases: $w \ge z$ and $w < z$, since they give rise to different regions for $D_{z,w}$ (see Figs. 9.2 (a)-(b)).

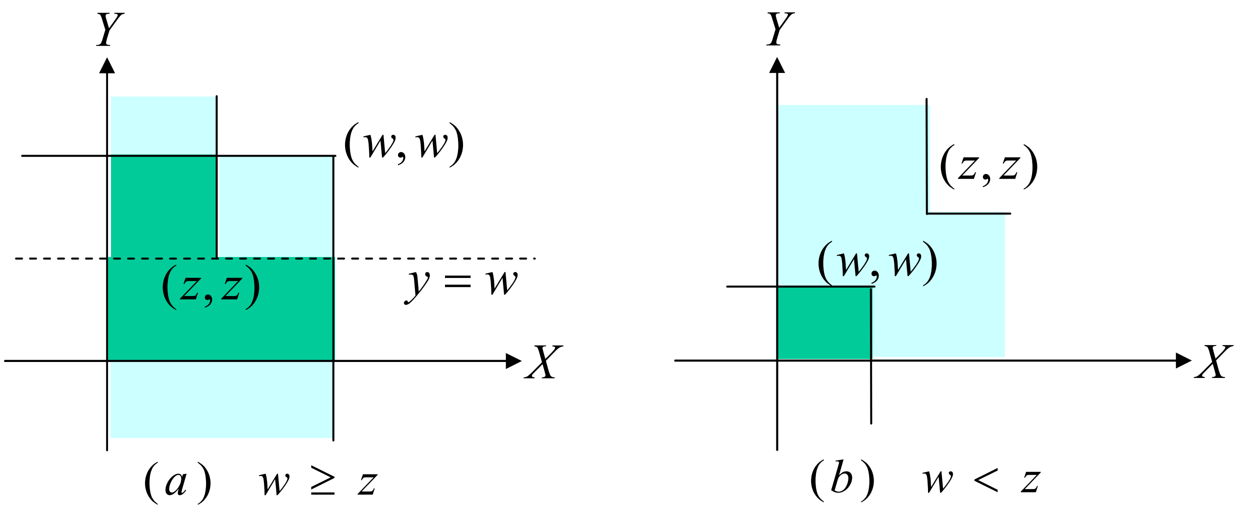


Fig. 9.2

For $w \geq z$, from Fig. 9.2 (a), the region $D_{z,w}$ is represented by the doubly shaded area. Thus

$$F_{ZW}(z,w) = F_{XY}(z,w) + F_{XY}(w,z) - F_{XY}(z,z), \quad w \geq z, \qquad (9\text{-}6)$$

and for $w < z$, from Fig. 9.2 (b), we obtain

$$F_{ZW}(z,w) = F_{XY}(w,w), \quad w < z. \qquad (9\text{-}7)$$

With

$$F_{XY}(x,y) = F_X(x)\,F_Y(y) = \frac{x}{\theta}\cdot\frac{y}{\theta} = \frac{xy}{\theta^2}, \qquad (9\text{-}8)$$

we obtain

$$F_{ZW}(z,w) = \begin{cases} (2w-z)z/\theta^2, & 0 < z < w < \theta, \\ w^2/\theta^2, & 0 < w < z < \theta. \end{cases} \qquad (9\text{-}9)$$

Thus

$$f_{ZW}(z,w) = \begin{cases} 2/\theta^2, & 0 < z < w < \theta, \\ 0, & \text{otherwise}. \end{cases} \qquad (9\text{-}10)$$

From (9-10), we also obtain

$$f_Z(z) = \int_z^\theta f_{ZW}(z,w)dw = \frac{2}{\theta}\left(1-\frac{z}{\theta}\right), \quad 0 < z < \theta, \tag{9-11}$$

and

$$f_W(w) = \int_0^w f_{ZW}(z,w)dz = \frac{2w}{\theta^2}, \quad 0 < w < \theta. \tag{9-12}$$

If $g(x,y)$ and $h(x,y)$ are continuous and differentiable functions, then as in the case of one random variable (see (5-30)) it is possible to develop a formula to obtain the joint p.d.f $f_{ZW}(z,w)$ directly. Towards this, consider the equations

$$g(x,y) = z, \qquad h(x,y) = w. \tag{9-13}$$

For a given point (z,w), equation (9-13) can have many solutions. Let us say

$$(x_1,y_1),\ (x_2,y_2),\ \cdots,\ (x_n,y_n),$$

represent these multiple solutions such that (see Fig. 9.3)

$$g(x_i, y_i) = z, \qquad h(x_i, y_i) = w. \tag{9-14}$$

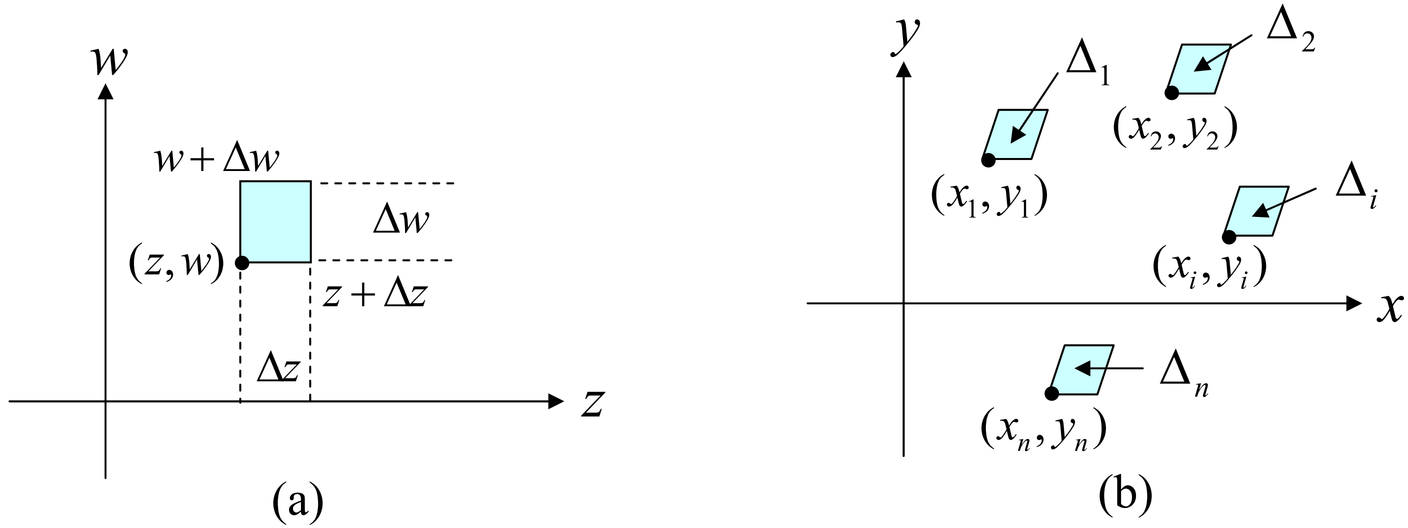


Fig. 9.3

Consider the problem of evaluating the probability

$$\begin{aligned} &P\left(z < Z \le z + \Delta z, w < W \le w + \Delta w\right) \\ &\quad = P\left(z < g(X,Y) \le z + \Delta z, w < h(X,Y) \le w + \Delta w\right). \end{aligned} \tag{9-15}$$

Using (7-9) we can rewrite (9-15) as

$$P\left(z < Z \le z + \Delta z, w < W \le w + \Delta w\right) = f_{ZW}(z, w)\Delta z \Delta w. \qquad (9\text{-}16)$$

But to translate this probability in terms of $f_{XY}(x, y)$, we need to evaluate the equivalent region for $\Delta z\, \Delta w$ in the xy plane. Towards this referring to Fig. 9.4, we observe that the point A with coordinates (z,w) gets mapped onto the point A' with coordinates (x_i, y_i) (as well as to other points as in Fig. 9.3(b)). As z changes to $z + \Delta z$ to point B in Fig. 9.4 (a), let B' represent its image in the xy plane. Similarly as w changes to $w + \Delta w$ to C, let C' represent its image in the xy plane.

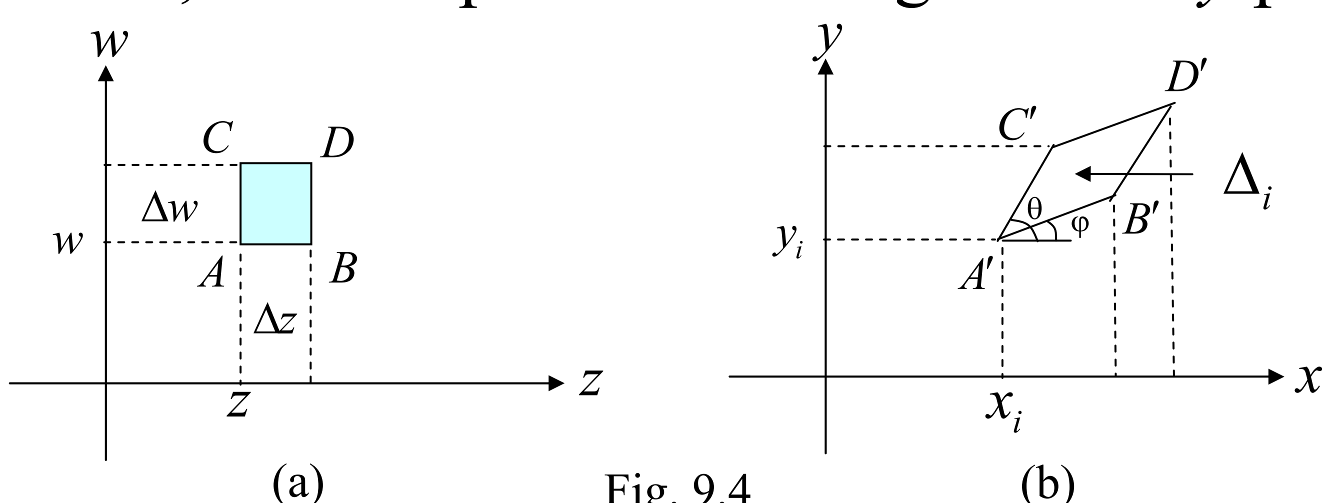


Fig. 9.4

Finally D goes to D', and $A'B'C'D'$ represents the equivalent parallelogram in the XY plane with area Δ_i. Referring back to Fig. 9.3, the probability in (9-16) can be alternatively expressed as

$$\sum_i P\big((X,Y)\in\Delta_i\big) = \sum_i f_{XY}(x_i, y_i)\Delta_i. \tag{9-17}$$

Equating (9-16) and (9-17) we obtain

$$f_{ZW}(z,w) = \sum_i f_{XY}(x_i, y_i)\frac{\Delta_i}{\Delta z \Delta w}. \tag{9-18}$$

To simplify (9-18), we need to evaluate the area Δ_i of the parallelograms in Fig. 9.3 (b) in terms of $\Delta z \Delta w$. Towards this, let g_1 and h_1 denote the inverse transformation in (9-14), so that

$$x_i = g_1(z,w), \qquad y_i = h_1(z,w). \tag{9-19}$$

As the point (z,w) goes to $(x_i, y_i) \equiv A'$, the point $(z+\Delta z, w) \to B'$, the point $(z, w+\Delta w) \to C'$, and the point $(z+\Delta z, w+\Delta w) \to D'$. Hence the respective x and y coordinates of B' are given by

$$g_1(z+\Delta z, w) = g_1(z,w) + \frac{\partial g_1}{\partial z}\Delta z = x_i + \frac{\partial g_1}{\partial z}\Delta z, \tag{9-20}$$

and

$$h_1(z+\Delta z, w) = h_1(z,w) + \frac{\partial h_1}{\partial z}\Delta z = y_i + \frac{\partial h_1}{\partial z}\Delta z. \tag{9-21}$$

Similarly those of C' are given by

$$x_i + \frac{\partial g_1}{\partial w}\Delta w, \quad y_i + \frac{\partial h_1}{\partial w}\Delta w. \tag{9-22}$$

The area of the parallelogram $A'B'C'D'$ in Fig. 9.4 (b) is given by

$$\begin{aligned} \Delta_i &= (A'B')(A'C')\sin(\theta - \varphi) \\ &= (A'B'\cos\varphi)(A'C'\sin\theta) - (A'B'\sin\varphi)(A'C'\cos\theta). \end{aligned} \tag{9-23}$$

But from Fig. 9.4 (b), and (9-20) - (9-22)

$$A'B'\cos\varphi = \frac{\partial g_1}{\partial z}\Delta z, \quad A'C'\sin\theta = \frac{\partial h_1}{\partial w}\Delta w, \tag{9-24}$$

$$A'B'\sin\varphi = \frac{\partial h_1}{\partial z}\Delta z, \quad A'C'\cos\theta = \frac{\partial g_1}{\partial w}\Delta w. \tag{9-25}$$

so that

$$\Delta_i = \left(\frac{\partial g_1}{\partial z}\frac{\partial h_1}{\partial w} - \frac{\partial g_1}{\partial w}\frac{\partial h_1}{\partial z} \right)\Delta z \Delta w \tag{9-26}$$

and

$$\frac{\Delta_i}{\Delta z \Delta w} = \left(\frac{\partial g_1}{\partial z}\frac{\partial h_1}{\partial w} - \frac{\partial g_1}{\partial w}\frac{\partial h_1}{\partial z} \right) = \det \begin{pmatrix} \frac{\partial g_1}{\partial z} & \frac{\partial g_1}{\partial w} \\ \frac{\partial h_1}{\partial z} & \frac{\partial h_1}{\partial w} \end{pmatrix} \tag{9-27}$$

The right side of (9-27) represents the Jacobian $J(z, w)$ of the transformation in (9-19). Thus

$$J(z,w) = \det \begin{pmatrix} \dfrac{\partial g_1}{\partial z} & \dfrac{\partial g_1}{\partial w} \\ \\ \dfrac{\partial h_1}{\partial z} & \dfrac{\partial h_1}{\partial w} \end{pmatrix}. \tag{9-28}$$

Substituting (9-27) - (9-28) into (9-18), we get

$$f_{ZW}(z,w) = \sum_i |J(z,w)|\, f_{XY}(x_i, y_i) = \sum_i \frac{1}{|J(x_i, y_i)|} f_{XY}(x_i, y_i), \tag{9-29}$$

since

$$|J(z,w)| = \frac{1}{|J(x_i, y_i)|} \tag{9-30}$$

where $J(x_i, y_i)$ represents the Jacobian of the original transformation in (9-13) given by

$$J(x_i, y_i) = \det \begin{pmatrix} \dfrac{\partial g}{\partial x} & \dfrac{\partial g}{\partial y} \\ \\ \dfrac{\partial h}{\partial x} & \dfrac{\partial h}{\partial y} \end{pmatrix}_{x = x_i,\, y = y_i}. \tag{9-31}$$

Next we shall illustrate the usefulness of the formula in (9-29) through various examples:

Example 9.2: Suppose X and Y are zero mean independent Gaussian r.vs with common variance σ^2.
Define $Z = \sqrt{X^2 + Y^2},\ W = \tan^{-1}(Y/X),$ where $|w| \le \pi/2.$
Obtain $f_{ZW}(z, w).$
Solution: Here

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}. \qquad (9\text{-}32)$$

Since

$$z = g(x, y) = \sqrt{x^2 + y^2}; w = h(x, y) = \tan^{-1}(y/x), \quad |w| \le \pi/2, \qquad (9\text{-}33)$$

if (x_1, y_1) is a solution pair so is $(-x_1, -y_1).$ From (9-33)

$$\frac{y}{x} = \tan w, \quad \text{or} \quad y = x \tan w. \qquad (9\text{-}34)$$

Substituting this into z, we get

$$z=\sqrt{x^2+y^2}=x\sqrt{1+\tan^2 w}=x\ \sec w, \quad \text{or} \quad x=z\ \cos w. \tag{9-35}$$

and

$$y = x\tan w = z\sin w. \tag{9-36}$$

Thus there are two solution sets

$$x_1 = z\cos w,\ \ y_1 = z\sin w,\ \ x_2 = -z\cos w,\ \ y_2 = -z\sin w. \tag{9-37}$$

We can use (9-35) - (9-37) to obtain $J(z,w)$. From (9-28)

$$J(z,w)=\begin{vmatrix} \dfrac{\partial x}{\partial z} & \dfrac{\partial x}{\partial w} \\[2ex] \dfrac{\partial y}{\partial z} & \dfrac{\partial y}{\partial w} \end{vmatrix} = \begin{vmatrix} \cos w & -z\sin w \\ \sin w & z\cos w \end{vmatrix} = z, \tag{9-38}$$

so that

$$|J(z,w)| = z. \tag{9-39}$$

We can also compute $J(x,y)$ using (9-31). From (9-33),

$$J(x,y) = \begin{vmatrix} \dfrac{x}{\sqrt{x^2+y^2}} & \dfrac{y}{\sqrt{x^2+y^2}} \\ \dfrac{-y}{x^2+y^2} & \dfrac{x}{x^2+y^2} \end{vmatrix} = \frac{1}{\sqrt{x^2+y^2}} = \frac{1}{z}. \qquad (9\text{-}40)$$

Notice that $|J(z,w)| = 1/|J(x_i,y_i)|$, agreeing with (9-30). Substituting (9-37) and (9-39) or (9-40) into (9-29), we get

$$f_{ZW}(z,w) = z\big(f_{XY}(x_1,y_1) + f_{XY}(x_2,y_2)\big)$$

$$= \frac{z}{\pi\sigma^2} e^{-z^2/2\sigma^2}, \qquad 0 < z < \infty, \quad |w| < \frac{\pi}{2}. \qquad (9\text{-}41)$$

Thus

$$f_Z(z) = \int_{-\pi/2}^{\pi/2} f_{ZW}(z,w)\,dw = \frac{z}{\sigma^2} e^{-z^2/2\sigma^2}, \qquad 0 < z < \infty, \qquad (9\text{-}42)$$

which represents a Rayleigh r.v with parameter σ^2, and

$$f_W(w) = \int_0^\infty f_{ZW}(z,w)\,dz = \frac{1}{\pi}, \quad |w| < \frac{\pi}{2}, \qquad (9\text{-}43)$$

which represents a uniform r.v in the interval $(-\pi/2, \pi/2)$. Moreover by direct computation

$$f_{ZW}(z,w) = f_Z(z) \cdot f_W(w) \tag{9-44}$$

implying that Z and W are independent. We summarize these results in the following statement: If X and Y are zero mean independent Gaussian random variables with common variance, then $\sqrt{X^2+Y^2}$ has a Rayleigh distribution and $\tan^{-1}(Y/X)$ has a uniform distribution. Moreover these two derived r.vs are statistically independent. Alternatively, with X and Y as independent zero mean r.vs as in (9-32), $X + jY$ represents a complex Gaussian r.v. But

$$X + jY = Ze^{jW}, \tag{9-45}$$

where Z and W are as in (9-33), except that for (9-45) to hold good on the entire complex plane we must have $-\pi < W < \pi$, and hence it follows that the magnitude and phase of a complex Gaussian r.v are independent with Rayleigh and uniform distributions $(U \sim (-\pi, \pi))$ respectively. The statistical independence of these derived r.vs is an interesting observation.

Example 9.3: Let X and Y be independent exponential random variables with common parameter λ. Define $U = X + Y$, $V = X - Y$. Find the joint and marginal p.d.f of U and V.
Solution: It is given that

$$f_{XY}(x, y) = \frac{1}{\lambda^2} e^{-(x+y)/\lambda}, \quad x > 0, \quad y > 0. \tag{9-46}$$

Now since $u = x + y$, $v = x - y$, always $|v| < u$, and there is only one solution given by

$$x = \frac{u+v}{2}, \quad y = \frac{u-v}{2}. \tag{9-47}$$

Moreover the Jacobian of the transformation is given by

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

and hence

$$f_{UV}(u,v) = \frac{1}{2\lambda^2} e^{-u/\lambda}, \qquad 0 < |v| < u < \infty, \tag{9-48}$$

represents the joint p.d.f of U and V. This gives

$$f_U(u) = \int_{-u}^{u} f_{UV}(u,v)\,dv = \frac{1}{2\lambda^2} \int_{-u}^{u} e^{-u/\lambda}\,dv = \frac{u}{\lambda^2} e^{-u/\lambda}, \quad 0 < u < \infty, \tag{9-49}$$

and

$$f_V(v) = \int_{|v|}^{\infty} f_{UV}(u,v)\,du = \frac{1}{2\lambda^2} \int_{|v|}^{\infty} e^{-u/\lambda}\,du = \frac{1}{2\lambda} e^{-|v|/\lambda}, \quad -\infty < v < \infty. \tag{9-50}$$

Notice that in this case the r.vs U and V are not independent.

As we show below, the general transformation formula in (9-29) making use of two functions can be made useful even when only one function is specified.

Auxiliary Variables:

Suppose

$$Z = g(X, Y), \tag{9-51}$$

where X and Y are two random variables. To determine $f_Z(z)$ by making use of the above formulation in (9-29), we can define an auxiliary variable

$$W = X \quad \text{or} \quad W = Y \tag{9-52}$$

and the p.d.f of Z can be obtained from $f_{ZW}(z, w)$ by proper integration.

Example 9.4: Suppose $Z = X + Y$ and let $W = Y$ so that the transformation is one-to-one and the solution is given by $y_1 = w, \ \ x_1 = z - w.$

The Jacobian of the transformation is given by

$$J(x,y) = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1$$

and hence

$$f_{ZW}(x,y) = f_{XY}(x_1,y_1) = f_{XY}(z-w,w)$$

or

$$f_Z(z) = \int f_{ZW}(z,w)dw = \int_{-\infty}^{+\infty} f_{XY}(z-w,w)dw, \qquad (9\text{-}53)$$

which agrees with (8.7). Note that (9-53) reduces to the convolution of $f_X(z)$ and $f_Y(z)$ if X and Y are independent random variables. Next, we consider a less trivial example.

Example 9.5: Let $X \sim U(0,1)$ and $Y \sim U(0,1)$ be independent. Define $Z = (-2 \ln X)^{1/2} \cos(2\pi Y)$. (9-54)

Find the density function of Z.

Solution: We can make use of the auxiliary variable $W = Y$ in this case. This gives the only solution to be

$$x_1 = e^{-(z\sec(2\pi w))^2/2}, \tag{9-55}$$

$$y_1 = w, \tag{9-56}$$

and using (9-28)

$$J(z,w) = \begin{vmatrix} \dfrac{\partial x_1}{\partial z} & \dfrac{\partial x_1}{\partial w} \\ \dfrac{\partial y_1}{\partial z} & \dfrac{\partial y_1}{\partial w} \end{vmatrix} = \begin{vmatrix} -z\sec^2(2\pi w)\ e^{-(z\sec(2\pi w))^2/2} & \dfrac{\partial x_1}{\partial w} \\ 0 & 1 \end{vmatrix} \tag{9-57}$$

$$= -z\sec^2(2\pi w) e^{-(z\sec(2\pi w))^2/2}.$$

Substituting (9-55) - (9-57) into (9-29), we obtain

$$f_{ZW}(z,w) = |z|\sec^2(2\pi\, w)\ \ e^{-(z\sec(2\pi\, w))^2/2}, \qquad -\infty < z < +\infty, \quad 0 < w < 1, \tag{9-58}$$

and

$$f_Z(z) = \int_0^1 f_{ZW}(z,w)dw = e^{-z^2/2}\int_0^1 |z|\sec^2(2\pi w)\ \ e^{-(|z|\tan(2\pi w))^2/2}dw. \quad (9\text{-}59)$$

Let $u = |z|\tan(2\pi w)$ so that $du = 2\pi\ |z|\sec^2(2\pi w)dw.$ Notice that as w varies from 0 to 1, u varies from $-\infty$ to $+\infty$. Using this in (9-59), we get

$$f_Z(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}\underbrace{\int_{-\infty}^{+\infty} e^{-u^2/2}\frac{du}{\sqrt{2\pi}}}_{1} = \frac{1}{\sqrt{2\pi}}e^{-z^2/2},\quad -\infty < z < \infty, \qquad (9\text{-}60)$$

which represents a zero mean Gaussian r.v with unit variance. Thus $Z \sim N(0,1)$. Equation (9-54) can be used as a practical procedure to generate Gaussian random variables from two independent uniformly distributed random sequences.

Example 9.6 : Let X and Y be independent identically distributed Geometric random variables with

$$P(X=k)=P(Y=K)=pq^k, \qquad k=0,1,2,\cdots.$$

(a) Show that $min\,(X\,,\,Y\,)$ and $X-Y$ are independent random variables.
(b) Show that $min\,(X\,,\,Y\,)$ and $max\,(X\,,\,Y\,)-min\,(X\,,\,Y\,)$ are also independent random variables.

Solution: (a) Let

$$Z=\min\,(X\,,\,Y\,)\,,\ \text{and}\ \ W=X-Y. \tag{9-61}$$

Note that Z takes only nonnegative values $\{0,1,2,\cdots\}$, while W takes both positive, zero and negative values $\{0,\pm 1,\pm 2,\cdots\}$. We have $P(Z=m,\ \ W=n)\ =P\{\min\,(X\,,\,Y\,)=m,\ \ X-Y\ =n\}$. But

$$Z=\min(X,Y)=\begin{cases} Y & X\ge Y \Rightarrow W=X-Y \ \text{ is nonnegative} \\ X & X<Y \Rightarrow W=X-Y \ \text{ is negative.} \end{cases}$$

Thus

$$\begin{aligned} P(Z=m,W=n) &= P\{\min(X,Y)=m, X-Y=n, (X\ge Y \cup X<Y)\} \\ &= P(\min(X,Y)=m, X-Y=n, X\ge Y) \\ &\quad + P(\min(X,Y)=m, X-Y=n, X<Y) \end{aligned} \tag{9-62}$$

$$
\begin{aligned}
P(Z=m, W=n) &= P(Y=m, X=m+n, X \geq Y) \\
&\quad + P(X=m, Y=m-n, X<Y) \\
&= \begin{cases} P(X=m+n)P(Y=m) = pq^{m+n}pq^{m}, m \geq 0, n \geq 0 \\ P(X=m)P(Y=m-n) = pq^{m}pq^{m-n}, m \geq 0, n < 0 \end{cases} \\
&= p^2 q^{2m+|n|}, \quad m = 0, 1, 2, \cdots \quad n = 0, \pm 1, \pm 2, \cdots
\end{aligned} \tag{9-63}
$$

represents the joint probability mass function of the random variables Z and W. Also

$$
\begin{aligned}
P(Z=m) = \sum_n P(Z=m, W=n) &= \sum_n p^2 q^{2m} q^{|n|} \\
&= p^2 q^{2m}(1+2q+2q^2+\cdots) \\
&= p^2 q^{2m}(1+\tfrac{2q}{1-q}) = pq^{2m}(1+q) \\
&= p(1+q)q^{2m}, \quad m = 0, 1, 2, \cdots.
\end{aligned} \tag{9-64}
$$

Thus Z represents a Geometric random variable since $1-q^2 = p(1+q)$, and

$$P(W=n)=\sum_{m=0}^{\infty} P(Z=m, W=n)=\sum_{m=0}^{\infty} p^2 q^{2m} q^{|n|}$$

$$=p^2 q^{|n|}(1+q^2+q^4+\cdots)=p^2 q^{|n|}\frac{1}{1-q^2}$$

$$=\frac{p}{1+q} q^{|n|}, \qquad n=0, \ \pm 1, \ \pm 2, \cdots. \tag{9-65}$$

Note that

$$P(Z=m, W=n)=P(Z=m)P(W=n), \tag{9-66}$$

establishing the independence of the random variables Z and W. The independence of $X-Y$ and $min\,(X, Y)$ when X and Y are independent Geometric random variables is an interesting observation.

(b) Let

$$Z=min\,(X, Y), \quad R=max\,(X, Y)-min\,(X, Y). \tag{9-67}$$

In this case both Z and R take nonnegative integer values $0, 1, 2, \cdots$. Proceeding as in (9-62)-(9-63) we get

$$
\begin{aligned}
P\{Z=m,R=n\} &= P\{\min(X,Y)=m, \max(X,Y)-\min(X,Y)=n, X\ge Y\} \\
&\quad + P\{\min(X,Y)=m, \max(X,Y)-\min(X,Y)=n, X<Y\} \\
&= P\{Y=m, X=m+n, X\ge Y) + P(X=m, Y=m+n, X<Y\} \\
&= P\{X=m+n, Y=m, X\ge Y) + P(X=m, Y=m+n, X<Y\} \\
&= \begin{cases} pq^{m+n}pq^{m} + pq^{m}pq^{m+n}, & m=0,1,2,\cdots, \quad n=1,2,\cdots \\ pq^{m+n}pq^{m}, & m=0,1,2,\cdots, \quad n=0 \end{cases} \\
&= \begin{cases} 2p^2q^{2m+n}, & m=0,1,2,\cdots, \quad n=1,2,\cdots \\ p^2q^{2m}, & m=0,1,2,\cdots, \quad n=0. \end{cases} \qquad (9\text{-}68)
\end{aligned}
$$

Eq. (9-68) represents the joint probability mass function of Z and R in (9-67). From (9-68),

$$
\begin{aligned}
P(Z=m) &= \sum_{n=0}^{\infty} P\{Z=m,R=n\} = p^2q^{2m}(1+2\sum_{n=1}^{\infty} q^n) = p^2q^{2m}\left(1+\frac{2q}{p}\right) \\
&= p(1+q)q^{2m}, \qquad m=0,1,2,\cdots \qquad (9\text{-}69)
\end{aligned}
$$

and

$$P(R=n)=\sum_{m=0}^{\infty}P\{Z=m,R=n\}=\begin{cases}\dfrac{p}{1+q}, & n=0\\[2mm] \dfrac{2p}{1+q}q^n, & n=1,2,\cdots.\end{cases} \tag{9-70}$$

From (9-68)-(9-70), we get

$$P(Z=m,R=n)=P(Z=m)P(R=n) \tag{9-71}$$

which proves the independence of the random variables Z and R defined in (9-67) as well.

10. Joint Moments and Joint Characteristic Functions

Following section 6, in this section we shall introduce various parameters to compactly represent the information contained in the joint p.d.f of two r.vs. Given two r.vs X and Y and a function $g(x, y)$, define the r.v

$$Z = g(X, Y) \tag{10-1}$$

Using (6-2), we can define the mean of Z to be

$$\mu_Z = E(Z) = \int_{-\infty}^{+\infty} z\, f_Z(z)\, dz. \tag{10-2}$$

However, the situation here is similar to that in (6-13), and it is possible to express the mean of $Z = g(X,Y)$ in terms of $f_{XY}(x,y)$ *without* computing $f_Z(z)$. To see this, recall from (5-26) and (7-10) that

$$P\big(z < Z \le z + \Delta z\big) = f_Z(z)\Delta z = P\big(z < g(X,Y) \le z + \Delta z\big)$$
$$= \sum \sum_{(x,y)\in D_{\Delta z}} f_{XY}(x,y)\Delta x \Delta y \qquad (10\text{-}3)$$

where $D_{\Delta z}$ is the region in *xy* plane satisfying the above inequality. From (10-3), we get

$$z\, f_Z(z)\Delta z = g(x,y) \sum \sum_{(x,y)\in D_{\Delta z}} f_{XY}(x,y)\Delta x \Delta y. \qquad (10\text{-}4)$$

As Δz covers the entire *z* axis, the corresponding regions $D_{\Delta z}$ are nonoverlapping, and they cover the entire *xy* plane.

By integrating (10-4), we obtain the useful formula

$$E(Z) = \int_{-\infty}^{+\infty} z\, f_Z(z)dz = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y) f_{XY}(x,y)dxdy. \qquad (10\text{-}5)$$

or

$$E[g(X,Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y) f_{XY}(x,y)dxdy. \qquad (10\text{-}6)$$

If X and Y are discrete-type r.vs, then

$$E[g(X,Y)] = \sum_i \sum_j g(x_i, y_j) P(X = x_i, Y = y_j). \qquad (10\text{-}7)$$

Since expectation is a linear operator, we also get

$$E\left(\sum_k a_k g_k(X,Y) \right) = \sum_k a_k E[g_k(X,Y)]. \qquad (10\text{-}8)$$

If X and Y are independent r.vs, it is easy to see that $Z = g(X)$ and $W = h(Y)$ are always independent of each other. In that case using (10-7), we get the interesting result

$$E[g(X)h(Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x)h(y)f_X(x)f_Y(y)dxdy$$

$$= \int_{-\infty}^{+\infty} g(x)f_X(x)dx \int_{-\infty}^{+\infty} h(y)f_Y(y)dy = E[g(X)]E[h(Y)]. \quad (10\text{-}9)$$

However (10-9) is in general not true (if X and Y are not independent).

In the case of one random variable (see (10- 6)), we defined the parameters mean and variance to represent its average behavior. How does one parametrically represent similar cross-behavior between two random variables? Towards this, we can generalize the variance definition given in (6-16) as shown below:

Covariance: Given any two r.vs X and Y, define

$$Cov(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right]. \tag{10-10}$$

By expanding and simplifying the right side of (10-10), we also get

$$Cov(X,Y) = E(XY) - \mu_X\mu_Y = E(XY) - E(X)E(Y)$$

$$= \overline{XY} - \overline{X}\,\overline{Y}. \tag{10-11}$$

It is easy to see that

$$\left|Cov(X,Y)\right| \le \sqrt{Var(X)Var(Y)}. \tag{10-12}$$

To see (10-12), let $U = aX + Y$, so that

$$Var(U) = E\left[\left\{a(X-\mu_X) + (Y-\mu_Y)\right\}^2\right]$$

$$= a^2 Var(X) + 2a\,Cov(X,Y) + Var(Y) \ge 0. \tag{10-13}$$

The right side of (10-13) represents a quadratic in the variable *a* that has no distinct real roots (Fig. 10.1). Thus the roots are imaginary (or double) and hence the discriminant

$$[Cov(X,Y)]^2 - Var(X)\,Var(Y)$$

must be non-positive, and that gives (10-12). Using (10-12), we may define the normalized parameter

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}, \quad -1 \le \rho_{XY} \le 1, \qquad (10\text{-}14)$$

or

$$Cov(X,Y) = \rho_{XY}\sigma_X\sigma_Y \qquad (10\text{-}15)$$

and it represents the correlation coefficient between X and Y.

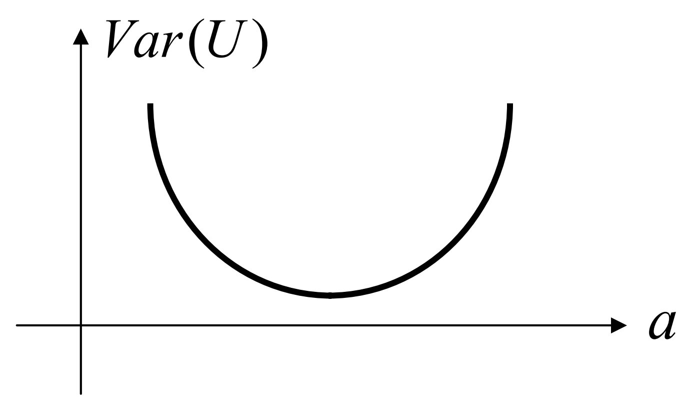


Fig. 10.1

Uncorrelated r.vs: If $\rho_{XY} = 0,$ then X and Y are said to be uncorrelated r.vs. From (11), if X and Y are uncorrelated, then

$$E(XY) = E(X)E(Y). \tag{10-16}$$

Orthogonality: X and Y are said to be orthogonal if

$$E(XY) = 0. \tag{10-17}$$

From (10-16) - (10-17), if either X or Y has zero mean, then orthogonality implies uncorrelatedness also and vice-versa. Suppose X and Y are independent r.vs. Then from (10-9) with $g(X) = X,\ h(Y) = Y,$ we get

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

and together with (10-16), we conclude that the random variables are uncorrelated, thus justifying the original definition in (10-10). Thus independence implies uncorrelatedness.

Naturally, if two random variables are statistically independent, then there cannot be any correlation between them $(\rho_{XY} = 0)$. However, the converse is in general not true. As the next example shows, random variables can be uncorrelated without being independent.

Example 10.1: Let $X \sim U(0,1)$, $Y \sim U(0,1)$. Suppose X and Y are independent. Define $Z = X + Y$, $W = X - Y$. Show that Z and W are dependent, but uncorrelated r.vs.

Solution: $z = x + y$, $w = x - y$ gives the only solution set to be

$$x = \frac{z+w}{2}, \quad y = \frac{z-w}{2}.$$

Moreover $0 < z < 2$, $-1 < w < 1$, $z + w \le 2$, $z - w \le 2$, $z > |w|$ and $|J(z,w)| = 1/2$.

Thus (see the shaded region in Fig. 10.2)

$$f_{ZW}(z,w) = \begin{cases} 1/2, & 0 < z < 2,\ -1 < w < 1,\ z + w \le 2,\ z - w \le 2,\ |w| < z, \\ 0, & \text{otherwise}, \end{cases} \tag{10-18}$$

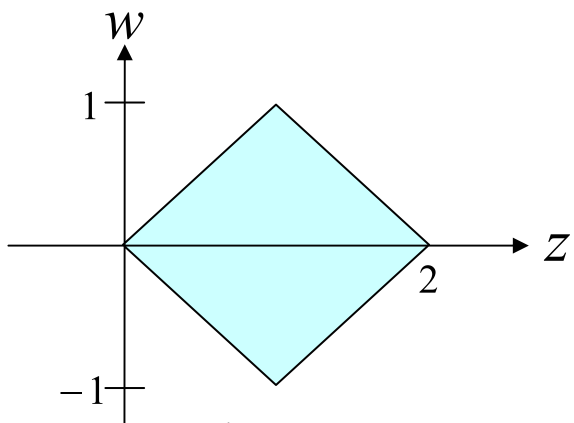


Fig. 10.2

and hence

$$f_Z(z) = \int f_{ZW}(z,w)\,dw = \begin{cases} \int_{-z}^{z} \frac{1}{2}\,dw = z, & 0 < z < 1, \\ \int_{z-2}^{2-z} \frac{1}{2}\,dw = 2 - z, & 1 < z < 2, \end{cases}$$

or by direct computation ($Z = X + Y$)

$$f_Z(z) = f_X(z) \otimes f_Y(z) = \begin{cases} z, & 0 < z < 1, \\ 2 - z, & 1 < z < 2, \\ 0, & \text{otherwise}, \end{cases} \tag{10-19}$$

and

$$f_W(w) = \int f_{ZW}(z,w)dz = \int_{|w|}^{2-|w|} \frac{1}{2}\, dz = \begin{cases} 1 - |\,w\,|, & -1 < w < 1, \\ 0, & \text{otherwise.} \end{cases} \tag{10-20}$$

Clearly $f_{ZW}(z,w) \neq f_Z(z) f_W(w)$. Thus Z and W are not independent. However

$$E(ZW) = E\big[(X+Y)(X-Y)\big] = E(X^2) - E(Y^2) = 0, \tag{10-21}$$

and

$$E(W) = E(X - Y) = 0,$$

and hence

$$Cov(Z,W) = E(ZW) - E(Z)E(W) = 0 \tag{10-22}$$

implying that Z and W are uncorrelated random variables.

Example 10.2: Let $Z = aX + bY$. Determine the variance of Z in terms of σ_X, σ_Y and ρ_{XY}.

Solution:

$$\mu_Z = E(Z) = E(aX + bY) = a\mu_X + b\mu_Y$$

and using (10-15)

$$\begin{aligned}\sigma_Z^2 = Var(Z) &= E\left[(Z-\mu_Z)^2\right] = E\left[\left(a(X-\mu_X) + b(Y-\mu_Y)\right)^2\right] \\ &= a^2 E(X-\mu_X)^2 + 2abE\left((X-\mu_X)(Y-\mu_Y)\right) + b^2 E(Y-\mu_Y)^2 \\ &= a^2\sigma_X^2 + 2ab\rho_{XY}\sigma_X\sigma_Y + b^2\sigma_Y^2. \end{aligned} \tag{10-23}$$

In particular if X and Y are independent, then $\rho_{XY} = 0$, and (10-23) reduces to

$$\sigma_Z^2 = a^2\sigma_X^2 + b^2\sigma_Y^2. \tag{10-24}$$

Thus the variance of the sum of independent r.vs is the sum of their variances $(a = b = 1)$.

Moments:

$$E[X^k Y^m] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^k y^m f_{XY}(x,y)dx\ dy, \tag{10-25}$$

represents the joint moment of order (k,m) for X and Y.

Following the one random variable case, we can define the joint characteristic function between two random variables which will turn out to be useful for moment calculations.

Joint characteristic functions:

The joint characteristic function between X and Y is defined as

$$\Phi_{XY}(u,v) = E\left(e^{j(Xu+Yv)}\right) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{j(Xu+Yv)} f_{XY}(x,y)dxdy. \tag{10-26}$$

Note that $|\Phi_{XY}(u,v)| \le \Phi_{XY}(0,0) = 1.$

It is easy to show that

$$E(XY) = \frac{1}{j^2} \frac{\partial^2 \Phi_{XY}(u,v)}{\partial u \partial v}\bigg|_{u=0,v=0}. \tag{10-27}$$

If X and Y are independent r.vs, then from (10-26), we obtain

$$\Phi_{XY}(u,v) = E(e^{juX})E(e^{jvY}) = \Phi_X(u)\Phi_Y(v). \tag{10-28}$$

Also

$$\Phi_X(u) = \Phi_{XY}(u,0), \quad \Phi_Y(v) = \Phi_{XY}(0,v). \tag{10-29}$$

More on Gaussian r.vs :

From Lecture 7, X and Y are said to be jointly Gaussian as $N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$, if their joint p.d.f has the form in (7-23). In that case, by direct substitution and simplification, we obtain the joint characteristic function of two jointly Gaussian r.vs to be

$$\Phi_{XY}(u,v) = E(e^{j(Xu+Yv)}) = e^{j(\mu_X u+\mu_Y v)-\frac{1}{2}(\sigma_X^2 u^2 + 2\rho\sigma_X\sigma_Y uv + \sigma_Y^2 v^2)}. \qquad (10\text{-}30)$$

Equation (10-14) can be used to make various conclusions. Letting $v = 0$ in (10-30), we get

$$\Phi_X(u) = \Phi_{XY}(u,0) = e^{j\mu_X u - \frac{1}{2}\sigma_X^2 u^2}, \qquad (10\text{-}31)$$

and it agrees with (6-47).

From (7-23) by direct computation using (10-11), it is easy to show that for two jointly Gaussian random variables

$$Cov(X,Y) = \rho\,\sigma_X\sigma_Y.$$

Hence from (10-14), ρ in $N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$ represents the actual correlation coefficient of the two jointly Gaussian r.vs in (7-23). Notice that $\rho = 0$ implies

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

Thus if X and Y are jointly Gaussian, uncorrelatedness does imply independence between the two random variables. Gaussian case is the only exception where the two concepts imply each other.

Example 10.3: Let X and Y be jointly Gaussian r.vs with parameters $N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$. Define $Z = aX + bY$. Determine $f_Z(z)$.

Solution: In this case we can make use of characteristic function to solve this problem.

$$\begin{aligned} \Phi_Z(u) &= E(e^{jZu}) = E(e^{j(aX+bY)u}) = E(e^{jauX + jbuY}) \\ &= \Phi_{XY}(au, bu). \end{aligned} \qquad (10\text{-}32)$$

From (10-30) with u and v replaced by au and bu respectively we get

$$\Phi_Z(u) = e^{j(a\mu_X + b\mu_Y)u - \frac{1}{2}(a^2\sigma_X^2 + 2\rho ab\sigma_X\sigma_Y + b^2\sigma_Y^2)u^2} = e^{j\mu_Z u - \frac{1}{2}\sigma_Z^2 u^2}, \tag{10-33}$$

where

$$\mu_Z \triangleq a\mu_X + b\mu_Y, \tag{10-34}$$

$$\sigma_Z^2 \triangleq a^2\sigma_X^2 + 2\rho ab\sigma_X\sigma_Y + b^2\sigma_Y^2. \tag{10-35}$$

Notice that (10-33) has the same form as (10-31), and hence we conclude that $Z = aX + bY$ is also Gaussian with mean and variance as in (10-34) - (10-35), which also agrees with (10-23).

From the previous example, we conclude that any linear combination of jointly Gaussian r.vs generate a Gaussian r.v.

In other words, linearity preserves Gaussianity. We can use the characteristic function relation to conclude an even more general result.

Example 10.4: Suppose X and Y are jointly Gaussian r.vs as in the previous example. Define two linear combinations

$$Z = aX + bY, \quad W = cX + dY. \tag{10-36}$$

what can we say about their joint distribution?
Solution: The characteristic function of Z and W is given by

$$\begin{aligned}\Phi_{ZW}(u,v) &= E(e^{j(Zu+Wv)}) = E(e^{j(aX+bY)u+j(cX+dY)v}) \\ &= E(e^{jX(au+cv)+jY(bu+dv)}) = \Phi_{XY}(au+cv, bu+dv).\end{aligned} \tag{10-37}$$

As before substituting (10-30) into (10-37) with u and v replaced by $au + cv$ and $bu + dv$ respectively, we get

$$\Phi_{ZW}(u,v) = e^{j(\mu_Z u+\mu_W v)-\frac{1}{2}(\sigma_Z^2 u^2+2\rho_{ZW}\sigma_X\sigma_Y uv+\sigma_W^2 v^2)}, \tag{10-38}$$

where

$$\mu_Z = a\mu_X + b\mu_Y, \tag{10-39}$$

$$\mu_W = c\mu_X + d\mu_Y, \tag{10-40}$$

$$\sigma_Z^2 = a^2\sigma_X^2 + 2ab\rho\sigma_X\sigma_Y + b^2\sigma_Y^2, \tag{10-41}$$

$$\sigma_W^2 = c^2\sigma_X^2 + 2cd\rho\sigma_X\sigma_Y + d^2\sigma_Y^2, \tag{10-42}$$

and

$$\rho_{ZW} = \frac{ac\sigma_X^2 + (ad+bc)\rho\sigma_X\sigma_Y + bd\sigma_Y^2}{\sigma_Z\sigma_W}. \tag{10-43}$$

From (10-38), we conclude that Z and W are also jointly distributed Gaussian r.vs with means, variances and correlation coefficient as in (10-39) - (10-43).

To summarize, any two linear combinations of jointly Gaussian random variables (independent or dependent) are also jointly Gaussian r.vs.

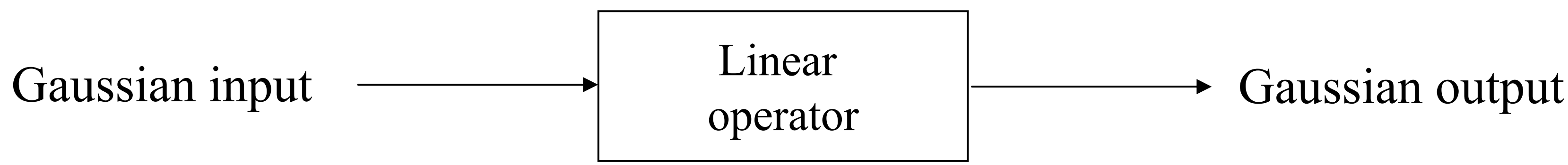


Fig. 10.3

Of course, we could have reached the same conclusion by deriving the joint p.d.f $f_{ZW}(z, w)$ using the technique developed in section 9 (refer (7-29)).

Gaussian random variables are also interesting because of the following result:

Central Limit Theorem: Suppose $X_1, X_2, \cdots, X_n$ are a set of zero mean independent, identically distributed (i.i.d) random variables with some common distribution. Consider their scaled sum

$$Y = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}. \tag{10-44}$$

Then asymptotically (as $n \to \infty$)

$$Y \to N(0,\sigma^2). \tag{10-45}$$

Proof: Although the theorem is true under even more general conditions, we shall prove it here under the independence assumption. Let σ^2 represent their common variance. Since

$$E(X_i) = 0, \tag{10-46}$$

we have

$$Var(X_i) = E(X_i^2) = \sigma^2. \tag{10-47}$$

Consider

$$\Phi_Y(u) = E(e^{jYu}) = E\left(e^{j(X_1+X_2+\cdots+X_n)u/\sqrt{n}}\right) = \prod_{i=1}^{n} E(e^{jX_i u/\sqrt{n}})$$

$$= \prod_{i=1}^{n} \Phi_{X_i}(u/\sqrt{n}) \tag{10-48}$$

where we have made use of the independence of the r.vs $X_1, X_2, \cdots, X_n$. But

$$E(e^{jX_i u/\sqrt{n}}) = E\left(1 - \frac{jX_i u}{\sqrt{n}} + \frac{j^2 X_i^2 u^2}{2!\,n} + \frac{j^3 X_i^3 u^3}{3!\,n^{3/2}} + \cdots\right) = 1 - \frac{\sigma^2 u^2}{2n} + o\left(\frac{1}{n^{3/2}}\right), \tag{10-49}$$

where we have made use of (10-46) - (10-47). Substituting (10-49) into (10-48), we obtain

$$\Phi_Y(u) = \left[1 - \frac{\sigma^2 u^2}{2n} + o\left(\frac{1}{n^{3/2}}\right)\right]^n, \tag{10-50}$$

and as

$$\lim_{n\to\infty} \Phi_Y(u) \to e^{-\sigma^2 u^2/2}, \tag{10-51}$$

since

$$\lim_{n\to\infty}\left(1-\frac{x}{n}\right)^n \to e^{-x}. \tag{10-52}$$

[Note that $o(1/n^{3/2})$ terms in (10-50) decay faster than $1/n^{3/2}$]. But (10-51) represents the characteristic function of a zero mean normal r.v with variance σ^2 and (10-45) follows.

The central limit theorem states that a large sum of independent random variables each with finite variance tends to behave like a normal random variable. Thus the individual p.d.fs become unimportant to analyze the collective sum behavior. If we model the noise phenomenon as the sum of a large number of independent random variables (eg: electron motion in resistor components), then this theorem allows us to conclude that noise behaves like a Gaussian r.v.

It may be remarked that the finite variance assumption is necessary for the theorem to hold good. To prove its importance, consider the r.vs to be Cauchy distributed, and let

$$Y = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}. \tag{10-53}$$

where each $X_i \sim C(\alpha)$. Then since

$$\Phi_{X_i}(u) = e^{-\alpha|u|}, \tag{10-54}$$

substituting this into (10-48), we get

$$\Phi_Y(u) = \prod_{i=1}^{n} \Phi_X(u/\sqrt{n}) = \left(e^{-\alpha|u|/\sqrt{n}}\right)^{n} \sim C(\alpha\sqrt{n}), \tag{10-55}$$

which shows that Y is still Cauchy with parameter $\alpha\sqrt{n}$. In other words, central limit theorem doesn't hold good for a set of Cauchy r.vs as their variances are undefined.

Joint characteristic functions are useful in determining the p.d.f of linear combinations of r.vs. For example, with X and Y as independent Poisson r.vs with parameters λ_1 and λ_2 respectively, let

$$Z = X + Y. \tag{10-56}$$

Then

$$\Phi_Z(u) = \Phi_X(u)\Phi_Y(u). \tag{10-57}$$

But from (6-33)

$$\Phi_X(u) = e^{\lambda_1(e^{ju}-1)}, \quad \Phi_Y(u) = e^{\lambda_2(e^{ju}-1)} \tag{10-58}$$

so that

$$\Phi_Z(u) = e^{(\lambda_1+\lambda_2)(e^{ju}-1)} \quad \sim \quad P(\lambda_1 + \lambda_2) \tag{10-59}$$

i.e., sum of independent Poisson r.vs is also a Poisson random variable.

11. Conditional Density Functions and Conditional Expected Values

As we have seen in section 4 conditional probability density functions are useful to update the information about an event based on the knowledge about some other related event (refer to example 4.7). In this section, we shall analyze the situation where the related event happens to be a random variable that is dependent on the one of interest.

From (4-11), recall that the distribution function of X given an event B is

$$F_X(x \mid B) = P\big(X(\xi) \le x \mid B\big) = \frac{P\big((X(\xi) \le x) \cap B\big)}{P(B)}. \qquad (11\text{-}1)$$

Suppose, we let

$$B = \{ y_1 < Y(\xi) \le y_2 \}. \tag{11-2}$$

Substituting (11-2) into (11-1), we get

$$F_X(x \mid y_1 < Y \le y_2) = \frac{P\big(X(\xi) \le x, y_1 < Y(\xi) \le y_2\big)}{P(y_1 < Y(\xi) \le y_2)}$$

$$= \frac{F_{XY}(x, y_2) - F_{XY}(x, y_1)}{F_Y(y_2) - F_Y(y_1)}, \tag{11-3}$$

where we have made use of (7-4). But using (3-28) and (7-7) we can rewrite (11-3) as

$$F_X(x \mid y_1 < Y \le y_2) = \frac{\int_{-\infty}^{x} \int_{y_1}^{y_2} f_{XY}(u, v)\, du dv}{\int_{y_1}^{y_2} f_Y(v)\, dv}. \tag{11-4}$$

To determine, the limiting case $F_X(x \mid Y = y)$, we can let $y_1 = y$ and $y_2 = y + \Delta y$ in (11-4).

This gives

$$F_X(x \mid y < Y \le y + \Delta y) = \frac{\int_{-\infty}^{x}\int_{y}^{y+\Delta y} f_{XY}(u,v)\,du\,dv}{\int_{y}^{y+\Delta y} f_Y(v)\,dv} \approx \frac{\int_{-\infty}^{x} f_{XY}(u,y)\,du\,\Delta y}{f_Y(y)\,\Delta y} \qquad (11\text{-}5)$$

and hence in the limit

$$F_X(x \mid Y = y) = \lim_{\Delta y \to 0} F_X(x \mid y < Y \le y + \Delta y) = \frac{\int_{-\infty}^{x} f_{XY}(u,y)\,du}{f_Y(y)}. \qquad (11\text{-}6)$$

(To remind about the conditional nature on the left hand side, we shall use the subscript $X \mid Y$ (instead of X) there). Thus

$$F_{X|Y}(x \mid Y = y) = \frac{\int_{-\infty}^{x} f_{XY}(u,y)\,du}{f_Y(y)}. \qquad (11\text{-}7)$$

Differentiating (11-7) with respect to x using (8-7), we get

$$f_{X|Y}(x \mid Y = y) = \frac{f_{XY}(x,y)}{f_Y(y)}. \qquad (11\text{-}8)$$

It is easy to see that the left side of (11-8) represents a valid probability density function. In fact

$$f_{X|Y}(x \mid Y = y) = \frac{f_{XY}(x, y)}{f_Y(y)} \geq 0 \tag{11-9}$$

and

$$\int_{-\infty}^{+\infty} f_{X|Y}(x \mid Y = y)dx = \frac{\int_{-\infty}^{+\infty} f_{XY}(x, y)dx}{f_Y(y)} = \frac{f_Y(y)}{f_Y(y)} = 1, \tag{11-10}$$

where we have made use of (7-14). From (11-9) - (11-10), (11-8) indeed represents a valid p.d.f, and we shall refer to it as the conditional p.d.f of the r.v X given $Y = y$. We may also write

$$f_{X|Y}(x \mid Y = y) = f_{X|Y}(x \mid y). \tag{11-11}$$

From (11-8) and (11-11), we have

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x, y)}{f_Y(y)}, \tag{11-12}$$

and similarly

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x, y)}{f_X(x)}. \tag{11-13}$$

If the r.vs X and Y are independent, then $f_{XY}(x, y) = f_X(x) f_Y(y)$ and (11-12) - (11-13) reduces to

$$f_{X|Y}(x \mid y) = f_X(x), \qquad f_{Y|X}(y \mid x) = f_Y(y), \tag{11-14}$$

implying that the conditional p.d.fs coincide with their unconditional p.d.fs. This makes sense, since if X and Y are independent r.vs, information about Y shouldn't be of any help in updating our knowledge about X.

In the case of discrete-type r.vs, (11-12) reduces to

$$P\left(X = x_i \mid Y = y_j\right) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}. \tag{11-15}$$

Next we shall illustrate the method of obtaining conditional p.d.fs through an example.

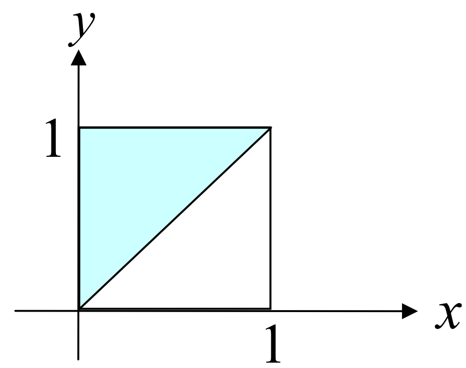


Fig. 11.1

Example 11.1: Given

$$f_{XY}(x, y) = \begin{cases} k, & 0 < x < y < 1, \\ 0, & \text{otherwise} \;, \end{cases} \qquad (11\text{-}16)$$

determine $f_{X|Y}(x \mid y)$ and $f_{Y|X}(y \mid x)$.

Solution: The joint p.d.f is given to be a constant in the shaded region. This gives

$$\iint f_{XY}(x, y)dxdy = \int_0^1 \int_0^y k \; dx \; dy = \int_0^1 k \; y \; dy = \frac{k}{2} = 1 \Rightarrow k = 2.$$

Similarly

$$f_X(x) = \int f_{XY}(x, y)dy = \int_x^1 k \; dy = k\,(1 - x), \quad 0 < x < 1, \qquad (11\text{-}17)$$

and

$$f_Y(y) = \int f_{XY}(x, y)dx = \int_0^y k \; dx = k \; y, \quad 0 < y < 1. \qquad (11\text{-}18)$$

From (11-16) - (11-18), we get

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x, y)}{f_Y(y)} = \frac{1}{y}, \qquad 0 < x < y < 1, \tag{11-19}$$

and

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x, y)}{f_X(x)} = \frac{1}{1-x}, \qquad 0 < x < y < 1. \tag{11-20}$$

We can use (11-12) - (11-13) to derive an important result. From there, we also have

$$f_{XY}(x, y) = f_{X|Y}(x \mid y) f_Y(y) = f_{Y|X}(y \mid x) f_X(x) \tag{11-21}$$

or

$$f_{Y|X}(y \mid x) = \frac{f_{X|Y}(x \mid y) f_Y(y)}{f_X(x)}. \tag{11-22}$$

But

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy = \int_{-\infty}^{+\infty} f_{X|Y}(x \mid y) f_Y(y) dy \tag{11-23}$$

and using (11-23) in (11-22), we get

$$f_{Y|X}(y \mid x) = \frac{f_{X|Y}(x \mid y) f_Y(y)}{\int_{-\infty}^{+\infty} f_{X|Y}(x \mid y) f_Y(y) dy}. \tag{24}$$

Equation (11-24) represents the p.d.f version of Bayes' theorem. To appreciate the full significance of (11-24), one need to look at communication problems where observations can be used to update our knowledge about unknown parameters. We shall illustrate this using a simple example.

Example 11.2: An unknown random phase θ is uniformly distributed in the interval $(0,2\pi)$, and $r = \theta + n$, where $n \sim N(0,\sigma^2)$. Determine $f(\theta \mid r)$.

Solution: Initially almost nothing about the r.v θ is known, so that we assume its a-priori p.d.f to be uniform in the interval $(0,2\pi)$.

In the equation $r = \theta + n,$ we can think of n as the noise contribution and r as the observation. It is reasonable to assume that θ and n are independent. In that case

$$f(r \mid \boldsymbol{\theta} = \theta) \sim N(\theta, \sigma^2) \tag{11-25}$$

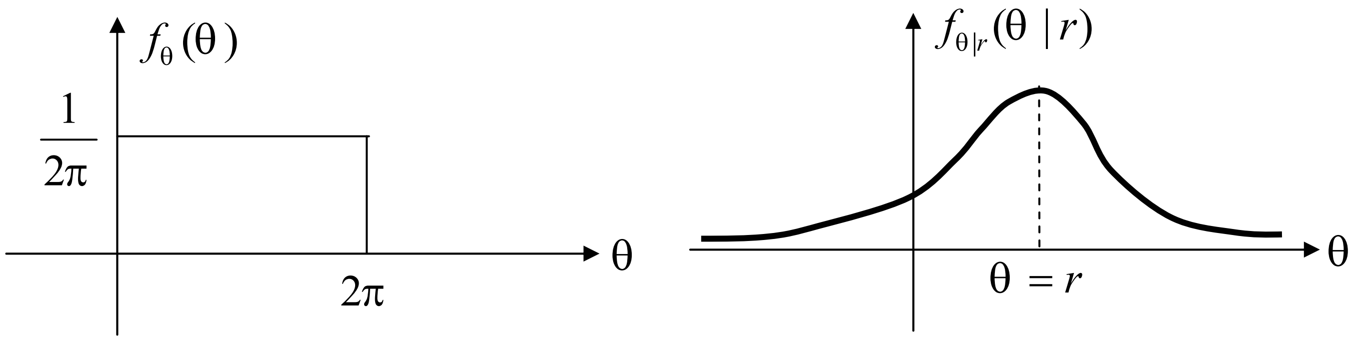
since it is given that $\boldsymbol{\theta} = \theta$ is a constant, $r = \theta + n$ behaves like n. Using (11-24), this gives the a-posteriori p.d.f of θ given r to be (see Fig. 11.2 (b))

$$\begin{aligned} f(\theta \mid r) &= \frac{f(r \mid \theta) f_\theta(\theta)}{\int_0^{2\pi} f(r \mid \theta) f_\theta(\theta) d\theta} = \frac{e^{-(r-\theta)^2/2\sigma^2}}{\frac{1}{2\pi}\int_0^{2\pi} e^{-(r-\theta)^2/2\sigma^2} d\theta} \\ &= \varphi(r) e^{-(\theta - r)^2/2\sigma^2}, \qquad 0 < \theta < 2\pi, \end{aligned} \tag{11-26}$$

where

$$\varphi(r) = \frac{2\pi}{\int_0^{2\pi} e^{-(r-\theta)^2/2\sigma^2} d\theta}.$$

Notice that the knowledge about the observation r is reflected in the a-posteriori p.d.f of θ in Fig. 11.2 (b). It is no longer flat as the a-priori p.d.f in Fig. 11.2 (a), and it shows higher probabilities in the neighborhood of $\theta = r$.



(a) a-priori p.d.f of θ (b) a-posteriori p.d.f of θ

Fig. 11.2

Conditional Mean:

We can use the conditional p.d.fs to define the conditional mean. More generally, applying (6-13) to conditional p.d.fs we get

$$E\left(g(X)\mid B\right)=\int_{-\infty}^{+\infty} g(x) f_X(x\mid B)dx. \tag{11-27}$$

and using a limiting argument as in (11-2) - (11-8), we get

$$\mu_{X|Y}=E\left(X\mid Y=y\right)=\int_{-\infty}^{+\infty} x\, f_{X|Y}(x\mid y)\,dx \tag{11-28}$$

to be the conditional mean of X given $Y=y$. Notice that $E(X\mid Y=y)$ will be a function of y. Also

$$\mu_{Y|X}=E\left(Y\mid X=x\right)=\int_{-\infty}^{+\infty} y\, f_{Y|X}(y\mid x)\; dy. \tag{11-29}$$

In a similar manner, the conditional variance of X given $Y=y$ is given by

$$\begin{aligned} Var(X\mid Y)=\sigma^2_{X|Y} &= E\left(X^2\mid Y=y\right)-\left(E(X\mid Y=y)\right)^2 \\ &= E\left((X-\mu_{X|Y})^2\mid Y=y\right). \end{aligned} \tag{11-30}$$

we shall illustrate these calculations through an example.

Example 11.3: Let

$$f_{XY}(x,y) = \begin{cases} 1, & 0 < |y| < x < 1, \\ 0, & \text{otherwise} . \end{cases} \qquad (11\text{-}31)$$

Determine $E(X \mid Y)$ and $E(Y \mid X)$.

Solution: As Fig. 11.3 shows, $f_{XY}(x,y) = 1$ in the shaded area, and zero elsewhere.

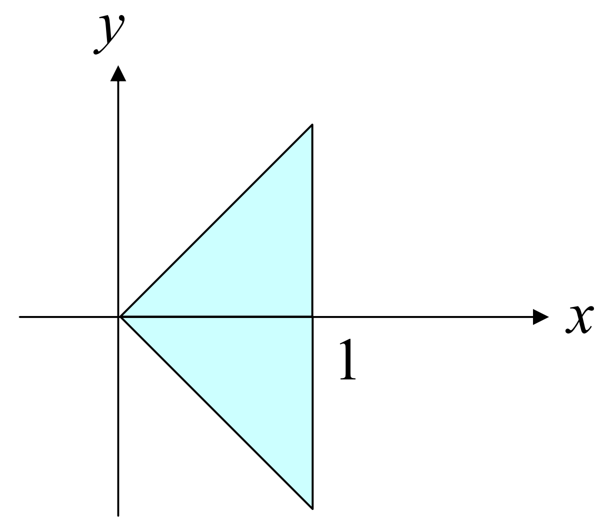


Fig. 11.3

From there

$$f_X(x) = \int_{-x}^{x} f_{XY}(x,y)dy = 2x, \quad 0 < x < 1,$$

and

$$f_Y(y) = \int_{|y|}^{1} 1 \ dx = 1 - |y|, \quad |y| < 1,$$

This gives

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{1}{1-|y|}, \quad 0 < |y| < x < 1, \qquad (11\text{-}32)$$

and

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{1}{2x}, \quad 0 < |y| < x < 1. \qquad (11\text{-}33)$$

Hence $E(X \mid Y) = \int x \, f_{X|Y}(x \mid y) dx = \int_{|y|}^{1} \frac{x}{(1-|y|)} dx$

$$= \frac{1}{(1-|y|)} \left. \frac{x^2}{2} \right|_{|y|}^{1} = \frac{1-|y|^2}{2(1-|y|)} = \frac{1+|y|}{2}, \quad |y| < 1. \qquad (11\text{-}34)$$

$$E(Y \mid X) = \int y f_{Y|X}(y \mid x) dy = \int_{-x}^{x} \frac{y}{2x} dy = \frac{1}{2x} \left. \frac{y^2}{2} \right|_{-x}^{x} = 0, \quad 0 < x < 1. \qquad (11\text{-}35)$$

It is possible to obtain an interesting generalization of the conditional mean formulas in (11-28) - (11-29). More generally, (11-28) gives

$$E\left(g(X) \mid Y = y\right) = \int_{-\infty}^{+\infty} g(x) f_{X|Y}(x \mid y) dx \ . \qquad (11\text{-}36)$$

But

$$E\left(g(X)\right) = \int_{-\infty}^{+\infty} g(x) f_X(x) dx = \int_{-\infty}^{+\infty} g(x) \int_{-\infty}^{+\infty} f_{XY}(x, y) dy dx$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x) f_{XY}(x, y) dx dy = \int_{-\infty}^{+\infty} \underbrace{\int_{-\infty}^{+\infty} g(x) f_{X|Y}(x \mid y) dx}_{E\left(g(X)|Y=y\right)} \ f_Y(y) dy$$

$$= \int_{-\infty}^{+\infty} E\left(g(X) \mid Y = y\right) f_Y(y) dy = E\left\{E\left(g(X) \mid Y = y\right)\right\} . \qquad (11\text{-}37)$$

Obviously, in the right side of (11-37), the inner expectation is with respect to X and the outer expectation is with respect to Y. Letting $g(X) = X$ in (11-37) we get the interesting identity

$$E(X) = E\{E(X \mid Y = y)\}, \qquad (11\text{-}38)$$

where the inner expectation on the right side is with respect to X and the outer one is with respect to Y. Similarly, we have

$$E(Y) = E\{E(Y \mid X = x)\}. \qquad (11\text{-}39)$$

Using (11-37) and (11-30), we also obtain

$$Var(X) = E\big(Var(X \mid Y = y)\big). \qquad (11\text{-}40)$$

Conditional mean turns out to be an important concept in estimation and prediction theory. For example given an observation about a r.v X, what can we say about a related r.v Y? In other words what is the best predicted value of Y given that $X = x$? It turns out that if "best" is meant in the sense of minimizing the mean square error between Y and its estimate $\hat{y}$, then the conditional mean of Y given $X = x$, i.e., $E(Y \mid X = x)$ is the best estimate for Y (see Lecture 16 for more on Mean Square Estimation).

We conclude this lecture with yet another application of the conditional density formulation.

Example 11.4 : **Poisson sum of Bernoulli random variables**

Let $X_i, \quad i = 1, 2, 3, \cdots$ represent independent, identically distributed Bernoulli random variables with

$$P(X_i = 1) = p, \qquad P(X_i = 0) = 1 - p = q$$

and N a Poisson random variable with parameter λ that is independent of all X_i. Consider the random variables

$$Y = \sum_{i=1}^{N} X_i, \qquad Z = N - Y. \tag{11-41}$$

Show that Y and Z are independent Poisson random variables.

Solution : To determine the joint probability mass function of Y and Z, consider

$$\begin{aligned} P(Y = m,\ Z = n) &= P(Y = m,\ N - Y = n) = P(Y = m,\ N = m + n) \\ &= P(Y = m \mid N = m + n) P(N = m + n) \\ &= P\Big(\sum_{i=1}^{N} X_i = m \,\Big|\, N = m + n\Big)\, P(N = m + n) \\ &= P\Big(\sum_{i=1}^{m+n} X_i = m\Big) P(N = m + n) \end{aligned} \tag{11-42}$$

(Note that $\sum_{i=1}^{m+n} X_i \sim B(m+n,\ p)$ and $X_i s$ are independent of N)

$$= \left(\frac{(m+n)!}{m!n!} p^m q^n \right) \left(e^{-\lambda} \frac{\lambda^{m+n}}{(m+n)!} \right)$$

$$= \left(e^{-p\lambda} \frac{(p\lambda)^m}{m!} \right) \left(e^{-q\lambda} \frac{(q\lambda)^n}{n!} \right)$$

$$= P(Y=m)P(Z=n). \tag{11-43}$$

Thus

$$Y \sim P(p\lambda) \quad \text{and} \quad Z \sim P(q\lambda) \tag{11-44}$$

and Y and Z are independent random variables.
Thus if a bird lays eggs that follow a Poisson random variable with parameter λ , and if each egg survives

with probability p, then the number of chicks that survive also forms a Poisson random variable with parameter $p\lambda$.