# Exploring Basic Probability
V. Krishnan
Presentation to CACT - September 28, 2001

**Abstract**
In this talk we will discuss some basic ideas on Probability like independence, conditional probability, Bayes' theorem, aspects of combinatorics, hypergeometric distribution, upper and lower bounds on the Gaussian tails, concluding with some continuous distributions. A few of the results are new.

## 1. Conditional Probability and Independence

The *conditional probability* of an event B conditioned on the occurrence of another event A is defined by,

$$\left.\begin{aligned} P\{B \mid A\} &= \frac{P\{A \cap B\}}{P\{A\}} \\ P\{A \cap B\} &= P\{B \mid A\}\, P\{A\} \end{aligned}\right\} \text{ if } P\{A\} \neq 0 \qquad (1.1)$$

If $P\{A\} = 0$ then the conditional probability $P\{B|A\}$ is not defined. Since this is a probability measure it also satisfies Kolmogorov axioms.

*Independence:* Two sets A and B can be called *functionally independent* if the occurrence of B does not in any way influence the occurrence of A and vice versa. Functional independence is a different concept from *statistical independence* that will be defined later. For example, the tossing of a coin is functionally independent of the tossing of a die because they do not depend on each other. However, the tossing of a coin and die are not mutually exclusive since any one can be tossed irrespective of the other. By the same token, pressure and temperature are not functionally independent because the physics of the problem, Boyle's law connects these quantities. They are certainly not mutually exclusive.

We will now define *statistical independence* of events. If A and B are two events in a field of events $\mathfrak{F}$, then A and B are statistically independent if and only if,

$$P\{A \cap B\} = P\{A\}\, P\{B\} \qquad (1.2)$$

Statistical independence neither implies nor implied by functional independence. Unless otherwise indicated we will call statistically independent events as independent without any qualifiers.

Example 1.1 (Statistically independent but functionally dependent)
Two dice, one red and the other blue are tossed. These tosses are functionally independent and we have the cartesian product of $6\times 6 = 36$ elementary events in the combined sample space with each event being equiprobable. We seek the probability of an event B defined by the sum of the numbers showing on the dice equals 9. There are four points $\{(6, 3), (5, 4), (4, 5), (3, 6)\}$ and hence $P\{B\} = {}^{4}\!/_{36} = {}^{1}\!/_{9}$. We now condition the event B with an event A defined as the red die shows odd numbers. The probability of the event A is $P\{A\} = {}^{18}\!/_{36} = {}^{1}\!/_{2}$. We want to determine whether the events A and B are statistically independent. These events are shown in Fig.(1.1)

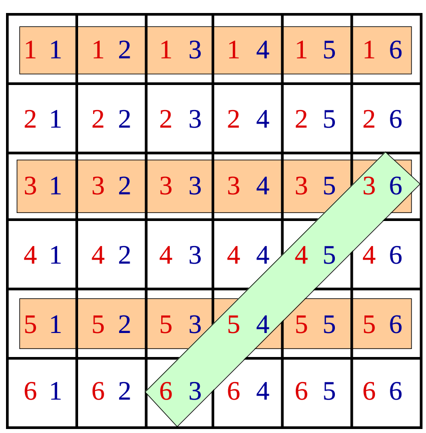

Fig. 1.1

From Fig.(1.1) $P\{A \cap B\} = P\{(3\ 6), (5\ 4)\} = 2/36 = 1/18$. and $= P\{A\} \times P\{B\} = 1/2 \times 1/9 = 1/18$ showing statistical independence. We compute $P\{B|A\}$ from the reduced sample space point of view. The conditioning event reduces the sample space from 36 points to 18 equiprobable points and the event $\{B|A\} = \{(3\ 6), (5\ 4)\}$. Hence $P\{B|A\} = 2/18 = 1/9 = P\{B\}$, or, the conditioning event has no influence on B. Here, even though the events A and B are functionally dependent, they are statistically independent.

Example 1.2 (Statistically dependent and functionally dependent)
However, if another set C is defined by the sum being equal to 8 as shown in Fig.(1.2) then $P\{C\} = 5/36$ .

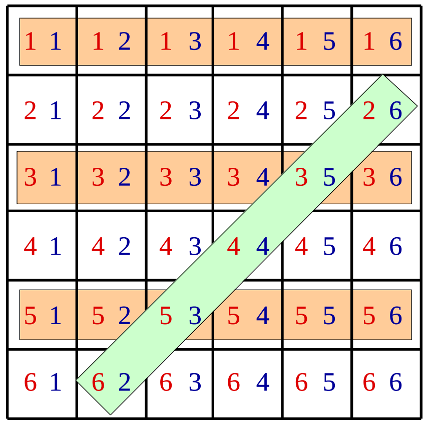

Fig. 1.2

Here the events C and A are not statistically independent because $P\{C\}.P\{A\} = 5/36 \times 1/2 = 5/72 \neq P\{C \cap A\} = 4/72$. In this example, we have the case where the events A and C are neither statistically independent nor functionally independent.

Example 1.3 (Statistically dependent but functionally independent)
We shall now take an example where the events are functionally independent but statistically dependent. An urn contains 15 balls out of which 4 are red, 5 are green and 6 are blue. Let X be the random variable representing the red balls, Y the random variable representing the green balls and Z the random variable representing the blue balls. Since X + Y + Z = 15 any two of the three random variables are functionally independent in the sense that they can be chosen independently of each other. We shall assume that X and Y are functionally independent of each other while Z = 15 – X – Y. We pick 4 balls out of this urn and calculate the joint probability $P_{XY}(x, y)$ of drawing red and green balls. This probability is given by,

$$P_{XY}(x, y) = \frac{\binom{5}{x}\binom{6}{y}\binom{4}{4-x-y}}{\binom{15}{4}} \qquad 0 \le x+y \le 4$$

We can now obtain the marginal probabilities $P_X(x)$ and $P_Y(y)$ as

$$P_X(x) = \sum_{y=0}^{4-x} P_{XY}(x, y) = \frac{\binom{5}{x}\binom{10}{4-x}}{\binom{15}{4}} \qquad 0 \le x \le 4$$

$$P_Y(y) = \sum_{y=0}^{4-y} P_{XY}(x, y) = \frac{\binom{6}{y}\binom{9}{4-y}}{\binom{15}{4}} \qquad 0 \le y \le 4$$

The probabilities $P_{Y|X}(y|X)$, Y conditioned on X, and $P_{X|Y}(x|Y)$, X conditioned on Y can be given as,

$$P_{Y|X}(y|X) = \frac{P_{XY}(x, y)}{P_X(x)} = \frac{\binom{6}{y}\binom{4}{4-x-y}}{\binom{10}{4-x}} \qquad 0 \le y \le 4-x,\ x = 0, 1, 2, 3, 4$$

$$P_{X|Y}(x|Y) = \frac{P_{XY}(x, y)}{P_Y(y)} = \frac{\binom{5}{x}\binom{4}{4-x-y}}{\binom{9}{4-y}} \quad 0 \le x \le 4-y, \; y = 0, 1, 2, 3, 4$$

Since $P_{Y|X}(y|X)$ is not equal to $P_Y(y)$ and $P_{X|Y}(x|Y)$ is not equal to $P_X(x)$ we conclude that X and Y are not statistically independent even though thet are functionally independent.

*Conditional Probability and Reduced sample space*

Conditional probability P{B|A} can also be interpreted as the probability of the event B in the reduced sample space given by the set difference (S – B). Sometimes it is simpler to use the reduced sample space to solve conditional probability problems. We shall use both these interpretations in obtaining conditional probabilities as shown in the following examples.

Example 1.4

The game of craps as played in Las Vegas has the following rules. A player rolls two dice. He wins on the first roll if he throws a 7 or a 11. He loses if the first throw is a 2, 3 or 12. If the first throw is a 4, 5, 6, 8, 9 or 10 it is called a point and the game continues. He goes on rolling until he throws the point for a win or a 7 for a loss. We have to find the probability of the player winning.

We will solve this problem both from the definition of conditional probability and the reduced sample space.

The following table shows the number of ways the sums 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 of the numbers appearing on the two dice can appear and their probabilities.

| Die A \ Die B | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 6 | 7 | 8 | 9 | 10 | 11 | 12 |

| Total | P{Total} |
|---|---|
| 2 | 1/36 |
| 3 | 2/36 |
| 4 | 3/36 |
| 5 | 4/36 |
| 6 | 5/36 |
| 7 | 6/36 |
| 8 | 5/36 |
| 9 | 4/36 |
| 10 | 3/36 |
| 11 | 2/36 |
| 12 | 1/36 |

*Solution using definition of conditional probability*

The probability of winning in the first throw is P{7} + P{11} = $\frac{6}{36} + \frac{2}{36} = \frac{8}{36}$ = 0.22222. The probability of losing in the first throw is P{2} + P{3} + P{12} = $\frac{1}{36} + \frac{2}{36} + \frac{1}{36} = \frac{4}{36}$ = 0.11111.

To calculate the probability of winning in the second throw we shall assume that i is the point with probability p. The probability of not winning in any given throw is given by the probability of not i and not 7 = r = 1 – p – $\frac{1}{6}$. We compute the conditional probability

P{win| i in first throw} = $p + rp + r^2p + \dots$

an infinite geometric series. This sum is given by,

P{win| i in first throw} = $\frac{p}{1-r}$, and

P{win in second or subsequent throws} = $\frac{p.p}{1-r}$

The probabilities of making the point 4, 5, 6, 8, 9, 10 is obtained from the table given above. Thus for i = 4,

$$P\{\text{win after first with } i = 4\} = \frac{\left(\frac{3}{36}\right)^2}{1 - \left(1 - \frac{3}{36} - \frac{1}{6}\right)} = \frac{1}{36}$$

$$P\{\text{win after first with } i = 5\} = \frac{\left(\frac{4}{36}\right)^2}{1 - \left(1 - \frac{4}{36} - \frac{1}{6}\right)} = \frac{2}{45}$$

$$P\{\text{win after first with } i = 6\} = \frac{\left(\frac{5}{36}\right)^2}{1 - \left(1 - \frac{5}{36} - \frac{1}{6}\right)} = \frac{25}{396}$$

and the same probabilities for 8, 9, 10. Thus the probability of winning in craps is given by:

$$P\{\text{win}\} = P\{\text{win in roll 1}\} + P\{\text{win in roll 2, 3, …}\}$$
$$= \left\{\frac{8}{36}\right\} + 2.\left\{\frac{1}{36} + \frac{2}{45} + \frac{25}{396}\right\} = \frac{244}{495} = 0.492929$$

*Solution using Reduced Sample Space*

We can now use the reduced sample space to arrive at the same result in a much simpler manner. After the point i has been rolled in the first throw the reduced sample space consists of (i–1+6) points for i = 4, 5, 6 and (13–i+6) points for i = 8, 9, 10 in which the game will terminate. We can see from the table that there are 6 ways of rolling 7 for a loss, (i–1, i = 4, 5, 6) ways of rolling the point for a win and (13–i, i = 8, 9, 10) ways of rolling the point for a win. Thus,

$$P\{\text{win in roll 2, 3, …} \mid i \text{ in roll 1}\} = P\{W \mid i\}$$
$$= \begin{cases} \dfrac{i-1}{i-1+6} & i = 4, 5, 6 \\[2ex] \dfrac{13-i}{13-i+6} & i = 8, 9, 10 \end{cases}$$

Thus, $P\{W \mid 4\} = \frac{3}{9}$, $P\{W \mid 5\} = \frac{4}{10}$, $P\{W \mid 6\} = \frac{5}{11}$, $P\{W \mid 8\} = \frac{5}{11}$, $P\{W \mid 9\} = \frac{4}{10}$, $P\{W \mid 10\} = \frac{3}{9}$.
Hence probability of a win after the first roll is $P\{W\ i\}.P\{i\}$. Hence,

$$P\{\text{win after first roll}\} = 2.\left[\frac{3}{9}.\frac{3}{36} + \frac{4}{10}.\frac{4}{36} + \frac{5}{11}.\frac{5}{36}\right] = \frac{134}{495}$$

$$P\{\text{win}\} = P\{\text{win in first roll}\} + P\{\text{win after first roll}\} = \frac{8}{36} + \frac{134}{495} = \frac{244}{495} = 0.492929$$

A result obtained with comparative ease.

## 2. Bayes' Theorem

The next logical question to ask is, "Given that the event B has occurred what are the probabilities that $A_1$, $A_2$, … $A_n$ are involved in it?" In other words, we have to find the conditional probabilities $P(A_1| B\}$, $P(A_2| B\}$,…, $P(A_n| B\}$. Bayes' theorem can be written as,

$$P\{A_k|B\} = \frac{P\{A_kB\}}{P\{B\}} = \frac{P\{B|A_k\}P\{A_k\}}{P\{B\}}$$
$$= \frac{P\{B|A_k\}P\{A_k\}}{\sum_{i=1}^{n} P\{B|A_i\}P\{A_i\}} \qquad (2.1)$$

This theorem connects the *a priori* probability $P\{A_i\}$ to the *a posteriori* probability $P\{A_i|B\}$. This useful theorem has wide applicability in communications and medical imaging.

Example 2.1 Monty Hall Problem

This problem is called the Monty Hall problem because of the game show host Monty Hall who designed this game. There are 3 doors A, B and C, and behind one of them is a car and behind the others are goats. The contestant is asked to select any door and the host opens one of the other two doors and shows a goat. He then offers the choice to the contestant of switching to the other unopened door or keep the original door. The question now is whether the probability of winning the car is improved if he switches or it is immaterial whether he switches or not.

The answer is counter intuitive in the sense that one maybe misled into thinking that it is immaterial whether one switches or not.

*Mathematical Analysis*

Let us analyze this problem from the point of view of Bayes' theorem. We shall first assume that the host knows the door behind which the car is located. Let us also assume that the host opens door B given that the contestant's choice is door A. The apriori probabilities of a car behind the doors A, B and C are,

$$P\{A\} = \frac{1}{3} \quad : \quad P\{B\} = \frac{1}{3} \quad : \quad P\{C\} = \frac{1}{3}$$

We can make the following observations. If the choice of the contestant is door A and the car is behind door A then the host will open the door B with probability of $1/2$ (since he has a choice between B and C). On the other hand, if the car is behind door B, then there is zero probability of his opening door B. If the car is behind door C, then he will open door B with probability 1. Thus we can write the following conditional probabilities.

$$P\{B \mid A\} = \frac{1}{2} \quad : \quad P\{B \mid B\} = 0 \quad : \quad P\{B \mid C\} = 1$$

We can now calculate the total probability P(B) of the host opening door B.

$$P\{B\} = P\{B \mid A\}P\{A\} + P\{B \mid B\}P\{B\} + P\{B \mid C\}P\{C\}$$
$$= \frac{1}{2}.\frac{1}{3} + 0.\frac{1}{3} + 1.\frac{1}{3} = \frac{1}{2}$$

Using Bayes' theorem eq. (2.1) we can now find the aposteriori probabilities of the car behind the doors A or C (B has already been opened by the host) conditioned on B.
These are:

$$P\{A \mid B\} = \frac{P\{BA\}}{P\{B\}} = \frac{P\{B \mid A\}P\{A\}}{P\{B\}} = \frac{1/2 . 1/3}{1/2} = \frac{1}{3}$$
$$P\{C \mid B\} = \frac{P\{BC\}}{P\{B\}} = \frac{P\{B \mid C\}P\{C\}}{P\{B\}} = \frac{1.1/3}{1/2} = \frac{2}{3}$$

Similar analysis holds for other cases of opening the doors A or C. We can see from the above result that the contestant must switch if he wants to double his probability of winning the car.

## 3. Combinatorics

Many problems in combinatorics can be modeled as placing r balls into n cells.

*Permutations*

The two ways of considering permutations are
(i) sampling with replacement and with ordering
(ii) sampling without replacement and with ordering

*Combinations*

Similar to the case of perturbations we have two ways of considering combinations,
(i) sampling without replacement and without ordering
(ii) sampling with replacement and without ordering.

We will only consider sampling with replacement and without ordering.

*Sampling with replacement and without ordering*

We shall now consider forming r-combinations with replacement and without ordering. In this case we pick r balls with replacement from balls in n distinct categories like, red, blue, green brown. We do not distinguish among balls of the same color and the distinction is among different colors. For example, if we are picking 6 balls from the 4 color categories then it can be arranged as follows.

| Red | Blue | Green | Brown |
|---|---|---|---|
| * * | * | | * * * |

There are now 2 red balls, 1 blue ball, no green balls and 3 brown balls. The bar symbol separates the four categories. The arrangement can not start with a bar nor end with a bar. Hence we have 4 – 1 = 3 bars. We now have to choose 6 balls from 6 balls plus 3 bars *without* replacement and without ordering. There are then $\binom{9}{6} = \binom{9}{3} = \frac{9 \times 8 \times 7}{3 \times 2 \times 1} = 84$ ways of picking 6 balls from 4 different categories.

Generalizing this result, if there are n distinct categories of balls and if we have to pick r balls *without regard to order and with replacement*, then

$$\text{Number of r-combinations} = \binom{n-1+r}{r} \qquad (3.1)$$

This is the celebrated Bose-Einstein statistics where the elementary particles in the same energy state are not distinguishable.

Example 3.1
Here is a very interesting example to illustrate the versatility of the formula in eq. (3.1). 4 dice each with 4 faces marked 1, 2, 3, 4 are tossed. We have to calculate the number of ways the tosses, will result in exactly 1 number, 2 numbers, 3 numbers and 4 numbers showing on the faces of the dice without regard to order and with replacement. The total number of ways is tabulated below in Tab.(3.1).

| | | | | |
|---|---|---|---|---|
| 1111 | 2222 | 3333 | 4444 | 4 single digits |
| 1112 | 2223 | 3334 | 4441 | |
| 1222 | 2333 | 3444 | 4111 | |
| 1113 | 2224 | 1333 | 2444 | 18 double digits |
| 1122 | 2233 | 3344 | 4411 | |
| 1133 | 2244 | - | - | |
| 1123 | 2234 | 3341 | 4412 | |
| 1134 | 2241 | 3312 | 4423 | 12 triple digits |
| 1142 | 2213 | 3324 | 4431 | |
| 1234 | - | - | - | 1 quadruple digit |

Table 3.1

We can see from the above example that there are 4 one-digits, 18 two-digits, 12 three-digits and 1 four-digits totaling 35 arrangements without regard to order and with replacement. The problem can be stated as follows. Four 4-face dice are tossed without regard to order and with replacement. Find the number of ways,

a. this can be achieved.
   With n = 4 and r = 4 in eq. (3.1), the number of ways is $\binom{4-1+4}{4} = \binom{7}{4} = 35$ that agrees with the total number in the table.

b. exactly one digit is present.
   The number of ways 1 digit can be drawn from 4 digits is $\binom{4}{1} = 4$ ways. The number of ways any particular 1 digit for example {1} will occupy a 4 digit number is a restricted case with 1 occupying one slot and the rest of the three are unrestricted. With n = 1 and r = 4 – 3 = 1 in eq. (3.1), the number of ways for this particular sequence {1} is $\binom{1-1+(4-3)}{(4-3)} = \binom{1}{1} = 1$. The total number using all the combinations are .$\binom{4}{1}\binom{1}{1} = 4$ ways. These 4 ways are (1111), (2222), (3333), (4444).

c. exactly 2 digits are present.
   The number of ways 2 digits can be drawn from 4 digits is $\binom{4}{2} = 6$ ways. The number of ways any particular 2 digits, for example, {1, 2} will occupy a 4 digit number is a restricted case with both 1 and 2 occupying any two slots and the other the two are unrestricted in the sense they can be occupied by

either 1, 2 or (1, 2). With n = 2 and r = 4 – 2 = 2 in eq. (3.1), the number of ways for this particular sequence {1, 2} is $\binom{2-1+(4-2)}{(4-2)} = \binom{3}{2} = 3$. The total number using all the $\binom{3}{2}$ combinations are $\binom{3}{2}\binom{4}{3} = 18$. ways. This agrees with the enumeration.

d. exactly 3 digits are present.
The number of ways 3 digits can be drawn from 4 digits is $\binom{4}{3} = 4$ ways. The number of ways 3 digits can be arranged in 4 slots such that all the three digits are present is again given by the restriction formula $\binom{3-1+(4-3)}{(4-3)} = \binom{3}{1} = 3$ ways. Hence the total number of ways is $\binom{4}{3}\binom{3}{1} = 12$. ways agreeing with the enumeration

e. exactly 4 digits present.
The number of ways 4 digits will be present is 1. This can be obtained from the argument in (c) above giving, $\binom{4}{4}\binom{4-1+(4-4)}{(4-4)} = \binom{4}{4}\binom{3}{0} = 1$ way.

*Generalization*
The above result can be generalized to the tossing of n n-sided dice. The total number of sample points without regard to order and with replacement is from eq. (3.1)

$$\binom{n-1+n}{n} = \binom{2n-1}{n} \tag{3.2}$$

Out of these the number of ways exactly k digits will occur can be derived as follows:
The number of ways k digits can be chosen from n is $\binom{n}{k}$ and if any particular sequence of k digits has to fill n slots then the number of ways this can be done is obtained by substituting n = k and r = (n – k) in eq. (3.1) yielding

$$\binom{k-1+(n-k)}{n-k} = \binom{n-1}{n-k}. \tag{3.3}$$

Since all the $\binom{n}{k}$ sequences are functionally independent, the number of ways N(n, k) is from eq. (3.3),

$$N(n, k) = \binom{n}{k}\binom{n-1}{n-k} \tag{3.4}$$

Using eq. (3.4) we have tabulated N(n, k) for n = 1,…, 8 and k = 1, … , 8 in Tab.(3.2)

n -sided die thrown n times

| k - Number of digits | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 2 | 0 | 1 | 6 | 18 | 40 | 75 | 126 | 196 |
| 3 | 0 | 0 | 1 | 12 | 60 | 200 | 525 | 1176 |
| 4 | 0 | 0 | 0 | 1 | 20 | 150 | 700 | 2450 |
| 5 | 0 | 0 | 0 | 0 | 1 | 30 | 315 | 1960 |
| 6 | 0 | 0 | 0 | 0 | 0 | 1 | 42 | 588 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 56 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| T | 1 | 3 | 10 | 35 | 126 | 462 | 1716 | 6435 |

Table. 3.2

In addition, we have the following interesting identity. Combining eqs.(3.2) and (3.4) we have,

$$\binom{2n-1}{n} = \sum_{k=0}^{n} \binom{n}{k}\binom{n-1}{n-k} \tag{3.5}$$

**4. Hypergeometric Distribution**

The binomial distribution is obtained while sampling with replacement. Let us select a sample of n microprocessor chips from a bin of N chips. It is known that K of these chips are defective. If we sample with replacement then the probability of any defective chip is $K/N$. The probability that k of them will be defective in a sample size of n is given by the binomial distribution since each drawing of chips constitute Bernoulli trials. On the other hand, if we sample without replacement, then the probability for the first chip is $K/N$, the probability for the second chip is $(K-1)/(N-1)$ and for each succeeding drawing the probabilities change. Hence the drawings of the chips do not constitute Bernoulli trials and binomial distribution does not hold. Hence we have to use alternate methods to arrive at the probability of finding k defective chips in n trials.

The number of sample points in this sample space is choosing n chips out of the total N chips without regard to order. This can be done in $\binom{N}{n}$ ways. Out of this sample size of n, there are k defective chips and (n–k) good chips. The number of ways of picking k defective chips out of the total of K defective chips is $\binom{K}{k}$ and the number of ways of drawing (n–k) good chips out of the remaining (N–K) good chips is $\binom{N-K}{n-k}$. We denote the total number of ways of drawing a sample of n chips with k bad chips as the event E. Since these pickings are functionally independent the number of ways is $\binom{K}{k}\binom{N-K}{n-k}$. Hence the probability of the event E is given by,

$$P\{E = k\} = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \tag{4.1}$$

This is known as the Hypergeometric Distribution.

Example 4.1

Hypergeometric distribution is used to estimate an unknown population from the data obtained. To estimate the population N of tigers in a wildlife sanctuary, K tigers are caught, tagged and released. After the lapse of a few months, a new batch of n tigers is caught and k of them are found to be tagged. It is assumed that the population of tigers does not change between the two catches. The total number of tigers N in the sanctuary is to be estimated.

In eq. (4.1), the known quantities are K, k and n. The total number N has to be estimated in equation

$$P_k(N) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

We will find the value of N that maximizes the probability $P_k\{N\}$. Such an estimate is called the *maximum likelihood estimate*. Since the numbers are very large we investigate the ratio:

$$\frac{P_k(N)}{P_k(N-1)} = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \Bigg/ \frac{\binom{K}{k}\binom{N-1-K}{n-k}}{\binom{N-1}{n}}$$

$$= \frac{\binom{N-K}{n-k}\binom{N-1}{n}}{\binom{N}{n}\binom{N-1-K}{n-k}}$$

$$= \frac{N-n}{N}\,\frac{N-K}{N-K-n+k}$$

This ratio $\frac{P_k(N)}{P_k(N-1)}$ is greater than 1 if,

$(N-n)\,(N-K) > N(N-K-n+k)$ or, $Nk < nK$, or, $N < \frac{nK}{k}$.

*Conclusion:*

If N is less than $\frac{nK}{k}$ then $P_k(N)$ is increasing and if N is greater than $\frac{nK}{k}$ then $P_k(N)$ is decreasing. Hence the maximum value of $P_k(N)$ occurs when N is equal to the nearest integer not exceeding $\frac{nK}{k}$. Thus if $K = 100$, $n = 50$ and $k = 20$ then the population of tigers is $\frac{100.50}{20} = 250$ tigers.

**5. Upper and Lower Bounds for Gaussian Tails Q(x)**

As mentioned in property (7), eq. (6.4.8), Q(x) is used to express the bit error probability in transmission of communication signals. Since Q(x) does not have a closed form solution, bounds can be established. Q(x) can expanded into an asymptotic series given by,

$$Q(x) = \frac{e^{-x^2}}{\sqrt{2\pi}}\,\frac{1}{x}\left\{1 - \frac{1}{x^2} + \frac{1.3}{x^4} + \ldots + \frac{(-1)^n 1.3.\ldots.(2n-1)}{x^{2n}} + \ldots\right\} \tag{5.1}$$

and for large x, Q(x) can be approximated by,

$$Q(x) \approx \frac{e^{-x^2}}{\sqrt{2\pi}}\,\frac{1}{x} \tag{5.2}$$

This is not a very good bound for small x. A very tight bound has been established in Ref[?] by first integrating Q(x) by parts as shown below:

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^{\infty} e^{-\xi^2/2}\,d\xi = \frac{1}{\sqrt{2\pi}}\int_x^{\infty} \frac{1}{\xi}\,\xi\, e^{-\xi^2/2}\,d\xi$$

$$= \frac{1}{\sqrt{2\pi}}\left\{\frac{1}{x}\,e^{-x^2/2} - \int_x^{\infty}\frac{1}{\xi^2}\,e^{-\xi^2/2}\,d\xi\right\}$$

and the expression for Q(x) can be manipulated to obtain,

$$Q(x) = \frac{1}{(1-\alpha)x + \alpha\sqrt{x^2+\beta}}\,\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2} \tag{5.3}$$

$\alpha$ and $\beta$ are optimized by minimizing the absolute relative error function $|\varepsilon(x)|$ given by,

$$|\varepsilon(x)| = \left|\frac{Q(x) - Q_0(x)}{Q(x)}\right|$$

The upper bound QU(x) is obtained for values of $\alpha = 0.344$ and $\beta = 5.5334$ and the lower bound QL(x) is obtained for values of $\alpha = 1/\pi$ and $\beta = 2\pi$. Thus,

$$QU(x) = \frac{1}{0.656\,x + 0.344\sqrt{x^2+5.334}}\,\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}$$

and

$$QL(x) = \frac{1}{\left(1-\frac{1}{\pi}\right)x + \frac{1}{\pi}\sqrt{x^2+2\pi}}\,\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}$$

The functions Q(x), QU(x) and QL(x) are shown in Fig. (5.1) and tabulated in Tab.(5.1). This table shows that the upper and lower bounds are for all x > 0. The percentage errors are also tabulated in Tab.(6.1) and shown in Fig.(5.2).

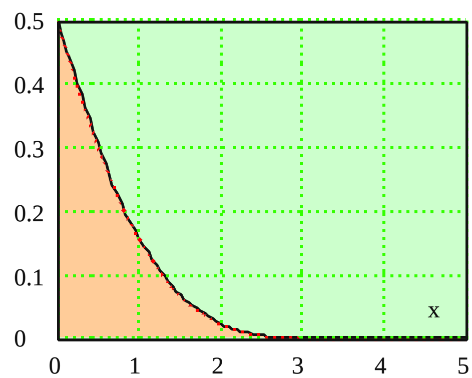


Fig. 5.1

| x | Lower Bound | True Value | Upper Bound |
|---|---|---|---|
| 0 | 0.5 | 0.5 | 0.50214035 |
| 0.5 | 0.30496416 | 0.30853754 | 0.30858864 |
| 1 | 0.15704991 | 0.15865525 | 0.15900716 |
| 1.5 | 0.06633866 | 0.0668072 | 0.06706087 |
| 2 | 0.0226461 | 0.02275013 | 0.02284873 |
| 2.5 | 0.00619131 | 0.00620967 | 0.00623602 |
| 3 | 0.00134729 | 0.0013499 | 0.00135514 |
| 3.5 | 0.00023233 | 0.00023263 | 0.00023343 |
| 4 | 0.00003164 | 0.00003167 | 0.00003177 |
| 4.5 | 0.0000034 | 0.0000034 | 0.00000341 |
| 5 | 0.00000029 | 0.00000029 | 0.00000029 |

Table 5.1

| x | %Error on Lower Bound | %Error on Upper Bound |
|---|---|---|
| 0 | 0 | - 0.42806922 |
| 0.5 | 1.15816594 | - 0.0165622 |
| 1 | 1.01184414 | - 0.22180666 |
| 1.5 | 0.70133395 | - 0.37970422 |
| 2 | 0.45727327 | - 0.43337916 |
| 2.5 | 0.29559943 | - 0.42447888 |
| 3 | 0.19344425 | - 0.38849723 |
| 3.5 | 0.12922798 | - 0.34469452 |
| 4 | 0.08836646 | - 0.30179685 |
| 4.5 | 0.06185728 | - 0.26313484 |
| 5 | 0.04427826 | - 0.22956021 |

Table 5.2

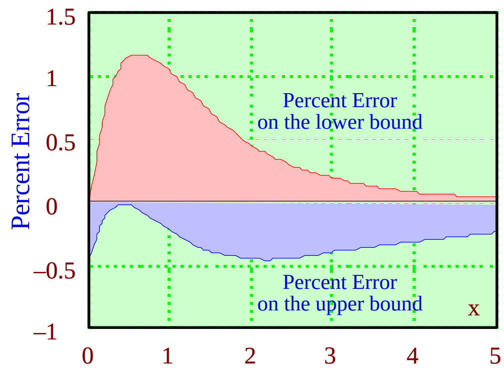

Fig. 5.2

The figures and tables illustrate clearly that the approximation for all x is excellent with the maximum error of about 1% occurring at low values of x.

## 6. Some Continuous Distributions

### Chi-Square Distribution

The Laplace, Erlang, Gamma and Weibull distributions are based on exponential distributions. The Chi-square distribution straddles both the exponential and Gaussian distributions. The general gamma distribution has parameters $(\alpha, \lambda)$. In the gamma distribution if we substitute $\alpha = n/2$ and $\lambda = 1/2$ we obtain the Chi-square distribution defined by,

$$f_X(x) = \begin{cases} \dfrac{x^{n/2-1} e^{-(x/2)}}{2^{n/2}\,\Gamma\left(\frac{n}{2}\right)} & x > 0 \\ 0 & x \le 0 \end{cases} \tag{6.1}$$

where n is called the degrees of freedom for the $\chi^2$-distribution. We can introduce another parameter $\sigma^2$ and rewrite eq. (7.7.1) as

$$f_X(x) = \begin{cases} \dfrac{x^{n/2-1} e^{-(x/2\sigma^2)}}{\left(2\sigma^2\right)^{n/2}\,\Gamma\left(\frac{n}{2}\right)} & x > 0 \\ 0 & x \le 0 \end{cases} \tag{6.2}$$

If we substitute $x = \chi^2$ in eq. (6.2) we obtain the familiar form of the Chi-square distribution given by,

$$f_X(x) = \begin{cases} \dfrac{\left(\chi^2\right)^{n/2-1} e^{-\left(\chi^2/2\sigma^2\right)}}{\left(2\sigma^2\right)^{n/2}\,\Gamma\left(\frac{n}{2}\right)} & x > 0 \\ 0 & x \le 0 \end{cases} \tag{6.3}$$

where $\Gamma(\alpha)$ is the gamma function defined by the integral,

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx\,, \quad \alpha > 0 \tag{6.4}$$

Integrating eq. (6.4 by parts with $u = x^{\alpha-1}$ and $dv = e^{-x}dx$, we have,

$$\begin{aligned} \Gamma(\alpha) &= \int_0^\infty x^{\alpha-1} e^{-x} dx\,, \quad \alpha > 0 \\ &= -x^{a-1} e^{-x}\Big|_0^\infty + (\alpha - 1)\int_0^\infty x^{\alpha-2} e^{-x} dx \\ &= (\alpha - 1)\Gamma(\alpha-1) \end{aligned} \tag{6.5}$$

If $\alpha$ is an integer n, then $\Gamma(n) = (n-1)\Gamma(n-1)$ and continued expansion yields $\Gamma(n) = n!$. Hence the gamma function can be regarded as the generalization of the factorial functionfor all positive real numbers. The most useful fractional argument for the gamma function is 1/2 with,

$$\Gamma\left(1/2\right) = \sqrt{\pi} \tag{6.6}$$

The *incomplete gamma function* $\Gamma_x(\alpha)$ is defined by,

$$\Gamma_x(\alpha) = \int_0^x \xi^{\alpha-1} e^{-\xi} d\xi \qquad (6.7)$$

has been tabulated. We will use eq. (6.7) later for the $\chi^2$-distribution function $F_X(x)$.

If n is even in eq. (6.3) then $\Gamma(n/2) = (n/2 - 1)!$ and if n is odd then $\Gamma(n/2)$ in eqs. (6.1) – (6.3) is obtained by iterating $(n/2 - 1)\,\Gamma(n/2 - 1)$ with the final iterant $\Gamma(1/2) = \sqrt{\pi}$.

Eq. (6.3) can also be obtained from the following: If $Y_1, Y_2, \ldots, Y_n$ are n i.i.d Gaussian random variables with zero mean and variance $\sigma^2$, then $X = \sum_{k=1}^{n} X_k^2$ is $\chi^2$- distributed with n-degrees of freedom.

The cumulative distribution function $F_X(x)$ is given by integrating eq. (6.3),

$$F_X(x) = \begin{cases} \displaystyle\int_0^x \frac{\xi^{n/2-1} e^{-(\xi/2\sigma^2)}}{(2\sigma^2)^{n/2}\,\Gamma(n/2)} d\xi & x > 0 \\ 0 & x \le 0 \end{cases} \qquad (6.8)$$

Eq. (6.8) can be given in terms of the incomplete gamma function $\Gamma_X(\alpha)$ given by eq. (6.7). Substituting $\eta = \xi/2\sigma^2$ and $d\xi = 2\sigma^2 d\eta$ in eq. (6.8) we obtain,

$$F_X(x) = \int_0^x \frac{\xi^{n/2-1} e^{-(\xi/2\sigma^2)}}{(2\sigma^2)^{n/2-1}\,\Gamma(n/2)} 2\sigma^2 d\xi \qquad (6.9)$$

$$= \int_0^{x/2\sigma^2} \frac{\eta^{n/2-1} e^{-(\eta)}}{\Gamma(n/2)} d\eta = \frac{\Gamma_{x/2\sigma^2}(n/2)}{\Gamma(n/2)}$$

When $x \to \infty$ then $\Gamma_{x/2\sigma^2}(n/2) \to \Gamma(n/2)$ and hence $F_X(x) = 1$. The cdf $F_X(x)$ in eq. (6.9) has been extensively tabulated.

The Chi-square distribution is shown in Fig. (6.1) for the number of degrees of freedom n = 2, 3, …, 10.

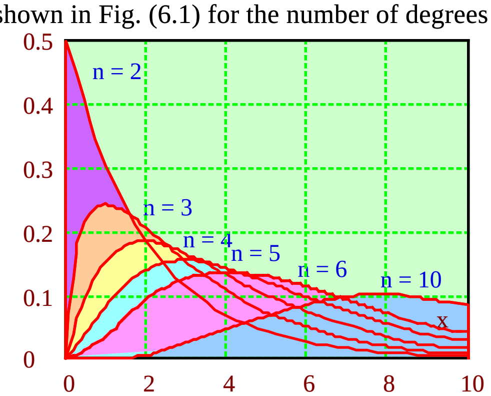


Fig. 6.1

The Chi-square distribution is used directly or indirectly in many tests of hypotheses. The most common use of the chi-square distribution is to test the goodness of a hypothesized fit. The goodness of fit test is performed to determine if an observed value of a statistic differs enough from a hypothesized value of a distribution to draw the inference whether the hypothesized quantity is not the true distribution.

Example 7.1
The density function of $f_Z(z)$ of $Z = X^2 + Y^2$ where X and Y are independent random variables distributed as $N(0, \sigma^2)$ can be found by substituting n = 2 in eq. (6.2) yielding,

$$f_Z(z) = \begin{cases} \dfrac{e^{-\left(z/2\sigma^2\right)}}{2\sigma^2} & z > 0 \\ 0 & z \le 0 \end{cases}$$

which is a $\chi^2$-density with 2 degrees of freedom. The corresponding cdf, $F_Z(z)$ is given by,

$$F_Z(z) = \begin{cases} 1 - e^{-\left(z/2\sigma^2\right)} & z > 0 \\ 0 & z \le 0 \end{cases}$$

**Chi-Distribution**

The Rayleigh and Maxwell distributions are special cases of the Chi distribution. Just as the sum of the squares of n zero mean Gaussian random variables with variance $\sigma^2$ is Chi-square distributed with n degrees of freedom we can consider that the sum of the square root of the sum of the squares of n zero mean Gaussian random variables with variance $\sigma^2$ is Chi distributed with n degrees of freedom. Accordingly, let $Z_n$ be the random variable given by,

$$Z = \sqrt{X_1^2 + X_2^2 + \dots + X_n^2} \tag{6.10}$$

where $\{X_i,\ i = 1, \dots, n\}$ are zero mean Gaussian random variables with variance $\sigma^2$. Then the chi-density $f_Z(z)$ can be obtained by substituting $x = z^2$ and $dx = 2zdz$ in the $\chi^2$-density given by,

$$f_X(x)\,dx = \frac{x^{n/2-1} e^{-\left(x/2\sigma^2\right)}}{\left(2\sigma^2\right)^{n/2}\Gamma\left(\frac{n}{2}\right)}\,dx = f_Z(z)\,dz$$

and $f_Z(z)$ is evaluated to yield,

$$f_Z(z) = \begin{cases} \dfrac{1}{2^{n/2-1}\,\Gamma\left(\frac{n}{2}\right)} \dfrac{z^{n-1}}{\sigma^n} e^{-1/2\,(z/\sigma)^2} & z > 0 \\ 0 & z \le 0 \end{cases} \tag{6.11}$$

where $\Gamma(n/2)$ is the gamma function defined in eq. (6.4).

A family of Chi-distributions is shown in Fig. (6.2) for $\sigma = 1.5$ and n = 1, 2, 3, 5, 7, 9, 11, 13, 15.

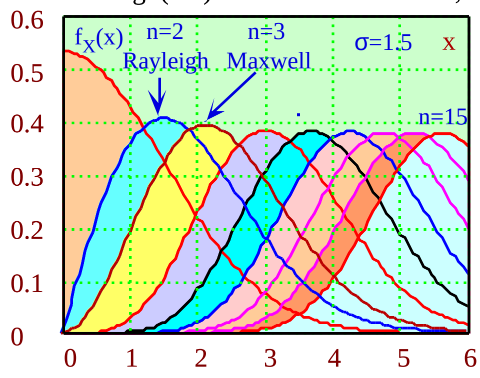


Fig. 6.2

The distribution function $F_Z(z)$ is given by integrating eq. (6.11),

$$F_Z(z) = \begin{cases} \dfrac{1}{2^{n/2-1}\,\Gamma\left(\frac{n}{2}\right)} \displaystyle\int_0^z \frac{\zeta^{n-1}}{\sigma^n} e^{-1/2\,(\zeta/\sigma)^2}\,d\zeta & z > 0 \\ 0 & z \le 0 \end{cases} \tag{6.12}$$

There is no closed form solution for eq. (7.8.3) but it can be expressed in terms of the incomplete gamma function $\Gamma_z(z)$ defined in eq. (6.7).

By substituting $y = {\zeta^2}/{2\sigma^2}$ and $d\zeta = (\sigma^2 dy)/\zeta$ in the in eq. (6.12) we can write,

$$
\begin{aligned}
F_Z(z) &= \frac{2^{n/2-1}}{2^{n/2-1}\Gamma\left(\frac{n}{2}\right)} \int_0^{z^2/2\sigma^2} y^{n/2-1} e^{-y}\, dy \\
&= \frac{2^{n/2-1}}{2^{n/2-1}\Gamma\left(\frac{n}{2}\right)} \Gamma_{(z^2/2\sigma^2)}\left(\frac{n}{2}\right) = \frac{\Gamma_{(z^2/2\sigma^2)}\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}
\end{aligned}
\tag{6.13}
$$

where $\Gamma_{(z^2/2\sigma^2)}\left(\frac{n}{2}\right)$ is the incomplete gamma function defined in eq. (6.7). Clearly as $z \to \infty$, $\Gamma_{(z^2/2\sigma^2)}\left(\frac{n}{2}\right) \to \Gamma(n/2)$ thus showing that $F_Z(\infty) = 1$.

There are two very important special cases of the Chi-distribution when n = 2 and n = 3.

*Rayleigh Distribution*
When n = 2 in the Chi distribution of eq. (7.8.2) we have,

$$
f_Z(z) = \begin{cases} \dfrac{z}{\sigma^2} e^{-\frac{1}{2}(z/\sigma)^2} & z > 0 \\ 0 & z \le 0 \end{cases}
\tag{6.14}
$$

This is called the Rayleigh density and finds wide application in communications fading channels. If we have two independent zero mean Gaussian random variables X and Y with variance $\sigma^2$, then if we make the transformation $X = Z\cos\theta$ and $Y = Z\sin\theta$, then $Z = Z = \sqrt{X^2 + Y^2}$ is Rayleigh distributed which is a chi-distribution with 2 degrees of freedom. $\theta$ is uniformly distributed in the interval $(0, 2\pi]$. The Rayleigh density is shown in Fig. (6.3) for $\sigma = \sqrt{10}$, 4, 5.

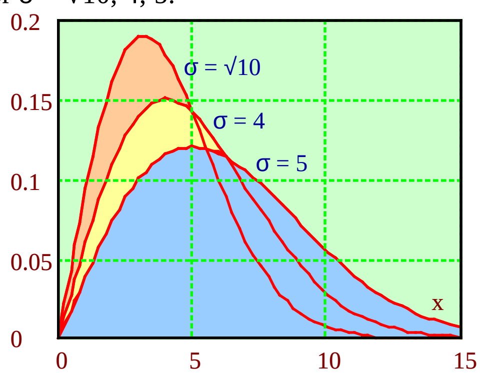


Fig. 6.3

The cdf corresponding to the Rayleigh density can be expressed in a closed form as follows:

$$
F_Z(z) = \int_0^z \frac{\zeta}{\sigma^2} e^{-\zeta^2/2\sigma^2}\, d\zeta \quad \text{Or,}
$$
$$
F_Z(z) = \begin{cases} 1 - e^{-z^2/2\sigma^2} & z > 0 \\ 0 & z \le 0 \end{cases}
\tag{6.15}
$$

*Maxwell Density*
This is one of the earliest distributions derived by Maxwell and Boltzmann. They derived that the speed distribution of gas molecules is given by,

$$
f(v) = 4\pi\left[\frac{m}{2\pi kT}\right]^{3/2} v^2 e^{-mv^2/2kT}
\tag{6.16}
$$

where k is the Boltzmann constant, m is the mass, T is the absolute temperature and v is the speed. This is based on the assumption that the molecules are distinguishable and they can occupy any energy state. This is a special case of the Chi-density of eq. (7.8.2) with n = 3 degrees of freedom and is given by,

$$f_X(x) = \sqrt{\frac{2}{\pi}}\,\frac{x^2}{\sigma^3}\,e^{-x^2/2\sigma^2} \qquad (6.17)$$

If we substitute $kT/m = \sigma^2$ in the Maxwell-Boltzmann distribution of eq. (6.16), we obtain eq. (6.17). The pdf $f_X(x)$ for the Maxwell and Rayleigh densities are shown in Fig. (6.4) for values of $\sigma = \sqrt{10}$, 4 and 5.

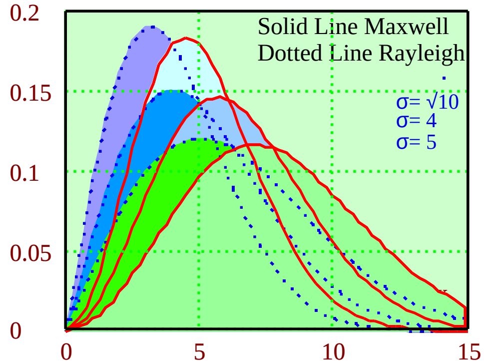

Fig. 6.4

The cdf $F_X(x)$ corresponding to the Maxwell density $f_X(x)$ is obtained by integrating eq. (6.17) and is given by,

$$F_X(x) = \sqrt{\frac{2}{\pi}}\int_0^x \frac{\xi^2}{\sigma^3}\,e^{-\xi^2/2\sigma^2}\,d\xi \qquad (6.18)$$

There is no closed form solution for eq. (7.8.9). However, integrating by parts, it cam be expressed as,

$$F_X(x) = \frac{2}{\sqrt{\pi}}\int_0^{\frac{x}{\sqrt{2}\,\sigma}} e^{-\xi^2/2\sigma^2}\,d\xi - \sqrt{\frac{2}{\pi}}\,\frac{x}{\sigma}\,e^{-x^2/2\sigma^2} \qquad (6.19)$$

$$= \operatorname{erf}\left(\frac{x}{\sqrt{2}\,\sigma}\right) - \sqrt{\frac{2}{\pi}}\,\frac{x}{\sigma}\,e^{-x^2/2\sigma^2} \qquad x > 0$$

where the error function erf(x) is defined as,

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-\xi^2}\,d\xi \qquad (6.20)$$

**Rice's Density**

In the Rayleigh distribution discussed in the last section, the random variables X and Y must be zero mean having the same variance $\sigma^2$. In many communication problems involving fading channels, X and Y will not be zero mean and will have mean values equal to $\mu_X$ and $\mu_Y$. If the random variable $Z = \sqrt{X^2 + Y^2}$ then $f_Z(z)$ has a Rice density with a noncentrality parameter $m = \sqrt{\mu_X^2 + \mu_Y^2}$ . It is given by,

$$f_Z(z) = \frac{z}{\sigma^2}\exp\left[\frac{-(z^2 + m^2)}{2\,\sigma^2}\right] I_0\left(\frac{mz}{\sigma^2}\right) \qquad (6.21)$$

where $I_0(z)$ is the zeroth order modified Bessel function of the first kind. The general $\nu$th order modified Bessel function is given by the series,

$$I_\nu(z) = \left(\frac{z}{\nu}\right)^\nu \sum_{k=0}^{\infty} \frac{\left(z^2/4\right)^k}{k!\,\Gamma(\nu + k + 1)} \qquad (6.22)$$

and the zeroth order is given by,

$$I_0(z) = \sum_{k=0}^{\infty} \frac{\left(z^2/4\right)^k}{(k!)^2} \qquad (6.23)$$

and $I_0(x)$, $I_1(x)$, $I_2(x)$, $I_3(x)$ are shown in Fig. (6.5).

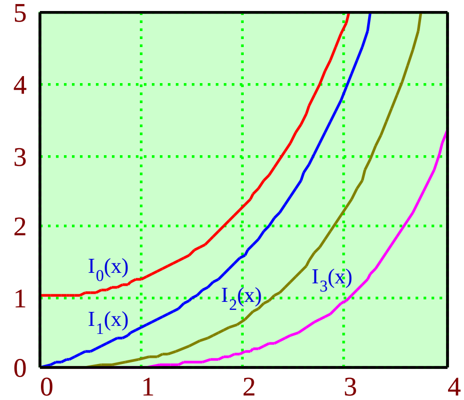

Fig. 6.5

From Fig. (6.5) we see that $I_0(0) = 1$ and hence when m = 0 (zero mean) then $f_X(x)$ is the Rayleigh distributed. Rice's Density is shown in Fig. (6.6) for m = 0, 2, 4, 6, 8, 10, 12, 14. m=0 is the Rayleigh density.

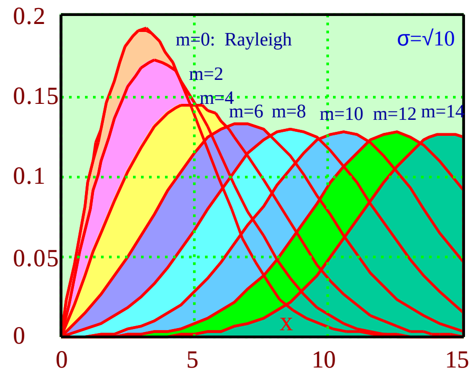

Fig. 6.6

The distribution function $F_X(x)$ corresponding to eq. (6.21) is given by integrating $f_Z(x)$.

$$F_Z(z) = \int_0^z \frac{\zeta}{\sigma^2} \exp\left[\frac{-(\zeta^2 + m^2)}{2\,\sigma^2}\right] I_0\left(\frac{m\zeta}{\sigma^2}\right) d\zeta \qquad (6.24)$$

There is no closed from solution for this integral.

**Nakagami Density**

The Nakagami distribution, like the Rayleigh density also belongs to the class of central Chi-square distributions. It is used for modeling data from multipath fading channels and it has been shown to fit empirical results more generally than the Rayleigh distributions. Sometimes, the pdf of the amplitude of a mobile signal can also be described by the Nakagami m-distribution.

The Nakagami distribution may be a reasonable means to characterize the backscattered echo from breast tissues to automate a scheme for separating benign and malignant breast masses. The shape parameter m has been shown to be useful in tissue characterization. Chi-square tests showed that the this distribution is a better fit to the envelope than the Rayleigh distribution. Two parameters, m (effective number) and Ω (effective cross section), associated with the Nakagami distribution are used for the effective classification of breast masses.

The Nakagami m-density is defined by,

$$f_X(x) = \frac{2}{\Gamma(m)}\left[\frac{m}{\Omega}\right]^m x^{2m-1}\, e^{-m x^2/\Omega} \tag{6.25}$$

where m is the shape parameter and $\Omega$ controls the spread of the distribution.
The interesting feature of this density is that unlike the Rice distribution, it is not dependent on the modified Bessel function. The Nakagami family is shown in Fig. (6.7) for $\Omega = 1$ and for m = 3, 2, 1.5, 1, 0.75, 0.5.

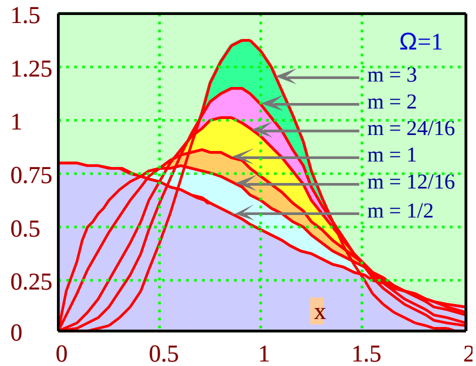

Fig. 6.7

The cdf $F_X(x)$ is given by,

$$F_X(x) = \int_0^x \frac{2}{\Gamma(m)}\left[\frac{m}{\Omega}\right]^m \xi^{2m-1}\, e^{-m\xi^2/\Omega}\, d\xi \tag{6.26}$$

This has no closed form solution. However, it cam be given in terms of the incomplete gamma function by substituting $y = m\,\xi^2/\Omega$ and $d\xi = \left(\Omega/2m\xi\right) dy$ in eq. (6.26). This results in,

$$\begin{aligned} F_X(x) &= \int_0^x \frac{2}{\Gamma(m)}\left[\frac{m\xi^2}{\Omega}\right]^m \frac{1}{\xi}\, e^{-m\xi^2/\Omega}\, d\xi \\ &= \int_0^{mx^2/\Omega} \frac{2}{\Gamma(m)}\left[\frac{m\xi^2}{\Omega}\right]^m e^{-m\xi^2/\Omega}\, \frac{1}{2\left(m\xi^2/\Omega\right)}\, dy \\ &= \int_0^{mx^2/\Omega} \frac{2}{\Gamma(m)}\, y^m\, e^{-y}\, \frac{1}{2y}\, dy \\ &= \int_0^{mx^2/\Omega} \frac{1}{\Gamma(m)}\, y^{m-1}\, e^{-y}\, dy = \frac{\Gamma_{mx^2/\Omega}(m)}{\Gamma(m)} \end{aligned} \tag{6.27}$$

Since $\Gamma_{mx^2/\Omega}(m) \rightarrow \Gamma(m)$ as $x \rightarrow \infty$ we have $F_X(\infty) = 1$.