1 Combinatorial Analysis

1.1 Introduction

This chapter deals with finding effective methods for counting the number of ways that things can occur. In fact, many problems in probability theory can be solved simply by counting the number of different ways that a certain event can occur. The mathematical theory of counting is formally known as combinatorial analysis.

1.2 The Basic Principle of Counting

Number of ways to perform two procedures in succession: If we are going to perform two procedures in succession and if the first procedure can be performed in n_1 ways, and if, for each of these ways, the second procedure can be performed in n_2 ways, then there are $n_1 n_2$ ways in which the two procedures can be performed successively.

Number of ways to perform several procedures in succession: Similarly, if we are performing r procedures successively, the ith procedure capable of being performed in n_i ways regardless of the ways in which the first $(i-1)$ procedures were performed, then the r procedures can be performed in $n_1 n_2 \cdots n_r$ different ways.

1.3 Permutations

Number of ways to order n distinct elements: Based on the reasoning mentioned earlier, if we have n distinct elements, there would be $n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$ ways of ordering them. We denote this quantity by $n!$ and in words by n factorial.

Number of ways to order n elements (some of which are not distinct): The number of distinct orderings of n objects, n_1 of which are type 1, n_2 of which are of type 2, ..., and n_r of which are of type r, is equal to:

$$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!} \ , \quad n_1 + n_2 + \ldots n_r = n$$

Number of ways to select r elements from n elements (order is important): The total number of ways of ordering r elements, chosen from n distinct elements, is equal to:

$$n(n-1)(n-2)\cdots(n-r+1)$$

This quantity can be also expressed as $n!/(n-r)!$. This is denoted by P_r^n, the number of permutations of n things taken r at a time.

1.4 Combinations

Number of ways to select r elements from n distinct elements (order is not important): It is possible to choose, without regard of order, r elements from n distinct elements in

$$\frac{n!}{r!(n-r)!}$$

different ways. This is an important quantity in combinatorics and will be denoted by $\binom{n}{r}$. It is also called a *Binomial Coefficient* and is sometimes written as C_n^r.

Note that,

$$\binom{n}{0} = \binom{n}{n} = 1$$

and

$$\binom{n}{i} = 0, \quad \text{for} \quad i < 0, \quad \text{and for} \quad i > n$$

A useful combinatorial identity is,

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}, 1 \leq r \leq n$$

This identity may be proved by the following combinatorial argument. Consider a group of n objects and fix attention on some particular one of these objects (call it object 1). Now, there are $\binom{n-1}{r-1}$ groups of size r that contain object 1 (since each such group is formed by selecting $r-1$ objects from the $n-1$ objects which are those remaining after excluding object 1). Also, there are $\binom{n-1}{r}$ groups of size r that do not contain object 1 (since each such group is formed by selecting r objects from the $n-1$ objects which are those remaining after excluding object 1). Note that there is a total of $\binom{n}{r}$ groups of size r, and a given group among these either contains or does not contain object 1. This means that the total number of possible ways to select r objects out of n is equal to the number of ways to select r objects from n when object 1 is included (total of $\binom{n-1}{r-1}$ groups) plus the number of ways to select r object from n when object 1 is not included (total of $\binom{n-1}{r}$ groups). This results in the above identity.

We also have,

$$\binom{n}{r} = \binom{n}{n-r}$$

<u>**Binomial Theorem**</u> is as follows:

$$(X+Y)^n = \sum_{k=0}^{n} \binom{n}{k} X^k Y^{n-k}$$

Combinatorial Proof of the Binomial Theorem: Consider the product

$$(X_1 + Y_1)(X_2 + Y_2) \dots (X_n + Y_n)$$

Its expansion consists of the sum of 2^n terms, each term being the product of n factors. Furthermore, each of the 2^n terms in the sum will contain as a factor either X_i or Y_i for each $i = 1, 2, ..., n$. For example,

$$(X_1 + Y_1)(X_2 + Y_2) = X_1X_2 + X_1Y_2 + Y_1X_2 + Y_1Y_2$$

Now, how many of the 2^n terms in the sum will have as factors k of the X_i's and $(n-k)$ of the Y_i's? As each term consisting of k of the X_i's and $(n-k)$ of the Y_i's corresponds to a choice of a group of size k from the n values $X_1, X_2, ..., X_n$, there are $\binom{n}{k}$ such terms. Thus, letting $X_i = X$, $Y_i = Y$, $i = 1, ..., n$, we get the desired result.

1.5 Multinomial Coefficients

Multinomial Theorem is as follows:

$$(x_1 + x_2 + \cdots + x_r)^n = \sum \binom{n}{n_1, n_2 \ldots, n_r} x_1^{n_1} x_2^{n_2} \cdots x_r^{n_r}$$

where $n_1 + n_2 + \cdots + n_r = n$, and

$$\binom{n}{n_1, n_2 \ldots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!}$$

and the summation is over all possible sets of integers $(n_1, n_2, \ldots, n_r)$ such that
(i) $0 \leq n_i \leq n, i = 1, 2, \ldots, r$ and
(ii) $n_1 + n_2 + \cdots + n_r = n$.

1.6 The Number of Integer Solutions of Equations

The Number of Positive Integer Solutions of Equations: There are $\binom{n-1}{r-1}$ distinct positive integer-valued vectors $(x_1, x_2, \ldots, x_r)$ satisfying:

$$x_1 + x_2 + \cdots + x_r = n \qquad x_i > 0, i = 1, \ldots, r$$

Proof: Refer to problem 7.

The Number of Nonnegative Integer Solutions of Equations:

There are $\binom{n+r-1}{r-1}$ distinct nonnegative integer-valued vectors $(x_1, x_2, \ldots, x_r)$ satisfying:

$$x_1 + x_2 + \cdots + x_r = n \qquad\qquad x_i \geq 0, i = 1, \ldots, r$$

Proof: Refer to problem 8.

1.7 Some Solved Problems

1. Twenty books are to be arranged on a shelf; eleven on travel, five on cooking, and four on gardening. The books in each category are to be grouped together. How many arrangements are possible?

   Solution: We have 11! arrangements for the travel, 5! arrangements for the cooking, and 4! arrangements for the gardening books. We can also permute the three different classes of books in 3! ways. Thus

   $$\text{total} = (11!)(5!)(4!)(3!)$$

2. A sandwich in a vending machine costs \$0.85. In how many ways can a customer put in two quarters, three dimes, and a nickel?

   Solution: We know that the number of ways to arrange n objects, n_1 of one kind, n_2 of a second kind, ..., and n_r of an rth kind, $\sum_{i=1}^{r} n_i = n$, is

   $$\frac{n!}{n_1!n_2!\cdots n_r!}$$

   Viewed in this context, an admissible coin sequence is a permutation of $n = n_1 + n_2 + n_3 = 6$ objects, where

   $$\begin{aligned} n_1 &= \text{number of quarters} = 2 \\ n_2 &= \text{number of dimes} = 3 \\ n_3 &= \text{number of nickels} = 1 \end{aligned}$$

   It follows that the number of different ways the coins can be put into the machine is 60:

   $$\frac{n!}{n_1!n_2!n_3!} = \frac{6!}{2!3!1!} = 60$$

3. Twelve people belong to a club. How many ways can they pick a president, vice-president, secretary, and treasurer?

   Solution: We think of filling the offices one at a time. There are 12 people we can pick for president. Having made the first choice, there are always 11 possibilities for vice-president, 10 for secretary, and 9 for treasurer. The *Multiplication rule* states that, if m experiments are performed in order and that, no matter what the outcomes of experiments $1, ..., k-1$ are, experiment k has n_k possible outcomes, then the total number of outcomes is $n_1 \cdot n_2 \cdots n_m$. So by the multiplication rule, the answer is

   $$\frac{12}{P} \cdot \frac{11}{V} \cdot \frac{10}{S} \cdot \frac{9}{T}$$

4. A committee of three consisting of two men and one woman is to be chosen from six men and three women. How many different committees can be chosen?

   Solution: The two men can be chosen in C_2^6 ways. The one woman can be chosen in C_1^3 ways. Thus the total number of different possible committees is

   $$(C_2^6)(C_1^3) = \binom{6}{2}\binom{3}{1} = \frac{6!}{4!2!} \cdot \frac{3!}{2!} = 45$$

5. In a class of 70 students, how many ways can the students be grouped so that there are 12 students in each of the first five groups and 10 students in the last one?

   Solution: There are C_{12}^{70} choices for the first group. Having chosen 12 for the first group, there remain C_{12}^{58} choices for the second group and so on. By the multiplication principle, the total is

$$\begin{aligned} \text{total} &= \binom{70}{12}\cdot\binom{58}{12}\cdot\binom{46}{12}\cdot\binom{34}{12}\cdot\binom{22}{12}\cdot\binom{10}{10} \\ &= \frac{70!}{58!12!}\cdot\frac{58!}{46!12!}\cdot\frac{46!}{34!12!}\cdot\frac{34!}{22!12!}\cdot\frac{22!}{10!12!}\cdot\frac{10!}{0!10!} \\ &= \frac{70!}{(12!)^5 10!} \end{aligned}$$

6. A house has 12 rooms. We want to paint 4 yellow, 3 purple, and 5 red. In how many ways can this be done?

   Solution: To generate all the possibilities, we can first decide the order in which the rooms will be painted, which can be done in 12! ways, then paint the first 4 on the list yellow, the next 3 purple, and the last 5 red. One example is

$$\frac{9}{Y}\,\frac{6}{Y}\,\frac{11}{Y}\,\frac{1}{Y}\,\frac{8}{P}\,\frac{2}{P}\,\frac{10}{P}\,\frac{5}{R}\,\frac{3}{R}\,\frac{7}{R}\,\frac{12}{R}\,\frac{4}{R}$$

   Now, the first four choices can be rearranged in 4! ways without affecting the outcome, the middle three in 3! ways, and the last five in 5! ways. Invoking the multiplication rule, we see that in a list of the 12! possible permutations each possible painting thus appears 4!3!5! times. Hence, the number of possible paintings is

$$\frac{12!}{4!3!5!} = \frac{12\cdot 11\cdot 10\cdot 9\cdot 8\cdot 7\cdot 6}{1\cdot 2\cdot 3\cdot 1\cdot 2\cdot 3\cdot 4} = 27720$$

   Another way of getting the last answer is to first pick 4 of 12 rooms to be painted yellow, which can be done in C_4^{12} ways, and then pick 3 of the remaining 8 rooms to be painted purple, which can be done in C_3^8 ways. (The 5 unchosen rooms will be painted red.) This is the same answer since

$$C_4^{12}C_3^8 = \frac{12!}{4!8!}\cdot\frac{8!}{3!5!} = \frac{12!}{4!3!5!}$$

7. How many ways can we divide n indistinguishable balls into r distinguishable groups in such a way that there is at least one ball per group?

   Solution: To solve this problem we imagine a line of n white balls and $r-1$ pieces of cardboard to place between them to indicate the boundaries of the groups. An example with $n = 14$ and $r = 6$ would be

$$ooo|o|oo|oooo|ooo|o$$

   Since we must pick $r-1$ of the $n-1$ spaces to put our cardboard dividers into, there are $\binom{n-1}{r-1}$ possibilities.
   When the groups can have size 0, the last scheme breaks down and we need to look at things in a

different way. We imagine having n white balls (w's) and $r-1$ black balls (b's) that will indicate the boundaries between groups. An example with $n = 9$, $r = 6$ would be

$$w\,w\,b\,b\,w\,w\,w\,w\,b\,w\,b\,w\,w\,b$$

which corresponds to groups of size $n_1 = 2$, $n_2 = 0$, $n_3 = 4$, $n_4 = 1$, $n_5 = 2$, and $n_6 = 0$. Each possible division of n indistinguishable balls into r boxes corresponds to one arrangement of the n white balls and $r-1$ black balls, which is in turn the same as the number of ways of getting n heads and $r-1$ tails in $n+r-1$ tosses, so there are $\binom{n+r-1}{r-1}$ outcomes.

8. Show that there are $\binom{n+r-1}{n}$ distinct nonnegative integer-valued vectors with components $(x_1, x_2, \ldots, x_r)$ satisfying

$$x_1 + x_2 + \ldots + x_r = n$$

Solution: Consider a vector consisting of n ones and $(r-1)$ zeros. To each permutation of this vector we correspond a solution of the above equation, namely, the solution where x_1 equals the number of ones to the left of the first zero, x_2 equals the number of ones between the first and second zeros, x_3 equals the number of ones between the second and third zeros, and so on until x_r, which equals the number of ones to the right of the last zero. For instance, if n =6, r =4, then the vector (1,1,0,0,1,1,1,0,1) corresponds to the solution x_1 =2, x_2 =0, x_3 =3, x_4 =1. It is easy to see that this correspondence between permutations of a vector of n ones and $(r-1)$ zeros and solutions of the above equation is a one-to-one correspondence. The result now follows because there are $\frac{(n+r-1)!}{n!(r-1)!}$ permutations of a vector of n ones and $(r-1)$ zeros. *Try to relate this problem to the previous problem.*

9. Show that

$$\binom{n}{1} + \binom{n}{3} + \cdots = \binom{n}{0} + \binom{n}{2} + \cdots$$

for any n.

Solution: Consider the expansion of $(x-y)^n$:

$$(x-y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k (-y)^{n-k}$$

Let $x = y$ =1. Then $x - y$ =0, and the previous equation reduces to

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

which can be written

$$\binom{n}{0} + \binom{n}{2} + \cdots = \binom{n}{1} + \binom{n}{3} + \cdots$$

10. Prove that

$$\binom{n}{1} + 2\binom{n}{2} + \cdots + n\binom{n}{n} = n2^{n-1}$$

Solution: This time we begin with the expansion of $(1+x)^n$:

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

Differentiating both sides of the previous equation with respect to x gives

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

Now, let x =1. This simplifies the left-hand side of the previous equation to $n2^{n-1}$, while the right-hand side reduces to

$$\sum_{k=0}^{n} k \binom{n}{k} = \binom{n}{1} + 2\binom{n}{2} + \cdots + n\binom{n}{n}$$

2 Axioms of Probability

2.1 Introduction

In this chapter we introduce the concept of the probability of an event and then show how these probabilities can be computed in certain situations. This is based on defining the concept of the sample space. Consider an experiment whose outcome is not predictable with certainty. However, although the outcome of the experiment will not be known in advance, let us suppose that the set of all possible outcomes is known. This set of all possible outcomes of an experiment is known as the sample space of the experiment.

2.2 Sample Space and Events

Experiment: An experiment is a procedure which results in an outcome. In general, the outcome of an experiment depends on the conditions under which the experiment is performed.

Deterministic Experiment: In a *deterministic experiment*, the observed result (outcome) is not subject to chance. This means that if we repeat a deterministic experiment under exactly the same conditions, we will get the same result.

Random Experiment: In a *random experiment*, the outcome is subject to chance. This means that if the experiment is repeated, the outcome may be different from time to time.

Statistical Regularity: Suppose that a random experiment is repeated for a large number of times. We say that the experiment has *statistical regularity* if the fraction of times that a particular event E is observed tends to some limit.

Note: In this course, we consider only experiments which are repeatable and the outcomes of the experiment exhibit statistical regularity.

Sample Space: For any random experiment, we define the *sample space* S to be the set of all possible outcomes of the experiment.

Elementary event: An elementary event E is a single element of S corresponding to a particular outcome of the experiment.

Event: An event is any subset of S (a collection of elementary events or outcomes).
Union of Some Events: The union of events $E_1, E_2, \ldots E_n$ is defined the event for which at least one of E_is occurs and denoted by $\bigcup_{i=1}^{n} E_i$.

Intersection of Some Events: The intersection of events $E_1, E_2, \ldots E_n$ is defined the event for which all of E_is occur and denoted by $\bigcap_{i=1}^{n} E_i$.

Complement of An Event: Consider an even E. The complement of E (denoted as E^c) is the event that E does not occur. We have, $EE^c = \emptyset, \;\; E \cup E^c = S$.

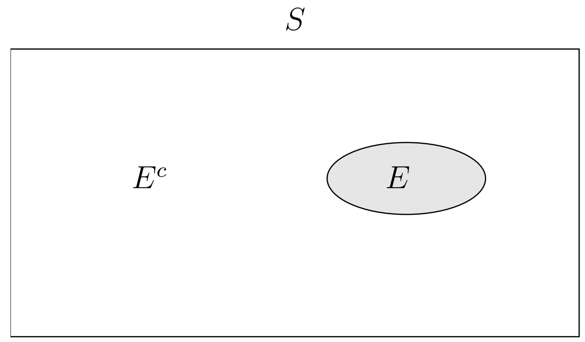


Figure 1: Complement of an event E.

Partitioning of Sample Space: The events $F_1, F_2, \ldots, F_n$ form a partition of S if they are disjoint and their union is S. This means that,

(i) $\bigcup_{i=1}^{n} F_i = S$

(ii) $F_i F_j = \emptyset, \;\; i \neq j$

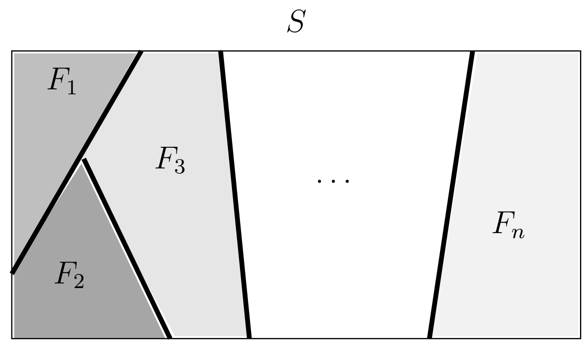


Figure 2: Partition of the sample space.

If $F_1, F_2, \ldots, F_n$ form a partition of the sample space, then, for any event $E \subset S$, we have,

$$E = \bigcup_{i=1}^{n} (EF_i), \;\; \text{where} \;\; (EF_i)(EF_j) = \emptyset, \; \text{for } i \neq j \qquad (1)$$

Some Properties of Union and Intersection

Commutative laws:	$E \cup F = F \cup E$	$EF = FE$
Associative laws:	$(E \cup F) \cup G = E \cup (F \cup G)$	$(EF)G = E(FG)$
Distributive laws:	$(E \cup F)G = EG \cup FG$	$(EF) \cup G = (E \cup G)(F \cup G)$

Note that, $E\emptyset = \emptyset, \quad ES = E, \quad E \cup \emptyset = E, \quad E \cup S = S$

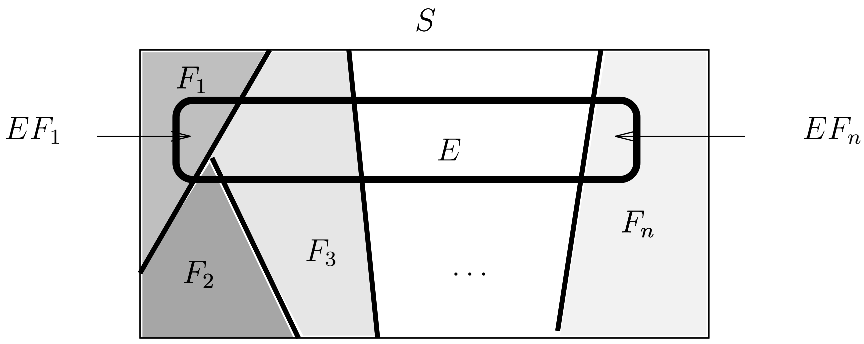


Figure 3: $E = \bigcup_{i=1}^{n} (EF_i)$.

DeMorgan's laws are as follows

$$\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i^c$$

$$\left(\bigcap_{i=1}^{n} E_i\right)^c = \bigcup_{i=1}^{n} E_i^c$$

Note that F and F^c parition the space, i.e., $FF^c = \emptyset$ and $F \cup F^c = S$, as a result, for any given E we have,

$$E = EF^c \cup EF, \quad \text{where} \quad (EF^c)(EF) = \emptyset$$

This is illustrated in Fig. 4.

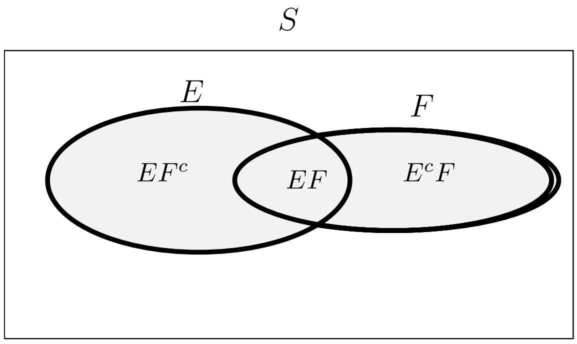


Figure 4: The intersection of events E and F showing $E = EF^c \cup EF, \quad \text{where} \quad (EF^c)(EF) = \emptyset$.

2.3 Axioms of Probability

Relative Frequency: Consider a random experiment with a sample space S and let E be a particular outcome of S. This means that the event (outcome) E is an element of the set S, i.e., $E \in S$. Assume that the experiment is repeated for n times and the total number of times that the event E has occurred is equal to $n(E)$. The ratio $n(E)/n$ is called the *relative frequency* of the event E. Relative frequency of an event has the property that $0 \le n(E)/n \le 1$, where $n(E)/n = 0$ if E occurs in none of the n trials and $n(E)/n = 1$ if E occurs in all of the n trials.

Probability: The *probability* of the event E is defined as,

$$P(E) = \lim_{n\to\infty} \frac{n(E)}{n} \tag{2}$$

The probability $P(E)$ of an event E satisfies the following axioms:

(i) $0 \le P(E) \le 1$.

(ii) $P(S) = 1$.

(iii) If $E_1, E_2, \ldots$ are disjoint events, i.e., $E_iE_j = \emptyset$ when $i \neq j$, then $P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$

As a result of above axioms:

$$P(\emptyset) = 0$$

2.4 Some Simple Propositions

The probability $P(E)$ of an event E satisfies the following properties:

(i) $P(E^c) = 1 - P(E)$.

(ii) If $E \subset F$, then $P(E) \le P(F)$.

(iii) $P(E \cup F) = P(E) + P(F) - P(EF)$.

(iv) $$P(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{i=1}^{n} P(E_i) - \sum_{i_1<i_2} P(E_{i_1}E_{i_2}) + \cdots + (-1)^{r+1} \sum_{i_1<i_2<\cdots i_r} P(E_{i_1}E_{i_2}\cdots E_{i_r}) + \cdots + (-1)^{n+1} P(E_1E_2\cdots E_n)$$

where the sum $(-1)^{r+1}\sum_{i_1<i_2<\cdots i_r} P(E_{i_1}E_{i_2}\cdots E_{i_r})$ is over $\binom{n}{r}$ terms.

2.5 Sample Spaces Having Equally Likely Outcomes

For a Sample Space having N equally likely outcomes,

$$P(\{1\}) = P(\{2\}) = \cdots = P(\{N\}) = \frac{1}{N}$$

Probability of any event E is

$$P(E) = \frac{\text{number of points in } E}{\text{number of points in } S}$$

2.6 Some Solved Problems

1. Let A, B and C be three arbitrary sets in a sample space S. In terms of these sets find set-theoretic expressions for the following events in terms of A, B and C: (i) Only A occurs, (ii) Both A and B but not C occur (iii) All three events occur (iv) At least one event occurs (v) At least two events occur (vi) Exactly one event occurs (vii) Exactly two events occur (viii) No event occurs (ix) Not more than two of the events occur.

   Solution:
   (i) AB^cC^c (ii) ABC^c
   (iii) ABC (iv) $A \cup B \cup C$
   (v) $(AB) \cup (AC) \cup (BC)$ (vi) $(AB^cC^c) \cup (BA^cC^c) \cup (CA^cB^c)$
   (vii) $(ABC^c) \cup (AB^cC) \cup (A^cBC)$ (viii) $A^cB^cC^c$
   (ix) $(ABC)^c = A^c \cup B^c \cup C^c$

2. A three digit number (000 to 999) is chosen at random. Find the probability that exactly one of the digits in the number is greater than 5. Find the probability that at least one of the digits in the number is greater than 5.

   Solution: One of the digits must be $6, 7, 8$ or 9 and the other two must be 0 to 5, inclusive. There are three ways to choose the digit that will be greater than 5 and 4 ways to fill the position. There are 6×6 ways of filling the remaining two positions. The total number of points satisfying the criteria is $3 \times 4 \times 6 \times 6 = 432$ points and the probability is .432.

3. A coin is tossed until for the first time the same result appears twice in succession. To every possible outcome requiring n tosses, attribute a probability of $1/2^{n-1}$. Describe the sample space of this experiment. Find the probability of the events: (i) The experiment stops before the 6 toss. (ii) An even number of tosses is required.

   Solution: Sample space:

		probability
HH	TT	1/4
HTT	THH	1/8
HTHH	THTT	1/16
...	...	.

   (i) Probability experiment ends before the sixth toss is $2 \times (1/4 + 1/8 + 1/16 + 1/32) = 15/16$.

   (ii) Probability experiment ends on an even number of tosses:

$$2 \times (\frac{1}{4} + \frac{1}{16} + \cdots) = 2 \times \frac{1/4}{1 - 1/4} = 2/3$$

4. A box contains five tags marked with the integers 1 to 5. Three tags are drawn from the box without replacement. Describe the sample space and the following events: (i) E_1 is the sum of the tag numbers is 12 (ii) E_2 corresponds to at least one tag of the tags being a 5 (iii)E_1E_2 (iv) Repeat the question if the experiment is with replacement.

   Solution: The sample space consists of 3-tuples of distinct integers from 1 to 5 and there are $5 \times 4 \times 3$ such points.

(i) E_1 is the set of 3-tuples whose integers add to 12 - there are six such points: (3,4,5) and its permutations.

(ii) E_2 is the set of points containing a 5. For a 5 in the first position, there are 4×3 ways of filling the other two positions. As there are 3 ways of placing the 5, the total number of points in E_2 is $3 \times 4 \times 3 = 36$.

(iii) $E_1 E_2 = E_1$.

(iv) The sample space now contains 5^3 points. E_1 consists of the points (5,5,2) and its permutations (3 points), (5,4,3) and its permutations (6 points) and (4,4,4) . E_2 consists of points of the form: (5,5,5), (5,5,X) where X<5 and its permutations (12 points) and (5,X,Y), both X and Y<5 and their permutations (48 points). $E_1 E_2$ consists of the points adding to 12 with at least one 5; namely the points (5,5,2) and its permutations (3 points) and (5,4,3) and its permutations (6 points).

5. An experiment consists of choosing two people and determining which day of the week their birthday falls on in the current year. Describe a sample space for this experiment and describe the following events (i) E_1 the birthdays fall on the same day (ii) E_2 at least one birthday is on a Friday.

   Solution: Without loss of generality assume there is an order in the experiment i.e. the two people are questioned sequentially. The the sample space consists of the 49 points:

$$\mathcal{S} = \{(x, y), \quad x, \ y = \text{Sunday, Monday, } \cdots, \text{ Saturday } \}$$

   where the first coordinate refers to the first person and the second, the second. Then

$$E_1 = \{(x, x)\}, \qquad | E_1 |= 7$$

   and

$$E_2 = \{(\text{Friday}, y), \ (x, \text{Friday})\}, \qquad | E_2 |= 13.$$

6. A die is tossed three times. Describe a suitable sample space for this experiment. Describe the following events and determine the number of points:

   (i) $A = \{$ exactly two sixes were obtained $\}$
   (ii) $B = \{$ the sum of the faces showing is 12 $\}$
   (iii) $C = \{$ all faces show an even number $\}$

   Describe also the events $AB, \ AC, \ ABC$.

   Solution: The sample space might consist of the 6^3 points $\{(i, j, k), \ 1 \leq i, j, k \leq 6\}$ where the ith coordinate reports the outcome on the ith toss.

   (i) $A = \{2$ sixes $\}$ and contains points of the form

$$\{(6, 6, 1), \ (6, 6, 2), \cdots (1, 6, 6), \ \cdots (5, 6, 6)\}$$

   Clearly $| A |= 15$.

   (ii) $B = \{$sum of faces is $12\}$ and contains points such as

$$(1, 6, 5), \ (1, 5, 6), \ (2, 6, 4), \ (2, 4, 6), \cdots$$

   and contains, by direct counting, 25 points.

   (iii) $C = \{$all faces even$\}$ and contains 3^3 points such as $(2, 2, 2), \ (2, 2, 4), \cdots$.

   By direct verification, $AB = \emptyset$, AC contains six points and $ABC = \emptyset$.

7. Ten people enter an elevator on the ground floor and get off, at random, at one of eight floors above ground. Describe a suitable sample space and find the probability that exactly one person gets off on the fifth floor.

   Solution: An appropriate sample space for this situation is that of dropping ten balls (people) into eight cells (floors) which consists of 8^{10} points i.e. ten-tuples with each coordinate place containing an integer from 1 to 8. Each point is equally likely and the probability that exactly one person gets off on the fifth floor is the number of points containing one 5 divided by 8^{10}. The number of such points is 10×7^9 since the coordinate position containing the 5 can be one of 10, and the other coordinate positions can contain any of 7 integers (anything but a 5).

8. You are dealt two cards from a standard deck of 52 cards. What is the probability you will obtain a pair (both cards have the same number or picture on them)? What is the probability the two cards will have the same suit? (It is assumed the first card is NOT replaced into the deck before the second card is dealt.)

   Solution: There are $\binom{52}{2}$ distinct pairs of cards possible (no order). Within a given number (or picture) card there are $\binom{4}{2}$ ways of obtaining a pair. There are 13 such numbers and the total number of ways of obtaining a pair is $13 \times \binom{4}{2}$. The probability is then

$$P(\text{pair}) = \frac{13 \times \binom{4}{2}}{\binom{52}{2}} .$$

   By a similar reasoning the probability of obtaining a pair with the same suit is

$$P(\text{pair with the same suit}) = \frac{4 \times \binom{13}{2}}{\binom{52}{2}} .$$

9. Choose a positive integer at random and cube it. What is the probability the last two digits of the cube are both ones?

   Solution: Consider the base 10 expansion of an integer $x + y \cdot 10 + z \cdot 10^2 + \cdots$ $(0 \le x, y, z \le 9)$, and cube it to give:

$$x^3 + 3x^2y \cdot 10 + 3xy^2 \cdot 10^2 + \cdots$$

   The only digit of the original number is x and the only digit, when cubed gives a 1 in the least significant position is 1. Thus $x = 1$. Similarly for the second least significant digit to be a 1 we must the least significant digit of $3x^2y$ be a 1 and the only value for which this happens, give $x = 1$ is $y = 7$. Thus only cubes of numbers ending in 71 gives numbers ending in 11. Thus the probability of the cube of a randomly chosen integer ending in 11 is approximately .01 (approximately because we could choose single digit numbers etc.

10. Each of nine balls are placed with equal probability in one of three boxes. Find the probability that:

    (a) there will be three balls in each box..

(b) There will be four balls in one box, three in another and two in the other.

Solution: (a) The sample space contains ordered 9-tuples, each position filled with one of three integers - total number of points is 3^9. The number of these points containing 3 ones, 3 twos and 3 threes is

$$\binom{9}{3} \times \binom{6}{3} \times \binom{3}{3}$$

and the probability follows.

(b) Similarly

$$P(4 \text{ in one box, } 3 \text{ in another and } 2 \text{ in the other}) = \frac{\binom{9}{4}\binom{5}{3}\binom{2}{2}}{3^9}.$$

11. In a batch of manufactured units, 2% of the units have the wrong weight (and perhaps also the wrong color), 5% have the wrong color (and perhaps also the wrong weight), and 1% have both the wrong weight and the wrong color. A unit is taken at random from the batch. What is the probability that the unit is defective in at least one of the two respects?

    Solution: Let A be the event that "the unit has the wrong weight" and B the event that "the unit has the wrong color". The model that we have chosen then shows that

    $$P(A) = 0.02; \quad P(B) = 0.05; \quad P(AB) = 0.01.$$

    Using the "Addition Theorem for Two Events" which states that

    $$P(A \cup B) = P(A) + P(B) - P(AB)$$

    we conclude that the desired probability is given by

    $$P(A \cup B) = P(A) + P(B) - P(AB) = 0.02 + 0.05 - 0.01 = 0.06.$$

12. Let us select one of the five-digit numbers 00000, 00001,..., 99999 at random. What is the probability that the number only contains the digits 0 and 1?

    Solution: We let each possible number correspond to one outcome. The number of possible cases is then 10^5. The number of favorable cases is seen to be 2^5, for there are two favorable digits in each of the five positions. Using the *First Definition of Probability* which states that if the probability of the outcomes in a given experiment are the same, then the probability of an event is the ratio of the number of favorable cases to the number of possible cases, we conclude that the desired probability is $2^5/10^5$=0.00032.

    Now let us compute the probability that the digits in the random number are all different. The number of favorable cases is now $10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 = 30240$ (any of the 10 digits can come first, then 1 of 9 digits can come second, and so on). Hence the probability is $30240/10^5 = 0.3024$.

13. Consider the event of rolling two dice. Determine the probability that at least one 4 appears on the dice.

Solution: The sample space is as follows:

$$\begin{array}{cccccc} (1,1) & (2,1) & (3,1) & (4,1) & (5,1) & (6,1) \\ (1,2) & (2,2) & (3,2) & (4,2) & (5,2) & (6,2) \\ (1,3) & (2,3) & (3,3) & (4,3) & (5,3) & (6,3) \\ (1,4) & (2,4) & (3,4) & (4,4) & (5,4) & (6,4) \\ (1,5) & (2,5) & (3,5) & (4,5) & (5,5) & (6,5) \\ (1,6) & (2,6) & (3,6) & (4,6) & (5,6) & (6,6) \end{array}$$

There 36 elements in the sample space. The number of elements of sample space having at least one four is equal to 11. The corresponding probability is 11/36. Note that we could also solve this problem using the expression for the probability of the union of two events.

14. Consider the event of rolling two dice. To visualize the set of outcomes, it is useful to use the following table:

$$\begin{array}{cccccc} (1,1) & (2,1) & (3,1) & (4,1) & (5,1) & (6,1) \\ (1,2) & (2,2) & (3,2) & (4,2) & (5,2) & (6,2) \\ (1,3) & (2,3) & (3,3) & (4,3) & (5,3) & (6,3) \\ (1,4) & (2,4) & (3,4) & (4,4) & (5,4) & (6,4) \\ (1,5) & (2,5) & (3,5) & (4,5) & (5,5) & (6,5) \\ (1,6) & (2,6) & (3,6) & (4,6) & (5,6) & (6,6) \end{array}$$

Compute the probability of the following events:

$$A = \{\text{sum} = 7\}, \quad B = \{8 < \text{sum} \leq 11\} \text{ and } C = \{10 < \text{sum}\}.$$

Solution: The sample space is as shown for the previous problem. The number of elements of the sample space resulting in event $A \equiv \{\text{sum} = 7\}$ is equal to $|A| = 6$ where $|A|$ shows the cardinality (number of element) of the set A. This results in $P(A) = 6/36 = 1/6$. For event $B \equiv \{8 < \text{sum} \leq 11\}$, we have, $|B| = 9 \implies P(B) = 9/36 = 1/4$ and for event $C \equiv \{10 < \text{sum}\}$, we have, $|C| = 3 \implies P(C) = 3/36 = 1/12$.

15. Given two events A and B, belonging to the same sample space, with $P(A) = 0.4$ and $P(B) = 0.7$. What are the maximum and minimum possible values for $P(AB)$?

Solution: The maximum value of $P(AB)$ is equal to $P(A) = 0.4$ and happens if $A \subset B$. The minimum value of $P(AB)$ is 0.1 and happens if $A \cup B = S$.

16. A and B have a and b dollars, respectively. They repeatedly toss a fair coin. If a head appears, A gets 1 dollar from B; if a tail appears, B gets 1 dollar from A. The game goes on until one of the players is ruined. Determine the probability that A is ruined.

Solution: The solution is easier to find if we consider a more general problem. Suppose that, at a given moment, A has n dollars, and hence B has $a + b - n$ dollars. Let P_n be the probability that, given this scenario, A is eventually ruined. We notice that after one tossing, A will have either $n+1$, or $n - 1$ dollars each with probability 1/2. Define the following events,

$W \equiv A$ has currently n dollars, A wins in the next tossing, and A is finally ruined.

$L \equiv A$ has currently n dollars, A losses in the next tossing, and A is finally ruined.

We have,

$$\begin{aligned} P(W) &= P_{n+1}/2 \\ P(L) &= P_{n-1}/2 \end{aligned}$$

$$P(A \text{ is ruined}) = P(W \cup L) = P(W) + P(L)$$

Note that W and L are disjoint. This results in,

$$P_n = \frac{1}{2} P_{n+1} + \frac{1}{2} P_{n-1}.$$

Both of the characteristics roots of this recursive equation are equal to one. From this observation we conclude that

$$P_n = C_1 + C_2 n$$

The constants C_1 and C_2 are determined by noting that $P_{a+b} = 0$ (for if A possesses all the money, he cannot be ruined) and $P_0 = 1$ (for if A has no money, he is ruined from the beginning!). Using these conditions, a simple calculation shows that

$$P_n = 1 - \frac{n}{a+b}$$

By taking $n = a$, we find that the probability is $b/(a+b)$ that A is ruined. In the same way, it is found that the probability is $a/(a+b)$ that B is ruined.
Suppose that A starts the game with $a = 10$ dollars and B with $b = 100$ dollars. The probability of ruin for A is then $100/110 = 0.91$.
Conclusion: Do not gamble with people who are richer than you!

17. In a standard deck of 52 cards there are 13 spades. Two cards are drawn at random. Determine the probability that both are spades.

    Solution: The number of favorable cases is equal to the number of combinations of 2 elements from among 13, and the number of possible cases is equal to the number of combinations of 2 elements from among 52. That is, we get

    $$P(A) = \binom{13}{2} \Big/ \binom{52}{2} = 1/17$$

18. Suppose we roll 6 dice. What is the probability of $A =$"We get exactly two 4's"?

    Solution: One way that A can occur is

    $$\frac{x}{1}\,\frac{4}{2}\,\frac{x}{3}\,\frac{4}{4}\,\frac{x}{5}\,\frac{x}{6}$$

    where x stands for "not a 4." Since the six events "die one shows x", "die two shows 4," ...,"die six shows x" are independent, the indicated pattern has probability

    $$\frac{5}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} = \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^4$$

    Here we have been careful to say "pattern" rather than "outcome" since the given pattern corresponds to 5^4 outcomes in the sample space of 6^6 possible outcomes for 6 dice. Each pattern that results in A corresponds to a choice of 2 of the 6 trials on which a 4 will occur, so the number of patterns is C_2^6. When we write out the probability of each pattern there will be two 1/6's and four 5/6's so each pattern has the same probability and

    $$P(A) = \binom{6}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^4$$

19. In a bridge hand, what is the probability each player will be dealt exactly one ace?

    Solution: Unlike the computations required for poker hands, here we need to consider simultaneously *four* 13-card hands. To visualize the basic counting technique, think of the 52 cards laid out in a row. Under 13 of the cards imagine an N; under 13 others, an S; under 13 others, an E; and under the remaining 13, a W. That permutation of N's, S's, E's, and W's would determine the hands received by North, South, East, and West, respectively.
    There are 52! ways to permute the 52 cards and there are 13! ways to permute each of the hands. Clearly, the total number of ways to deal the four hands will equal to,

    $$\frac{52!}{13!13!13!13!}$$

    By a similar argument, the aces can be distributed, one to a player, in 4!/1!1!1!1! ways, and for each of those distributions, the remaining 48 cards can be dealt in 48!/12!12!12!12! ways.
    The **Fundamental Principle** states that if operation A can be performed in m different ways and operation B in n different ways, the sequence (operation A, operation B) can be performed in $m \cdot n$ different ways.
    Thus, by the Fundamental Principle, the probability of each player receiving exactly one ace is

    $$\frac{\dfrac{4!}{(1!)^4} \cdot \dfrac{48!}{(12!)^4}}{\dfrac{52!}{(13!)^4}} = \frac{4!(13)^4 48!}{52!}$$

    or 0.105.

20. In the game of bridge the entire deck of 52 cards is dealt out to 4 players. What is the probability that one of the players receives 13 spades?

    Solution: There are $\binom{52}{13, 13, 13, 13}$ possible divisions of the cards among the four distinct players. As there are $\binom{39}{13, 13, 13}$ possible divisions of the cards leading to a fixed player having all 13 spades, it follows that the desired probability is given by

    $$\frac{4\binom{39}{13, 13, 13}}{\binom{52}{13, 13, 13, 13}} \approx 6.3 \times 10^{-12}$$

    One could also solve this problem by taking into account that 13 cards can be selected in $\binom{52}{13}$ ways and one of these selection corresponds to 13 spades, then the desired probability is equal to,

    $$\frac{4}{\binom{52}{13}}$$

    where the factor 4 accounts for the fact that there are four players each having the same probability of receiving the 13 spades.

21. If 2 balls are "randomly drawn" from a bowl containing 6 white and 5 black balls, what is the probability that one of the drawn balls is white and the other is black?

    Solution: If we regard the order in which the balls are selected as being significant, then the sample space consists of $11 \cdot 10 = 110$ points. Furthermore, there are $6 \cdot 5 = 30$ ways in which the first ball selected is white and the second black. Similarly, there are $5 \cdot 6 = 30$ ways in which the first ball is black and the second white. Hence, assuming that "randomly drawn" means that each of the 110 points in the sample space are equally likely to occur, we see that the desired probability is

    $$\frac{30+30}{110} = \frac{6}{11}$$

    This problem could also have been solved by regarding the outcome of the experiment as the (unordered) set of drawn balls. From this point of view, there would be $\binom{11}{2}$=55 points in the sample space. Using this second representation of the experiment, we see that the desired probability is

    $$\frac{\binom{6}{1}\binom{5}{1}}{\binom{11}{2}} = \frac{6}{11}$$

    which agrees with the earlier answer.

22. How many integers between 100 and 999 have distinct digits, and how many of those are odd?

    Solution: Think of the integers as being an arrangement of a hundreds digit, a tens digit, and a units digit. The hundreds digit can be filled in any of 9 ways (0s are inadmissible), the tens place in any of 9 ways (anything but what appears in the hundreds place), and the units place in any of 8 ways (the first two digits must not be repeated). Thus, the number of integers between 100 and 999 with distinct digits is $9 \cdot 9 \cdot 8$, or 648.

    $$\frac{(9)}{100\text{s}} \quad \frac{(9)}{10\text{s}} \quad \frac{(8)}{1\text{s}}$$

    To compute the number of *odd* integers with distinct digits, we first consider the units place, where any of 5 integers can be positioned (1,3,5,7, or 9). Then, turning to the hundreds place, we have 8 choices (the 0 is inadmissible; so is whatever appeared in the units place). The same number of choices is available for the tens place. Multiplying these numbers together gives $8 \cdot 8 \cdot 5 = 320$ as the number of odd integers in the range 100-999.

23. In a deck of cards there are 52 cards consisting of 4 suits with 13 denominations in each. A poker deal contains 5 randomly selected cards. A "full house" means that the player receives 3 cards of one denomination and 2 cards of another denomination; "three-of-a-kind" means that he gets 3 cards of one denomination, 1 of another denomination and 1 of a third denomination. Determine the probability of a full house and the probability of a three-of-a-kind.

    Solution: (a) Full House

    The number of possible poker deals is $\binom{52}{5}$. The favorable cases are found as follows.

The denominations for the 3-cards and the 2-cards can be selected in $13 \cdot 12$ ways. There are $\binom{4}{3}$ ways of selecting 3 cards from 4 cards with the same denomination; analogously, there are $\binom{4}{2}$ ways of taking out 2 cards. Hence the number of favorable cases is $13 \cdot 12 \cdot \binom{4}{3} \cdot \binom{4}{2}$, and the probability we want becomes

$$13 \cdot 12 \binom{4}{3}\binom{4}{2} \Big/ \binom{52}{5} \approx \frac{1}{694}$$

(b) Three-of-a-kind

The denomination for the 3-cards, the 1-card and the 1-card can be chosen in $13 \cdot \binom{12}{2}$ ways. Hence we find, in about the same way as in (a), the probability

$$13 \cdot \binom{12}{2}\binom{4}{3}\binom{4}{1}\binom{4}{1} \Big/ \binom{52}{5} \approx \frac{1}{47}$$

24. Suppose we select two marbles at random from a bag containing four white and six red marbles. What is the probability that we select one of each color?

Solution: There are C_2^{10} outcomes in the sample space. We can choose one white marble in C_1^4 ways and one red marble in C_1^6 ways.Thus the desired probability is

$$P(\text{one white, one red}) = \frac{\binom{4}{1}\binom{6}{1}}{\binom{10}{2}} = \frac{24}{45}$$

Alternatively, we may call the event "one white, one red" the event (w, r) $\cup$ (r, w) and since these are disjoint events, the probability is

$$(4/10)(6/9) + (6/10)(4/9) = 24/45$$

25. A random experiment consists of selecting at random an integer between 100 and 499. What is the probability of the following events: (a) "the selected integer contains at least one 1", and (b) "the selected integer contains exactly two 2's"?

Solution: The number of possible outcomes is 400.
(a) Let E be the desired event. It is more simple to compute $P(\bar{E})$ first. For $\bar{E}$ to happen the first digit can be 2, 3, or 4, and the other two digits can be any of 2, 3, . . . , 9. Thus

$$P(E) = 1 - P(\bar{E}) = 1 - \frac{3\,.\,9\,.\,9}{400} = 1 - 0.6075 = 0.3923$$

(b) For this event we can have the following cases: x 2 2 , 2 x 2, 2 2 x. In the first case x can be 1, 3, or 4, and in the two latter cases x can be any digit except 2.Thus

$$P(\text{exactly two 2's}) = \frac{3 + 9 + 9}{400} = .0525$$

26. Suppose we pick 4 balls out of an urn with 12 red balls and 8 black balls. What is the probability of $B =$ "We get two balls of each color"?

    Solution: There are
$$\binom{20}{4} = \frac{20 \cdot 19 \cdot 18 \cdot 17}{1 \cdot 2 \cdot 3 \cdot 4} = 5 \cdot 19 \cdot 3 \cdot 17 = 4845$$
    ways of picking 4 balls out of the 20. To count the number of outcomes in B, we note that there are $\binom{12}{2}$ ways to choose the red balls and $\binom{8}{2}$ ways to choose the black balls, so the multiplication rule implies
$$|B| = \binom{12}{2}\binom{8}{2} = \frac{12 \cdot 11}{1 \cdot 2} \cdot \frac{8 \cdot 7}{1 \cdot 2} = 6 \cdot 11 \cdot 4 \cdot 7 = 1,848$$
    It follows that $P(B)$ =1848/4845=0.3814.

27. In a game of bridge, each of the four players gets 13 cards. If North and South have 8 spades between them, what is the probability that East has 3 spades and West has 2?

    Solution: We can imagine that first North and South take their 26 cards and then East draws his 13 cards from the 26 that remain. Since there are 5 spades and 21 nonspades, the probability he receives 3 spades and 10 nonspades is $\binom{5}{3}\binom{21}{10} / \binom{26}{13}$. To compute the last probability it is useful to observe that
$$\frac{\binom{5}{3}\binom{21}{10}}{\binom{26}{13}} = \frac{\frac{5!}{3!2!} \cdot \frac{21!}{10!11!}}{\frac{26!}{13!13!}} = \frac{\frac{13!}{3!10!} \cdot \frac{13!}{2!11!}}{\frac{26!}{5!21!}} = \frac{\binom{13}{3}\binom{13}{2}}{\binom{26}{5}}$$
    To arrive at the answer on the right-hand side directly, think of 26 blanks, the first thirteen being East's cards, the second thirteen being West's. We have to pick 5 blanks for spades, which can be done in $\binom{26}{5}$ ways, while the number of ways of giving East 3 spades and West 2 spades is $\binom{13}{3}\binom{13}{2}$. After some cancellation the right-hand side is
$$\frac{13 \cdot 2 \cdot 11 \cdot 13 \cdot 6}{26 \cdot 5 \cdot 23 \cdot 22} = \frac{22308}{65780} = 0.3391$$
    Multiplying the last answer by 2, we see that with probability 0.6783 the five outstanding spades will be divided 3-2, that is, one opponent will have 3 and the other 2. Similar computations show that the probabilities of 4-1 and 5-0 splits are 0.2827 and 0.0391.

28. In an urn there are N slips of paper marked $1, 2, .., N$. Slips are drawn at random, one at a time, until the urn is empty. If slip no. i is obtained at the ith drawing, we say that a "rencontre" has occurred. Find the probability of at least one rencontre. This problem is called the problem of rencontre or the matching problem.

    Solution: If $A_i =$"rencontre at the ith drawing", we can write the required probability P as
$$P = P(\bigcup_{1}^{N} A_i)$$

The *Addition Theorem for Three Events* states that:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$

By generalizing this theorem to N events, we have the general addition formula

$$P(\bigcup_1^N A_i) = \sum_i P(A_i) - \sum_{i<j} P(A_i A_j) + \sum_{i<j<k} P(A_i A_j A_k) + \ldots + (-1)^{N-1} P(\bigcap_1^N A_i)$$

In the first sum there are N terms, in the second sum $\binom{N}{2}$ terms, and so on. Consider a general term $P(A_{i_1} A_{i_2} ... A_{i_r})$, which expresses the probability of rencontre in drawings $i_1, i_2, ..., i_r$. Let us compute the probability. The total number of possible cases for the N drawings is $N!$. Favorable cases are where the slips $i_1, i_2, ..., i_r$ appear in the drawings with these numbers, while the remaining slips can appear in any order in the other drawings; this gives $(N-r)!$ possibilities. Hence we have

$$P(A_{i_1} A_{i_2} ... A_{i_r}) = (N-r)!/N!$$

If this is inserted into the expression given before, we find

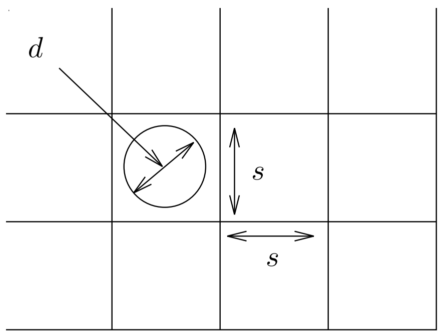
$$P(\bigcup_1^N A_i) = N \cdot \frac{(N-1)!}{N!} - \binom{N}{2} \cdot \frac{(N-2)!}{N!} + \cdots + (-1)^{N-1} \binom{N}{N} \frac{1}{N!}$$

or, after a reduction,

$$P = 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{N-1} \frac{1}{N!}$$

For large N this is approximately equal to $1 - e^{-1} = 0.63$.
The problem of rencontre was first discussed by Montmort at the beginning of the eighteenth century.

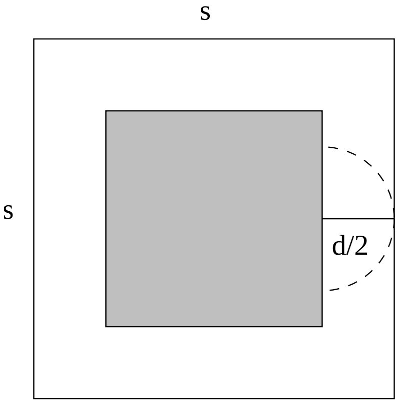
29. A carnival operator wants to set up a ring-toss game. Players will throw a ring of diameter d on to a grid of squares, the side of each square being of length s (see the following figure). If the ring lands entirely inside a square, the player wins a prize. To ensure a profit, the operator must keep the player's chances of winning down to something less than one in five. How small can the operator make the ratio d/s?



Solution: First, it will be assumed that the player is required to stand far enough away so that no skill is involved and the ring is falling at random on the grid. We see that in order for the ring not to touch any side of the square, the ring's center must be somewhere in the interior of a smaller square, each side of which is a distance $d/2$ from one of the grid lines. Since the area of a grid square is s^2

and the area of an interior square is $(s-d)^2$, the probability of a winning toss can be written as a simple ratio:

$$P(\text{ring touches no lines}) = \frac{(s-d)^2}{s^2}$$



But the operator requires that

$$\frac{(s-d)^2}{s^2} \leq 0.20$$

Solving for d/s gives

$$\frac{d}{s} \geq 1 - \sqrt{0.20} = 0.55$$

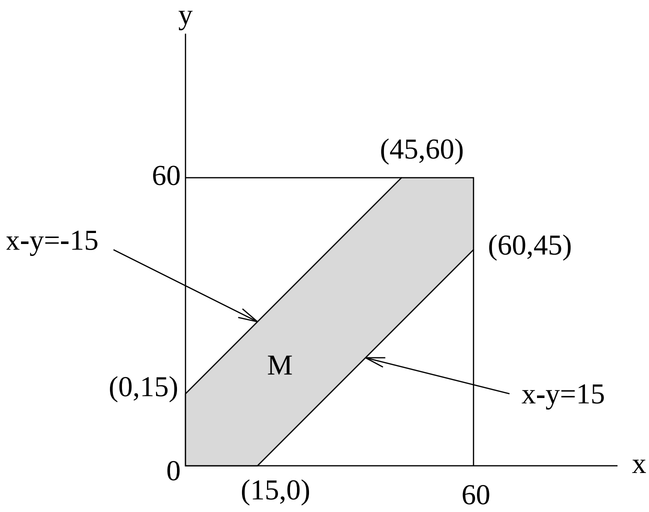
That is, if the diameter of the ring is at least 55% as long as the side of one of the squares, the player will have no more than a 20 % chance of winning.

30. Two friends agree to meet "sometime around 12:30." But neither of them is particularly punctual - or patient. What will actually happen is that each will arrive at random sometime in the interval from 12:00 to 1:00. If one arrives and the other is not there, the first person will wait 15 min or until 1:00, whichever comes first, and then leave. What is the probability the two will get together?

    Solution: To simplify notation, we can represent the time period from 12:00 to 1:00 as the interval from 0 to 60 min. Then if x and y denote the two arrival times, the sample space is the 60×60 square shown in the figure below. Furthermore, the event M, "the two friends meet," will occur if and only if $|x-y| \leq 15$ or, equivalently, if and only if $-15 \leq x-y \leq 15$. These inequalities appear as the shaded region in the figure below.

    Notice that the areas of the two triangles above and below M are each equal to $\frac{1}{2}(45)(45)$. It follows that the two friends have a 44% chance of meeting:

$$P(M) = \frac{\text{area of M}}{\text{area of S}} = \frac{(60)^2 - 2\left[\frac{1}{2}(45)(45)\right]}{(60)^2} \approx 0.44$$



3 Conditional Probability and Independence

3.1 Introduction

In this chapter, we introduce one of the most important concepts in probability theory, that of conditional probability. The importance of this concept is twofold. In the first place, we are often interested in calculating probabilities when some partial information concerning the result of the experiment is available; in such a situation the desired probabilities are conditional. Second, even when no partial information is available, conditional probabilities can often be used to compute the desired probabilities more easily.

3.2 Conditional Probabilities

Conditional Probability: Consider two events E and F which are somehow interrelated (are dependent on each other). The conditional probability of E given F is defined as,

$$P(E|F) = \frac{P(EF)}{P(F)}, \quad \text{if} \quad P(F) \neq 0.$$

If $P(F) = 0$, then the conditional probability is not defined.

The multiplication rule

$$P(E_1E_2E_3\cdots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2)\cdots P(E_n|E_1E_2\ldots E_{n-1})$$

3.3 Bayes' formula

For any two events E and F,

$$\begin{aligned} P(E) &= P(EF) + P(EF^c) \\ &= P(E|F)P(F) + P(E|F^c)P(F^c) \\ &= P(E|F)P(F) + P(E|F^c)[1 - P(F)] \end{aligned}$$

Bayes' formula: Suppose that $F_1, F_2, \ldots, F_n$ are mutually exclusive events such that $\bigcup_{i=1}^{n} F_i = S$, then

$$\begin{aligned} P(F_j|E) &= \frac{P(EF_j)}{P(E)} \\ &= \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)} \end{aligned}$$

3.4 Independent Events

Two Independent Events: Two events E and F are *independent* if

$$P(EF) = P(E)P(F)$$

then $P(E|F) = P(E)$ and $P(F|E) = P(F)$.
If E and F are independent, then so are E and F^c, as well as E^c and F and also E^c and F^c.
Three Independent Events: The three events E, F and G are independent if

$$\begin{gathered} P(EFG) = P(E)P(F)P(G) \\ P(EF) = P(E)P(F) \\ P(EG) = P(E)P(G) \\ P(FG) = P(F)P(G) \end{gathered}$$

Independent Events: Similarly, the events $E_i, i = 1, 2, \ldots, n$, are called independent if and only if for any collection of r distinct indices $(\alpha_1, \alpha_2, \ldots, \alpha_r)$ chosen from the set $1, 2, \ldots, n$, we have,

$$P(E_{\alpha_1} E_{\alpha_2} \cdots E_{\alpha_r}) = P(E_{\alpha_1}) P(E_{\alpha_2}) \cdots P(E_{\alpha_r})$$

3.5 $P(\cdot|F)$ Is a Probability

Conditional Probabilities satisfy all of the properties of ordinary probabilities. That is:

(a) $0 \leq P(E|F) \leq 1$.

(b) $P(S|F) = 1$.

(c) If $E_1, E_2, \ldots$ are disjoint events, then $P\left(\bigcup_i E_i|F\right) = \sum_i P(E_i|F)$

3.6 Some Solved Problems

1. Consider a die with 1 painted on three sides, 2 painted on two sides, and 3 painted on one side. If we roll this die ten times what is the probability we get five 1's, three 2's and two 3's?

   Solution: The answer is
$$\frac{10!}{5!3!2!}\left(\frac{1}{2}\right)^5\left(\frac{1}{3}\right)^3\left(\frac{1}{6}\right)^2$$
   If we have a group of n objects to be divided into m groups of size $n_1, ..., n_m$ with $n_1 + \cdots + n_m = n$ this can be done in
$$\frac{n!}{n_1!n_2!\cdots n_m!}\text{ ways}$$
   The first factor, in the answer above, gives the number of ways to pick five rolls for 1's, three rolls for 2's, and two rolls for 3's. The second factor gives the probability of any outcome with five 1's, three 2's, and two 3's. Generalizing from this example, we see that if we have k possible outcomes for our experiment with probabilities $p_1, ..., p_k$ then the probability of getting exactly n_i outcomes of type i in $n = n_1 + ... + n_k$ trials is
$$\frac{n!}{n_1!\cdots n_k!}p_1^{n_1}\cdots p_k^{n_k}$$
   since the first factor gives the number of outcomes and the second the probability of each one. We referred to this as the *Multinomial Distribution.*

2. Suppose an urn contains $r = 1$ red chip and $w = 1$ white chip. One is drawn at random. If the chip selected is red, that chip along with $k = 2$ additional red chips are put back into the urn. If it is white, the chip is simply returned to the urn, then a second chip is drawn. What is the probability that both selections are red?

   Solution: If we let R_1 be the event "red chip is selected on first draw" and R_2 be "red chip is selected on second draw," it should be clear that
$$P(R_1) = \frac{1}{2} \quad \text{and} \quad P(R_2|R_1) = \frac{3}{4}$$
   We also know that
$$P(AB) = P(A|B)P(B)$$
   Hence, substituting the probabilities we found into the above equation gives
$$\begin{aligned} P(R_1R_2) &= P(\text{both chips are red}) = P(R_1)P(R_2|R_1) \\ &= \left(\frac{1}{2}\right)\left(\frac{3}{4}\right) = \frac{3}{8} \end{aligned}$$

3. At a party n men take off their hats. The hats are then mixed up and each man randomly selects one. We say that a match occurs if a man selects his own hat.
   1. What is the probability of no matches?
   2. What is the probability of exactly k matches?

   Solution: We can solve this problem using the corresponding formulas for matching problem. However, in the following, we examine a different approach using conditional probability.

Let E denote the event that no matches occur, and to make explicit the dependence on n write $P_n = P(E)$. We start by conditioning on whether or not the first man selects his own hat-call these events M and M^c. Then

$$P_n = P(E) = P(E|M)P(M) + P(E|M^c)P(M^c) \quad (1)$$

Clearly, $P(E|M) = 0$, and so

$$P_n = P(E|M^c)\frac{n-1}{n}$$

Now, $P(E|M^c)$ is the probability of no matches when $n-1$ men select from a set of $n-1$ hats that does not contain the hat of one of these men. This can happen in either of two mutually exclusive ways. Either there are no matches and the extra man does not select the extra hat (this being the hat of the man that chose first), or there are no matches and the extra man does select the extra hat. The probability of the first of these events is just P_{n-1}, which is seen by regarding the extra hat as "belonging" to the extra man. As the second event has probability $[1/(n-1)]P_{n-2}$, we have

$$P(E|M^c) = P_{n-1} + \frac{1}{n-1}P_{n-2}$$

and thus, from Equation (1),

$$P_n = \frac{n-1}{n}P_{n-1} + \frac{1}{n}P_{n-2}$$

or equivalently

$$P_n - P_{n-1} = -\frac{1}{n}(P_{n-1} - P_{n-2}) \quad (2)$$

However, as P_n is the probability of no matches when n men select among their own hats, we have

$$P_1 = 0 \qquad P_2 = \frac{1}{2}$$

and so, from Equation (2),

$$P_3 - P_2 = -\frac{(P_2 - P_1)}{3} = -\frac{1}{3!} \text{ or } P_3 = \frac{1}{2!} - \frac{1}{3!}$$
$$P_4 - P_3 = -\frac{(P_3 - P_2)}{4} = \frac{1}{4!} \text{ or } P_4 = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!}$$

and, in general, we see that

$$P_n = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \cdots + \frac{(-1)^n}{n!}$$

To obtain the probability of exactly k matches, we consider any fixed group of k men. The probability that they, and only they, select their own hats is

$$\frac{1}{n}\frac{1}{n-1}\cdots\frac{1}{n-(k-1)}P_{n-k} = \frac{(n-k)!}{n!}P_{n-k}$$

where P_{n-k} is the probability that the other $n-k$ men, selecting among their own hats, have no matches. As there are $\binom{n}{k}$ choices of a set of k men, the desired probability of exactly k matches is

$$\frac{P_{n-k}}{k!} = \frac{\frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^{n-k}}{(n-k)!}}{k!}$$

4. In a batch of 50 units there are 5 defectives. A unit is selected at random, and thereafter one more from the remaining ones. Find the probability that both are defective.

   Solution: Let A be the event that "the first unit is defective" and B the event that "the second unit is defective". It is seen that $P(A) = 5/50$. If A occurs, there remain 49 units, 4 of which are defective. Hence we conclude that $P(B|A) = 4/49$, and using the following formula

   $$P(AB) = P(A)P(B|A) = P(B)P(A|B)$$

   we arrive at the probability we seek:

   $$P(AB) = \frac{5}{50} \cdot \frac{4}{49} = \frac{2}{245}$$

   Let us now draw a third unit, and let us evaluate the probability that the first two units are defective and that the third one is good. If C is the event that "the third unit is good", and we employ the top formula twice to get the following formula for three events

   $$P(ABC) = P(AB)P(C|(AB)) = P(A)P(B|A)P(C|(AB))$$

   we arrive at our final answer

   $$P(ABC) = \frac{5}{50} \cdot \frac{4}{49} \cdot \frac{45}{48} = \frac{3}{392}$$

5. In a factory, units are manufactured by machines H_1,H_2,H_3 in the proportions $25 : 35 : 40$. The percentages $5\%, 4\%$ and 2%, respectively, of the manufactured units are defective. The units are mixed and sent to the customers. (a) Find the probability that a randomly chosen unit is defective. (b) Suppose that a customer discovers that a certain unit is defective. What is the probability that it has been manufactured by machine H_1?

   Solution: (a) The *Total Probability Theorem* states that if the events H_1,H_2,...,H_n are mutually exclusive, have positive probabilities, and together fill the probability space Ω completely, any event A satisfies the formula

   $$P(A) = \sum_{i=1}^{n} P(H_i)P(A|H_i)$$

   Using this theorem, and taking $H_i =$ "unit produced by machine H_i" and A="unit is defective", we find

   $$P(A) = 0.25 \cdot 0.05 + 0.35 \cdot 0.04 + 0.40 \cdot 0.02 = 0.0345.$$

   (b) *Bayes' Theorem* states that, under the same conditions as the Total Probability Theorem (see part (a)),

   $$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{\sum_{j=1}^{n} P(H_j)P(A|H_j)}$$

   Hence Bayes' theorem gives the answer ($A =$ defective unit)

   $$P(H_1|A) = \frac{0.25 \cdot 0.05}{0.25 \cdot 0.05 + 0.35 \cdot 0.04 + 0.40 \cdot 0.02} = 0.36$$

6. In a certain country there are two nationalities living together, the Bigs and the Smalls. Among the Bigs 80% are tall, and among the Smalls 1%. A visiting tourist encounters a person at random who turns out to be tall. Determine the probability that this person belongs to the Bigs.

   Solution: Let H_1 = "the person belongs to the Bigs", H_2 = "the person belongs to the Smalls", A="the person is tall". Bayes' theorem shows that

   $$\mathrm{P}(H_1|A) = \frac{P(H_1) \cdot 0.80}{P(H_1) \cdot 0.80 + P(H_2) \cdot 0.01}.$$

   The formulation of the problem is inadequate, since it is necessary to know the probabilities $P(H_1)$ and $P(H_2)$, the proportions of Bigs and Smalls in the country.
   If the proportions are the same, so that $P(H_1) = P(H_2) = 1/2$, the probability becomes $80/81 = 0.99$. But if the Bigs are so few that $P(H_1) = 0.001$ and $P(H_2) = 0.999$, the probability is instead

   $$0.001 \cdot 0.80/(0.001 \cdot 0.80 + 0.999 \cdot 0.01) = 80/1079 = 0.0741.$$

7. Al flips 3 coins and Betty flips 2. Al wins if the number of Heads he gets is more than the number Betty gets. What is the probability Al will win?

   Solution: Let A be the event that Al wins, let B_i be the event that Betty gets i Heads, and let C_j be the event that Al gets j Heads. By considering the four outcomes of flipping two coins it is easy to see that

   $$P(B_0) = 1/4 \quad P(B_1) = 1/2 \quad P(B_2) = 1/4$$

   while considering the eight outcomes for three coins leads to

   $$\begin{aligned} P(A|B_0) &= P(C_1 \cup C_2 \cup C_3) = 7/8 \\ P(A|B_1) &= P(C_2 \cup C_3) = 4/8 \\ P(A|B_2) &= P(C_3) = 1/8 \end{aligned}$$

   Since AB_i, $i = 0, 1, 2$ are disjoint and their union is A, we have

   $$P(A) = \sum_{i=0}^{2} P(AB_i) = \sum_{i=0}^{2} P(A|B_i)P(B_i)$$

   since $P(AB_i) = P(A|B_i)P(B_i)$ by the definition of conditional probability. Plugging in the values, we obtain,

   $$P(A) = \frac{1}{4} \cdot \frac{7}{8} + \frac{2}{4} \cdot \frac{4}{8} + \frac{1}{4} \cdot \frac{1}{8} = \frac{7+8+1}{32} = \frac{1}{2}$$

8. Suppose for simplicity that the number of children in a family is $1, 2$, or 3, with probability $1/3$ each. Little Bobby has no brothers. What is the probability he is an only child?

   Solution: Let B_1, B_2, B_3 be the events that a family has one, two, or three children, and let A be the event that a family has only one boy. We want to compute $P(B_1|A)$. By the definition of conditional probability, we have,

   $$P(B_1|A) = P(B_1 A)/P(A)$$

To evaluate the numerator we again use the definition of conditional probability

$$P(B_1A) = P(B_1)P(A|B_1) = \frac{1}{3}\cdot\frac{1}{2} = \frac{1}{6}$$

Similarly, $P(B_2A) = P(B_2)P(A|B_2) = \frac{1}{3}\cdot\frac{2}{4} = \frac{1}{6}$ and

$$P(B_3A) = P(B_3)P(A|B_3) = \frac{1}{3}\cdot\frac{3}{8} = \frac{1}{8}$$

Now $P(A) = \sum_i P(B_iA)$ so

$$P(B_1|A) = \frac{P(B_1A)}{P(A)} = \frac{1/6}{1/6+1/6+1/8} = \frac{8}{8+8+6} = \frac{4}{11}$$

9. A company buys tires from two suppliers, 1 and 2. Supplier 1 has a record of delivering tires containing 10% defectives, whereas supplier 2 has a defective rate of only 5%. Suppose 40% of the current supply came from supplier 1. If a tire is taken from this supply and observed to be defective, find the probability that it came from Supplier 1.

   Solution: Let B_i denote the event that a tire comes from supplier i, $i = 1,2$, and note that B_1 and B_2 form a partition of the sample space for the experiment of selecting one tire. Let A denote the event that the selected tire is defective. Then,

$$\begin{aligned} P(B_1|A) &= \frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1)+P(B_2)P(A|B_2)} \\ &= \frac{.40(.10)}{.40(.10)+(.60)(.05)} \\ &= \frac{0.04}{.04+.03} = \frac{4}{7} \end{aligned}$$

   Supplier 1 has a greater probability of being the party supplying the defective tire than does Supplier 2.

10. Urn I contains 2 white and 4 red balls, whereas urn II contains 1 white and 1 red ball. A ball is randomly chosen from urn I and put in to urn II, and a ball is then randomly selected from urn II.
    1) What is the probability that the ball selected from urn II is white?
    2) What is the conditional probability that the transferred ball was white, given that a white ball is selected from urn II?

    Solution: Letting W be the event that the transferred ball was white, and E the event that the ball selected from II was white we obtain,

$$\begin{aligned} 1)\;\; P(E) &= P(E|W)P(W) + P(E|W^c)P(W^c) \\ &= \left(\frac{2}{3}\right)\left(\frac{2}{6}\right) + \left(\frac{1}{3}\right)\left(\frac{4}{6}\right) \\ &= \frac{4}{9} \end{aligned}$$

$$
\begin{aligned}
2) P(W|E) &= \frac{P(WE)}{P(E)} \\
&= \frac{P(E|W)P(W)}{P(E)} \\
&= \frac{1}{2}
\end{aligned}
$$

11. A laboratory blood test is 95 per cent effective in detecting a certain disease when it is, in fact, present. However, the test also yields a "false positive" result for 1 per cent of the healthy persons tested. (That is, if a healthy person is tested, then, with probability 0.01, the test result will imply he has the disease.) If 0.5 per cent of the population actually has the disease, what is the probability a person has the disease given that his test result is positive?

    Solution: Let D be the event that the tested person has the disease and E the event that his test result is positive. The desired probability $P(D|E)$ is obtained by

$$
\begin{aligned}
P(D|E) &= \frac{P(DE)}{P(E)} \\
&= \frac{P(E|D)P(D)}{P(E|D)P(D) + P(E|D^c)P(D^c)} \\
&= \frac{(.95)(.005)}{(.95)(.005) + (.01)(.995)} \\
&= \frac{95}{294} \approx .323
\end{aligned}
$$

    Thus only 32 per cent of those persons whose test results are positive actually have the disease. As you may be surprised at this result (as it is expected this figure to be much higher, since the blood test seems to be a good one), it is probably worthwhile to present a second argument which, although less rigorous than the preceding one, is probably more revealing.

    Since 0.5 per cent of the population actually has the disease, it follows that, on the average, 1 person out of every 200 tested will have it. The test will correctly confirm that this person has the disease with probability 0.95. Thus, on the average, out of every 200 persons tested the test will correctly confirm that 0.95 persons have the disease. On the other hand, however, out of the (on the average) 199 healthy people, the test will incorrectly state that (199)(0.01) of these people have the disease. Hence, for every 0.95 diseased persons that the test correctly states are ill, there are (on the average) (199)(0.01) healthy persons that the test incorrectly states are ill. Hence the proportion of time that the test result is correct when it states that a person is ill is

$$
\frac{0.95}{0.95 + (199)(0.01)} = \frac{95}{294} \approx 0.323
$$

12. A purchaser of electrical components buys them in lots of size 10. It is his policy to inspect 3 components randomly from a lot and to accept the lot only if all 3 are non-defective. If 30 per cent of the lots have 4 defective components and 70 per cent have only 1, what proportion of lots does the purchaser reject?

    Solution: Let A denote the event that the purchaser accepts a lot. Now, using the Total Probability Theorem

$$
P(A) = P(A|\text{lot has 4 defectives})\frac{3}{10} + P(A|\text{lot has 1 defective})\frac{7}{10}
$$

$$= \frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}}\left(\frac{3}{10}\right) + \frac{\binom{1}{0}\binom{9}{3}}{\binom{10}{3}}\left(\frac{7}{10}\right)$$
$$= \frac{54}{100}$$

Hence 46 per cent of the lots are rejected.

13. Suppose events A, B, and C are mutually independent. Form a composite event from A and B and call it E. Is E independent of C?

    Solution: Yes.

14. Suppose A and B are independent events. Does it follow that A^c and B^c are also independent? That is, does $P(A \cap B) = P(A)P(B)$ guarantee that $P(A^c \cap B^c) = P(A^c)P(B^c)$?

    Solution: The answer is yes, the proof being accomplished by equating two different expression for $P(A^c \cap B^c)$. First, we know that

    $$P(A^c \cup B^c) = P(A^c) + P(B^c) - P(A^c B^c)$$

    But the union of two complement is also the complement of their intersection. Therefore,

    $$P(A^c \cup B^c) = 1 - P(AB)$$

    Combining the two equations above, we get

    $$1 - P(AB) = 1 - P(A) + 1 - P(B) - P(A^c B^c)$$

    Since A and B are independent, $P(AB) = P(A) \cdot P(B)$, so

    $$\begin{aligned} P(A^c B^c) &= 1 - P(A) + 1 - P(B) - (1 - P(A)P(B)) \\ &= (1 - P(A))(1 - P(B)) \\ &= P(A^c)P(B^c) \end{aligned}$$

    the latter factorization implying that A^c and B^c are, themselves, independent.

15. We throw a fair four-sided die twice. Let E be the event that "the sum of the dice is 4", and let F be the event that "the first die thrown has value 2". Are the events E and F independent?

    Solution: The sample space for this problem consists of 16 equally likely outcomes as shown graphical in figure 5. Also shown are the two events E and F. The event EF is a single outcome (2,2) with probability 1/16.

    On the other hand, $p(E) = 3/16$ and $P(F) = 4/16$, thus

    $$\frac{3}{16} \cdot \frac{4}{16} = \frac{3}{64} \neq \frac{1}{16}$$

    The events E and F are therefore not independent.

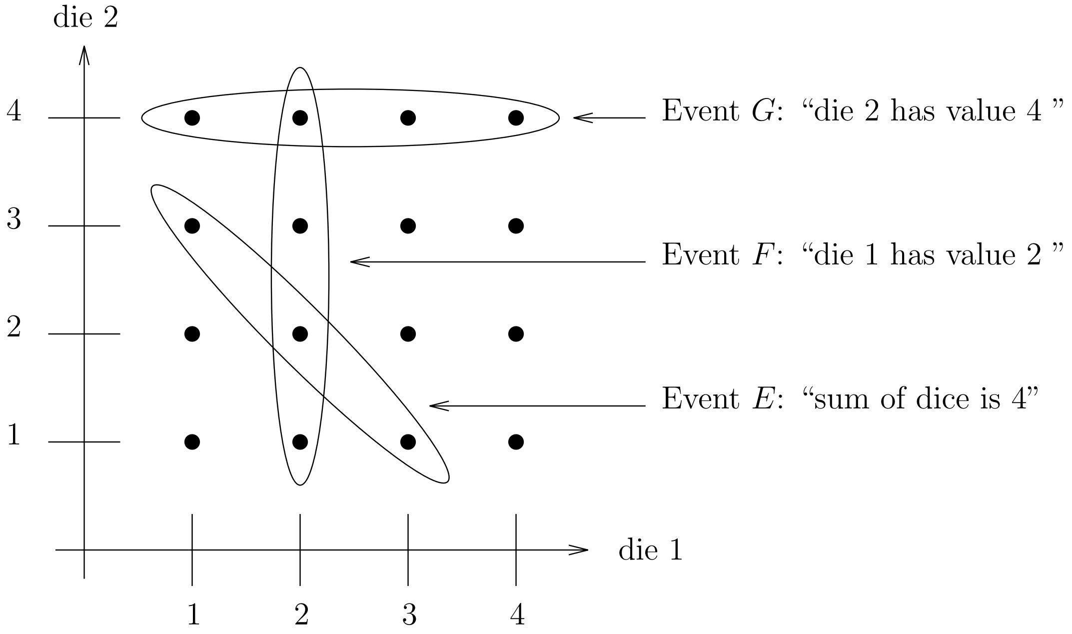
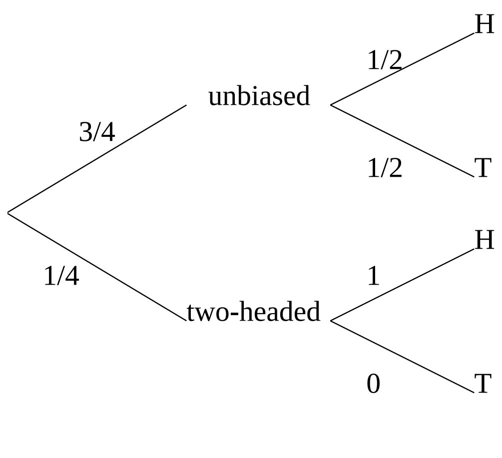


Figure 5: A sample space for two throws of a four-sided die.



16. Two coins are available, one unbiased and the other two-headed. Choose a coin at random and toss it once; assume that the unbiased coin is chosen with probability 3/4. Given that the result is head, find the probability that the two-headed coin was chosen.

    Solution: The "tree diagram" shown below represents the experiment.

    We may take Ω to consist of the four possible paths through the tree, with each path assigned a probability equal to the product of the probabilities assigned to each branch. Notice that we are given the probabilities of the events $B_1 = \{\text{unbiased coin chosen}\}$ and $B_2 = \{\text{two-headed coin chosen}\}$, as well as the conditional probabilities $P(A|B_i)$, where $A = \{\text{coin comes up head}\}$. This is sufficient to determine the probabilities of all events.
    Now we can compute $P(B_2|A)$ using the definition of conditional probability to obtain

$$\begin{aligned} P(B_2|A) &= \frac{P(B_2A)}{P(A)} \\ &= \frac{P\{\text{two-headed coin chosen and coin comes up head}\}}{P\{\text{coin comes up head}\}} \\ &= \frac{(1/4)(1)}{(3/4)(1/2)+(1/4)(1)} = \frac{2}{5} \end{aligned}$$

17. Russian roulette is played with a revolver equipped with a rotatable magazine of six shots. The revolver is loaded with one shot. The first duelist, A, rotates the magazine at random , points the revolver at his head and presses the trigger. If, afterwards, he is still alive, he hands the revolver to the other duelist, B, who acts in the same way as A. The players shoot alternately in this manner, until a shot goes off. Determine the probability that A is killed.

    Solution: Let T be the event that "A is killed on the first trial". We have

$$P(\text{A is killed}) = P(T) + P(T^c)P(\text{A is killed}|T^c).$$

    But if A survives the first trial, the roles of A and B are interchanged and so P(A is killed $|T^c$) =P(B is killed)=1-P(A is killed). Inserting this above, we find

$$P(\text{A is killed}) = \frac{1}{6} + \frac{5}{6}[1 - P(\text{A is killed})]$$

    Solving this equation we find P(A is killed)=6/11.

18. A toy manufacturer buys ball bearings from three different suppliers - 50% of his total order comes from supplier 1, 30% from supplier 2, and the rest from supplier 3. Past experience has shown that the quality control standards of the three suppliers are not all the same. Of the ball bearings produced by supplier 1, 2% are defective, while suppliers 2 and 3 produce defective bearings 3% and 4% of the time, respectively. What proportion of the ball bearings in the toy manufacturer's inventory are defective?

    Solution: Let A_i be the event "bearing came from supplier i," i=1,2,3. Let B be the event "bearing in toy manufacturer's inventory is defective." Then

$$P(A_1) = 0.5, \quad P(A_2) = 0.3, \quad P(A_3) = 0.2$$

    and

$$P(B|A_1) = 0.02, \quad P(B|A_2) = 0.03, \quad P(B|A_3) = 0.04$$

The Total Probability Theorem states that:
Let $\{A_1\}_{i=1}^n$ be a set of events defined over S such that $S = \bigcup_{i=1}^n A_i, A_iA_j = \emptyset$ for $i \neq j$, and $P(A_i) > 0$ for i=1,2,...,n. For any event B, we have,

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

Hence, combining the probabilities according to the theorem gives

$$\begin{aligned} P(B) &= (0.02)(0.5) + (0.03)(0.3) + (0.04)(0.2) \\ &= 0.027 \end{aligned}$$

meaning the manufacturer can expect 2.7% of his stock of ball bearings to be defective.

19. A man has n keys on a chain, one of which opens the door to his apartment. Having celebrated a little too much one evening, he returns home, only to find himself unable to distinguish one key from another. Resourceful, he works out a fiendishly clever plan: He will choose a key at random and try it. If it fails to open the door, he will discard it and choose one of the remaining $n-1$ keys at random, and so on. Clearly, the probability that he gains entrance with the first key he selects is $1/n$. What is the probability the door opens with the second key he tries?

    Solution: It would be tempting here to answer $1/(n-1)$, but in this case our intuition would be in error. Actually, $1/(n-1)$ is a right answer, but to a different question. To see why, let K_i, $i = 1, 2, \ldots, n$, denote the event "ith key tried opens door." Then $P(K_1)$ is certainly $1/n$, but the event "second key tried opens door" can occur only if the first key *does not* open the door. That is,

$$P(K_2) = P(K_2K_1^c)$$

    Since we also know that

$$P(AB) = P(A|B)P(B) \tag{3}$$

    applying this equation to the right-hand side of the top equation, we see that the probability that the second key tried opens the door is the same as the probability for the first key, $1/n$:

$$\begin{aligned} P(K_2K_1^c) &= P(K_2|K_1^c)P(K_1^c) \\ &= \left(\frac{1}{n-1}\right)\left(\frac{n-1}{n}\right) \\ &= \frac{1}{n} \end{aligned}$$

    Thus, the ratio $1/(n-1)$ does answer a *conditional* probability $P(K_2|K_1^c)$.

20. Urn I contains two white chips (w_1,w_2) and one red chip (r_1); urn II has one white chip (w_3) and two red chips (r_2,r_3). One chip is drawn at random from urn I and transferred to urn II. Then one chip is drawn from urn II. Suppose a red chip is selected from urn II. What is the probability the chip *transferred* was white?

    Solution: Let A_1 and A_2 denote the events "red chip is transferred from urn I" and "white chip is transferred from urn I." Let B be the event "red chip is drawn from urn II." What we are asking for is $P(A_2|B)$.

We note that $P(A_1) = \frac{1}{3}$, $P(A_2) = \frac{2}{3}$, $P(B|A_1) = \frac{3}{4}$, and $P(B|A_2) = \frac{2}{4}$.
Bayes' Theorem states:
Let $\{A_i\}_{i=1}^n$ be a set of n events, each with positive probability, that partition S in such a way that $\bigcup_{i=1}^n A_i = S$ and $A_iA_j = \emptyset$ for $i \neq j$. For any event B (also defined on S), where $P(B) > 0$,

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

for any $1 \leq j \leq n$.

Therefore, we can substitute the probabilities we have found into Bayes' formula to obtain:

$$\begin{aligned} P(A_2|B) &= \frac{P(B|A_2)P(A_2)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)} \\ &= \frac{\left(\frac{2}{4}\right)\left(\frac{2}{3}\right)}{\left(\frac{3}{4}\right)\left(\frac{1}{3}\right) + \left(\frac{2}{4}\right)\left(\frac{2}{3}\right)} \\ &= \frac{4}{7} \end{aligned}$$

21. Urn I contains five red chips and four white chips; urn II contains four red chips and five white chips. Two chips are to be transferred from urn I to urn II. Then a single chip is to be drawn from urn II. What is the probability the chip drawn from the second urn will be white?

    Solution: Let W be the event "white chip is drawn from urn II." Let A_i, i=0, 1, 2, denote the event "i white chips are transferred from urn I to urn II." By the Total Probability Theorem, we obtain,

$$P(W) = P(W|A_0)P(A_0) + P(W|A_1)P(A_1) + P(W|A_2)P(A_2)$$

    Note that $P(W|A_i) = (5 + i)/11$ and that $P(A_i)$ is gotten directly from the *hypergeometric distribution*. Therefore,

$$\begin{aligned} P(W) &= \left(\frac{5}{11}\right)\frac{\binom{4}{0}\binom{5}{2}}{\binom{9}{2}} + \left(\frac{6}{11}\right)\frac{\binom{4}{1}\binom{5}{1}}{\binom{9}{2}} + \left(\frac{7}{11}\right)\frac{\binom{4}{2}\binom{5}{0}}{\binom{9}{2}} \\ &= \left(\frac{5}{11}\right)\left(\frac{10}{36}\right) + \left(\frac{6}{11}\right)\left(\frac{20}{36}\right) + \left(\frac{7}{11}\right)\left(\frac{6}{36}\right) \\ &= \frac{53}{99} \end{aligned}$$

22. The highways connecting two resort areas at A and B are shown in figure below:
    There is a direct route through the mountains and a more-circuitous route going through a third resort area at C in the foothills. Travel between A and B during the winter months is not always possible, the roads sometimes being closed due to snow and ice.

    Suppose we let E_1, E_2, and E_3 denote the events that highways AB, AC, and CB are passable, respectively, and we know from past years that on a typical winter day,

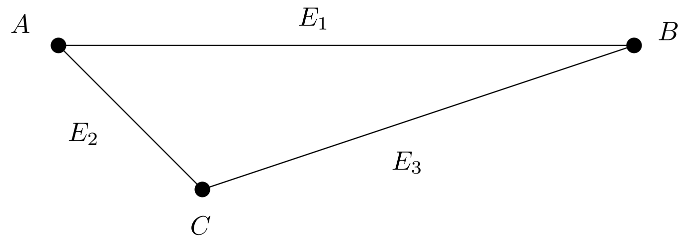


Figure 6:

$$\begin{aligned}
P(E_1) &= \frac{2}{5}, P(E_3|E_2) = \frac{4}{5} \\
P(E_2) &= \frac{3}{4}, P(E_1|E_2E_3) = \frac{1}{2} \\
P(E_3) &= \frac{2}{3}
\end{aligned}$$

What is the probability that a traveller will be able to get from A to B?

Solution: If E denotes the event that we *can* get from A to B, then

$$E = E_1 \cup (E_2E_3)$$

It follows that

$$P(E) = P(E_1) + P(E_2E_3) - P\left[E_1(E_2E_3)\right]$$

We know that,

$$\begin{aligned}
P(E) &= P(E_1) + P(E_3|E_2)P(E_2) - P\left[E_1|(E_2E_3)\right]P(E_2E_3) \\
&= P(E_1) + P(E_3|E_2)P(E_2) - P\left[E_1|(E_2E_3)\right]P(E_3|E_2)P(E_2) \\
&= \frac{2}{5} + \left(\frac{4}{5}\right)\left(\frac{3}{4}\right) - \left(\frac{1}{2}\right)\left(\frac{4}{5}\right)\left(\frac{3}{4}\right) \\
&= 0.7
\end{aligned}$$

23. A crooked gambler has nine dice in her coat pocket: Three are fair and six are biased. The biased ones are loaded in such a way that the probability of rolling a 6 is $\frac{1}{2}$. She takes out one die at random and rolls it twice. Let A be the event "6 appears on the first roll" and B be the event "6 appears on the second roll." Are A and B independent?

    Solution: Our intuition here would most probably answer yes - but, appearances not withstanding, this is not a typical dice problem. Repeated throws of a die *do* qualify as independent events *if* the probabilities associated with the different faces are known. In this situation, though, those probabilities are *not* known and depend in a random way on which die the gambler draws from her pocket.
    To see formally what effect having two different dice has on the relationship between A and B, we must appeal to the Total Probability Theorem. Let F and L denote the events "fair die selected" and "loaded die selected," respectively. Then

$$\begin{aligned}
P(AB) &= P(6 \text{ on first roll} \cap 6 \text{ on second roll}) \\
&= P(AB|F)P(F) + P(AB|L)P(L)
\end{aligned}$$

Conditional on either F or L, A and B are independent, so

$$\begin{aligned} P(AB) &= \left(\frac{1}{6}\right)\left(\frac{1}{6}\right)\left(\frac{3}{9}\right)+\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{6}{9}\right) \\ &= \left(\frac{19}{108}\right) \end{aligned}$$

Similarly,

$$\begin{aligned} P(A) &= P(A|F)P(F)+P(A|L)P(L) \\ &= \left(\frac{1}{6}\right)\left(\frac{3}{9}\right)+\left(\frac{1}{2}\right)\left(\frac{6}{9}\right) \\ &= \left(\frac{7}{18}\right) = P(B) \end{aligned}$$

But note that

$$P(AB) = \frac{19}{108} = \frac{57}{324} \neq P(A)\cdot P(B) = \left(\frac{7}{18}\right)\left(\frac{7}{18}\right) = \left(\frac{49}{324}\right)$$

proving that A and B are *not* independent.

24. Laplace's Rule of Succession. There are k+1 biased coins in a box. The ith coin will, when flipped, turn up heads with probability i/k, i=0,1,...,k. A coin is randomly selected from the box and is then repeatedly flipped. If the first n flips all result in heads, what is the conditional probability that the (n+1)st flip will do likewise?

    Solution: Let E_i denote the event that the ith coin is initially selected, i=0,1,...,k; let F_n denote the event that the first n flips all result in heads; and let F be the event that the (n+1)st flip is a head. The desired probability, $P(F|F_n)$, is now obtained as follows,

$$P(F|F_n) = \sum_{i=0}^{k} P(F|F_nE_i)P(E_i|F_n)$$

    Now, given that the ith coin is selected, it is reasonable to assume that the outcomes will be conditionally independent with each one resulting in a head with probability i/k. Hence

$$P(F|F_nE_i) = P(F|E_i) = \frac{i}{k}$$

    Also,

$$\begin{aligned} P(E_i|F_n) &= \frac{P(E_iF_n)}{P(F_n)} \\ &= \frac{P(F_n|E_i)P(E_i)}{\sum_{j=0}^{k} P(F_n|E_j)P(E_j)} \\ &= \frac{(i/k)^n[1/(k+1)]}{\sum_{j=0}^{k}(j/k)^n[1/(k+1)]} \end{aligned}$$

Hence

$$P(F|F_n) = \frac{\sum_{i=0}^{k} (i/k)^{n+1}}{\sum_{j=0}^{k} (j/k)^{n}}$$

But if k is large, we can use the integral approximations

$$\frac{1}{k}\sum_{i=0}^{k} (i/k)^{n+1} \approx \int_0^1 x^{n+1}dx = \frac{1}{n+2}$$

$$\frac{1}{k}\sum_{j=0}^{k} (j/k)^{n} \approx \int_0^1 x^{n}dx = \frac{1}{n+1}$$

and so, for k large,

$$P(F|F_n) \approx \frac{n+1}{n+2}$$

25. A bag contains 3 white balls, 6 red balls and 3 black balls. Three balls are drawn at random, without replacement. What is the probability that the three balls are the same color?

    Solution: Let W be the event the three balls are all white, R the event the three balls are all red, B the event the three balls are all black and C the event the balls are of the same color. Then by the theorem on total probability,

    $$P(C) = P(C|W)P(W) + P(C|R)P(R) + P(C|B)P(B) =$$

    $$\left[1 \cdot \binom{3}{3} + 1 \cdot \binom{6}{3} + 1 \cdot \binom{3}{3}\right] \cdot \frac{1}{\binom{12}{3}}$$

26. From a standard deck of 52 cards, one red one is taken. Thirteen cards are then chosen and found to be of the same color. What is the probability they are black?

    Solution: Given that the 13 cards are of the same color, they could be red (R) or black (B). Denote by C the event that all 13 cards are of the same color. It is required to find $P(B|C)$. By Bayes theorem:

    $$P(B|C) = \frac{P(C|B)P(B)}{P(C|B)P(B) + P(C|R)P(R)} = \frac{\binom{26}{13}/\binom{51}{13}}{\binom{26}{13}/\binom{51}{13} + \binom{25}{13}/\binom{51}{13}} .$$

    Since $P(C|R) = P(C|B) = 1$, then

    $$P(B|C) = \frac{1}{1 + \binom{25}{13}/\binom{26}{13}} = \frac{1}{1 + \frac{1}{2}} = \frac{2}{3} .$$

    . and hence $P(C) = \binom{26}{13}/\binom{51}{13} + \binom{25}{13}/\binom{51}{13}$.

27. A fair die is tossed. If the face shows i then a fair coin is tossed i times. What is the probability of obtaining at least 3 heads on any given turn?

Solution: Let A be the event of obtaining at least three heads on any outcome of the experiment and let B_i be the event of obtaining i on the toss of a die, $i = 1, 2, \cdots, 6$. By the theorem on total probability,

$$P(A) = \sum_{i=1}^{6} P(A|B_i)P(B_i) \ .$$

Now $P(B_i) = 1/6$ and $P(A|B_1) = P(A|B_2) = 0$ and

$$P(A|B_3) = \frac{1}{8}$$
$$P(A|B_4) = \left(\binom{4}{3} + \binom{4}{4}\right)\left(\frac{1}{2}\right)^4$$
$$P(A|B_5) = \left(\binom{5}{3} + \binom{5}{4} + \binom{5}{5}\right)\left(\frac{1}{2}\right)^5$$
$$P(A|B_6) = \left(\binom{6}{3} + \binom{6}{4} + \binom{6}{5} + \binom{6}{6}\right)\left(\frac{1}{2}\right)^6$$

and the result follows.

28. If A and B are independent events and the probability that both A and B occur is 0.16 while the probability that neither occur is 0.36, determine $P(A)$ and $P(B)$.

Solution: From the problem statement, $P(AB) = P(A)P(B) = 0.16$ and $P\{(A \cup B)^c\} = 0.36$ from which it follows that

$$P(A \cup B) = 0.64 = P(A) + P(B) - 0.16$$

which gives two equations in two unknowns:

$$(i) \;\; P(A) + P(B) = 0.80 \qquad (ii) \;\; P(A)P(B) = 0.16$$

which are easily solved to give $P(A) = 0.40$ and $P(B) = 0.40$.

29. Of three coins, 2 are two-headed and the other is fair. From these three coins, one is chosen at random and tossed n times. If the results is all heads, what is the probability the coin tossed is two-headed?

Solution: Let A be the event the coin is fair, B the event the coin is two-headed and C the event that n heads are obtained in n tosses of the coin. Then $P(C|B) = 1$, $P(C|A) = 1/2^n$, $P(A) = 1/3$ and $P(B) = 2/3$. By Bayes theorem:

$$P(B|C) = \frac{P(C|B)P(B)}{P(C|B)P(B) + P(C|A)P(A)} = \frac{1 \cdot \frac{2}{3}}{1 \cdot \frac{2}{3} + \frac{1}{2^n} \cdot \frac{1}{3}} = \frac{2^{n+1}}{2^{n+1} + 1} \ .$$

30. A fair coin is thrown n times.

(a) Show that the conditional probability of a head on any specified trial, given a total of k heads over the n trials, is k/n.

(b) Suppose $0 \le r \le k$ and $0 < m < n$ are integers. Show that the conditional probability of r heads over the first m trials, given a total of k heads over all n trials, is

$$\binom{m}{r}\binom{n-m}{k-r}\Big/\binom{n}{k} .$$

Solution: (a) Let A_i be the event that a head is obtained on the ith trial and B_k the event of obtaining k heads in n trials. Then

$$P(A_iB_k) = \binom{n-1}{k-1}\frac{1}{2^{n-1}}\frac{1}{2} \quad \text{and} \quad P(B_k) = \binom{n}{k}\frac{1}{2^n}$$

Thus

$$P(A_i|B_k) = \frac{\binom{n-1}{k-1}\frac{1}{2^n}}{\binom{n}{k}\frac{1}{2^n}} = \frac{k}{n} .$$

(b) Let C_r be the event of obtaining r heads in the first m places. Then $P(B_kC_r)$ is the probability of k heads in n throws with r of the k among the first m. Then

$$P(C_r|B_k) = \frac{\binom{m}{r}\binom{n-m}{k-r}}{2^n}\Big/\frac{\binom{n}{k}}{2^n} = \frac{\binom{m}{r}\binom{n-m}{k-r}}{\binom{n}{k}} .$$

31. Stores A,B and C have 50, 75 and 100 employees of which 50%, 60% and 70% respectively, are women. Resignations are equally likely among all employees, regardless of gender. One employee resigns and this is a woman. What is the probability she worked in store C?

Solution: By straightforward application of Bayes theorem:

$$\begin{aligned} P(C|W) &= \frac{P(W|C)P(C)}{P(W|C)P(C) + P(W|A)P(A) + P(W|B)P(B)} \\ &= \frac{0.7 \times (100/225)}{0.7 \times (100/225) + 0.5 \times (50/225) + 0.6 \times (75/225)} \\ &= \frac{70}{70+25+45} = \frac{1}{2} . \end{aligned}$$

32. Two hunters shoot at a deer which is hit by exactly one bullet. If the first hunter hits his target with probability 0.3 and the second with probability 0.6, what is the probability the second hunter killed the deer? (The answer is not 2/3.)

Solution: Let H_2 be the event that hunter 2 shot the deer and H_2^c the event that hunter 2 did not shoot the deer. Let B be the event that only one hunter shot the deer. Then $P(B) = 0.3 \times 0.4 + 0.7 \times 0.6 = 0.12 + 0.42 = 0.54$ and the conditional probability

$$P(H_2|B) = \frac{0.42}{0.12 + 0.42} = \frac{42}{54} = \frac{7}{9} .$$

33. You are about to have an interview for the Harvard Law School. 60% of the interviewers are conservative and 40% are liberal. 50% of the conservatives smoke cigars but only 25% of the liberals do. Your interviewer lights up a cigar. What is the probability he is a liberal?

Solution: Let L = liberal and C = conservative and S = smoker. Then by Bayes' theorem:

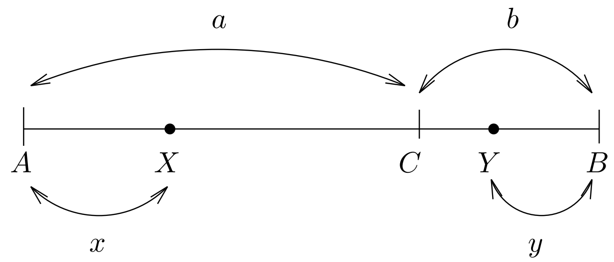
$$\begin{aligned} P(L|S) &= \frac{P(S|L)P(L)}{P(S|L)P(L)+P(S|C)P(C)} \\ &= \frac{0.25\times 0.4}{0.25\times 0.4+0.5\times 0.6} = 0.25\ . \end{aligned}$$

34. One slot machine pays off 1/2 of the time while another pays off 1/4 of the time. A player picks one of the machines and plays it six times, winning three of the times. What is the probability he is playing the machine that pays off 1/4 of the time?

Solution: Let S_1, S_2 be slot machine 1 and 2 respectively, and $W_{3,6}$ the event of winning 3 out of 6 with the machine being played. Then by Bayes theorem:

$$\begin{aligned} P(S_2|W_{3,6}) &= \frac{P(W_{3,6}|S_2)P(S_2)}{P(W_{3,6}|S_2)P(S_2)+P(W_{3,6}|S_1)P(S_1)} \\ &= \frac{\binom{6}{3}(1/4)^3(3/4)^3(1/2)}{\binom{6}{3}(1/4)^3(3/4)^3(1/2)+\binom{6}{3}(1/2)^6(1/2)} = \frac{27}{91}\ . \end{aligned}$$

35. Consider a line AB divided into two parts by a point C, where the length of segment AC is greater than or equal to the length of segment CB (see the following figure). Suppose a point X is chosen at random along the segment AC and a point Y is chosen at random along the segment CB. Let x and y denote the distances of X and Y from A and B, respectively. What is the probability the three segments AX, XY, and YB will fit together to form a triangle?



Solution: The key here is to recognize that three conditions must be met if the segments are to form a triangle - each segment must be shorter than the sum of the other two:

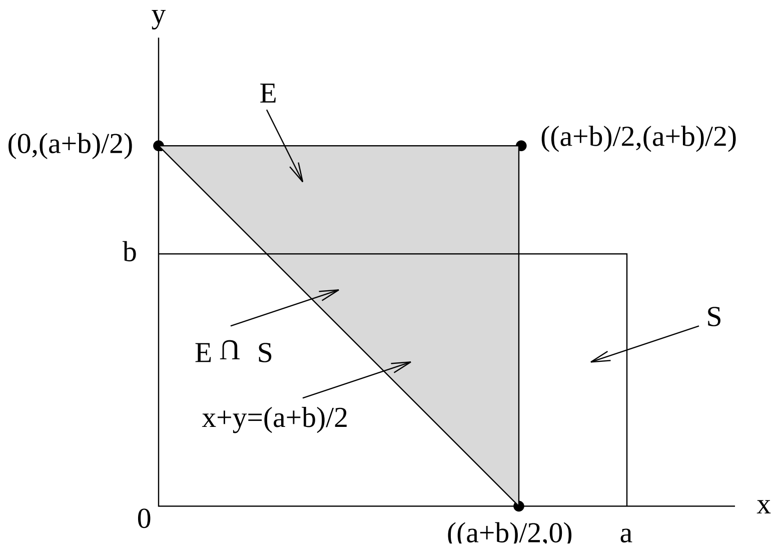
1. $x < (a+b-x-y)+y = a+b-x$
2. $a+b-x-y < x+y$
3. $y < x+(a+b-x-y)$

Intuitively, it seems clear that the probability of the segments forming a triangle will be greatest when C is the midpoint of AB: As b gets smaller, y tends to get smaller, and the likelihood of condition (2) being true diminishes.

To make the argument precise, we need to determine what proportion of the sample space $S = \{(x,y) : 0 \leq x \leq a, 0 \leq y \leq b\}$ is included in the (x,y)-values satisfying conditions (1),(2), and (3). Note that

$$\begin{aligned} &1. \quad x < (a+b-x-y)+y \;\Rightarrow\; x < \frac{a+b}{2} \\ &2. \quad a+b-x-y < x+y \;\Rightarrow\; x+y > \frac{a+b}{2} \\ &3. \quad y < x+(a+b-x-y) \;\Rightarrow\; y < \frac{a+b}{2} \end{aligned}$$

The (x,y)-values satisfying all three of these inequalities make up the interior of the triangle shown in the figure. Call that interior E.



It follows that the probability of the segments forming a triangle will equal the area of $E \cap S$ divided by the area of S:

$$\begin{aligned} P(\text{segments form triangle}) &= \frac{\frac{1}{2}b^2}{ab} \\ &= \frac{b}{2a} \end{aligned}$$

As expected, the probability is greatest when C is midway between A and B, and it decreases fairly rapidly as C approaches B.

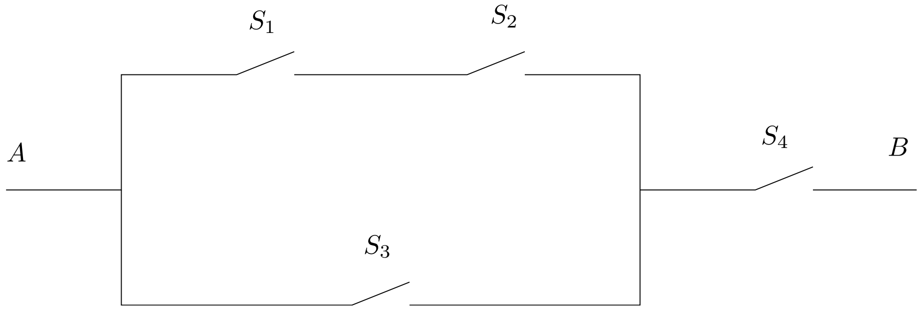
36. Consider the following connection of switches:

    Define the event: $E_i, i = 1,2,3,4$ as: Switch S_i is closed. Assume that $P(E_1) = P(E_2) = P(E_3) = P(E_4) = a$. Define the event $\mathcal{E}$ as the event that point A is connected to point B. Compute the probability of the event $\mathcal{E}$.

    Solution: Consider the connection of switches shown in the above figure. Define the events:

    E_1: Switch S_1 is closed. $\Longrightarrow$ E_1^c: Switch S_1 is open.

    E_2: Switch S_2 is closed. $\Longrightarrow$ E_2^c: Switch S_2 is open.



E_3: Switch S_3 is closed. $\Longrightarrow$ E_3^c: Switch S_3 is open.

E_4: Switch S_4 is closed. $\Longrightarrow$ E_4^c: Switch S_4 is open.

Assume that $P(E_1) = P(E_2) = P(E_3) = P(E_4) = a$, and consequently, $P(E_1^c) = P(E_2^c) = P(E_3^c) = P(E_4^c) = 1 - a$.

Define the event $\mathcal{E}$ as the event that point A is connected to point B. What is $P(\mathcal{E})$.

<u>First method:</u>

$\mathcal{E} = [(E_1E_2) \cup E_3]E_4$

$P(\mathcal{E}) = P[(E_1E_2) \cup E_3] \times P(E_4)$

$P[(E_1E_2) \cup E_3] = P(E_1E_2) + P(E_3) - P(E_1E_2E_3) = a + a^2 - a^3$

$P(\mathcal{E}) = a(a + a^2 - a^3) = a^2 + a^3 - a^4$

<u>Second method (fallacy):</u>

$P\{E_4[E_3 \cup (E_1E_2)]\} = P\{E_4[(E_3 \cup E_1)(E_3 \cup E_2)]\}$

$P(E_3 \cup E_1) = P(E_3) + P(E_1) - P(E_3E_1) = a + a - a^2 = 2a - a^2$

$P(E_3 \cup E_2) = P(E_3) + P(E_2) - P(E_3E_2) = a + a - a^2 = 2a - a^2$

$P\{E_4[(E_3 \cup E_1)(E_3 \cup E_2)]\} = P(E_4) \times P(E_3 \cup E_1) \times P(E_3 \cup E_2) = a(2a - a^2)^2$

It is seen that the two answers are different. Why?!

<u>Third method:</u>

$\mathcal{E}^c = \{[(E_1E_2) \cup E_3]E_4\}^c = E_4^c \cup \{[E_3 \cup (E_1E_2)]\}^c = E_4^c \cup [E_3^c(E_1^c \cup E_2^c)]$

$P(\mathcal{E}^c) = P\{E_4^c \cup [E_3^c(E_1^c \cup E_2^c)]\} = P(E_4^c) + P[E_3^c(E_1^c \cup E_2^c)] - P[E_4^cE_3^c(E_1^c \cup E_2^c)]$

$P(E_4^c) = 1 - a$

$P(E_1^c \cup E_2^c) = (1 - a) + (1 - a) - (1 - a)^2 = 1 - a^2$ (this is indeed equal to $P[(E_1E_2)^c]$).

$P[E_4^cE_3^c(E_1^c \cup E_2^c)] = (1 - a)^2(1 - a^2)$

Substituting results in

$P(\mathcal{E}^c) = P(E_4^c) + P[E_3^c(E_1^c \cup E_2^c)] - P[E_4^cE_3^c(E_1^c \cup E_2^c)] = (1 - a) + (1 - a)(1 - a^2) - (1 - a)^2(1 - a^2) = 1 - (a^2 + a^3 - a^4)$.

Which is in agreement with the previous result.

<u>Fourth method:</u>

Let 1 correspond to a closed switch and 0 correspond to an open switch. Using these notations, the sample space is represented as:

		S_4	S_3	S_2	S_1	
	e_1	0	0	0	0	
	e_2	0	0	0	1	
	e_3	0	0	1	0	
	e_4	0	0	1	1	
	e_5	0	1	0	0	
	e_6	0	1	0	1	
	e_7	0	1	1	0	
	e_8	0	1	1	1	
	e_9	1	0	0	0	
	e_{10}	1	0	0	1	
	e_{11}	1	0	1	0	
OK	e_{12}	1	0	1	1	$a^3(1-a)$
OK	e_{13}	1	1	0	0	$a^2(1-a)^2$
OK	e_{14}	1	1	0	1	$a^3(1-a)$
OK	e_{15}	1	1	1	0	$a^3(1-a)$
OK	e_{16}	1	1	1	1	a^4

The $e_1, e_2, \ldots, e_{16}$ are the elementary events.

$P(\mathcal{E}) = P(e_{12} \cup e_{13} \cup e_{14} \cup e_{15} \cup e_{16})$.

As the elementary events are disjoint, the probabilities add. As a result,

$P(\mathcal{E}) = P(e_{12})+P(e_{13})+P(e_{14})+P(e_{15})+P(e_{16}) = a^3(1-a)+a^2(1-a)^2+a^3(1-a)+a^3(1-a)+a^4 = a^2 + a^3 - a^4$.

Which is in agreement with the previous result.

4 Random Variables

4.1 Random Variables

It is frequently the case when an experiment is performed that we are mainly interested in some function of the outcome as opposed to the actual outcome itself. For instance, in tossing dice we are often interested in the sum of the two dice and are not really concerned about the separate values of each die. These quantities of interest, or more formally, these real-valued functions defined on the sample space, are known as random variables.

Random Variable: A *random variable* is a number which selects value in an unpredictable manner. In other words, a random variable is the result of a random experiment whose outcome is a number. A more formal definition of a random variable is as follows.

Let S be the sample space associated with some experiment E. A random variable X is a function that assigns a real number $X(s)$ to each elementary event $s \in S$. In this case, as the outcome of the experiment, i.e., the specific s, is not predetermined, then the value of $X(s)$ is not fixed. This means that the value of the random variable is determined by the specific outcome of the experiment.

The domain of a random variable X is the sample space S and the range space R_X is a subset of the real line, i.e., $R_X \subseteq \Re$.

There are two types of random variables: discrete and continuous.

It is conventional to use capital letters such as $X, Y, S, T, \ldots$, to denote a random variable and the corresponding lower case letter, $x, y, s, t, \ldots$, to denote particular values taken by the random variable.

4.2 Discrete Random Variables

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

Probability mass function: For a discrete random variable X, we define the *probability mass function* $p(a)$ of X by

$$p(a) = P\{X = a\}$$

Since X must take on one of the values x_i, we have

$$\sum_i p(x_i) = \sum_i p\{x = x_i\} = 1$$

Cumulative Distribution Function (CDF): For the random variable X, we define the function $F(x)$

by the equation,

$$F(x) = P(X \leq x).$$

And for any a

$$F(a) = \sum_{all\ \ x \leq a} p(x)$$

4.3 Expected Value

The expected value of the discrete random variable X, which we denote by $E(X)$, is given by

$$E(X) = \sum_{all\ \ x} xp(x)$$

4.4 Expectation of a Function of a Random Variable

A function of a discrete random variable X, say $Y = g(X)$, is also a discrete random variable. We have,

$$E(Y) = E[g(X)] = \sum_i g(x_i)p(x_i)$$

If a and b are constants, then

$$E[aX + b] = aE[X] + b$$

4.5 Variance

The *variance* of the random variable X, which we denote by $\text{Var}(X)$, is given by,

$$\begin{aligned} \text{Var}(X) &= E\{[X - \mu]^2\} \\ &= E[X^2] - \mu^2 \end{aligned}$$

where $\mu = E[X]$

For any two constants a and b

$$\text{Var}(aX + b) = a^2\text{Var}(X)$$

Standard deviation is defined as the square root of the variance. That is,

$$\text{SD}(X) = \sqrt{\text{Var}(X)}$$

Note: Unlike expected value of the sum which is always equal to the sum of the expected values, the variance of a sum is not necessarily equal to the sum of variances. A sufficient condition under which variance of the sum is equal to the sum of the variances, is that the random variables being summed are independent of each other. Note that variance of a constant is zero, and scaling a random variable by a constant a, multiples its variance by a^2.

4.6 The Bernoulli and Binomial Random Variables

Bernoulli random variable: The *Bernoulli* random variable corresponds to an event E which has two possible outcomes, *success* $(X = 1)$ and *failure* $(X = 0)$ and its probability mass function is

$$p(0) = P\{X = 0\} = 1 - p$$

$$p(1) = P\{X = 1\} = p$$

for $0 \leq p \leq 1$. Event E is also called a binary event, as it has two possible outcomes.

Binomial random variable: We repeat the Bernoulli experiment n times. The n trials are independent. Such an independent repetition of the same experiment under identical conditions is called a *Bernoulli Trial.* Let random variable X denotes the total number of times that a binary event E has occurred (success). We have,

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

The binomial coefficient enters here to account for the total number of possible ways that the i times of success can be located among the n times of performing the experiment. We let $0! = 1$ by definition.

4.6.1 Properties of Binomial Random Variables

The mean of the distribution is:

$$\begin{aligned}
E[X] &= \sum_{i=0}^{n} iP(i) = \sum_{i=1}^{n} i \binom{n}{i} p^i (1-p)^{n-i} \\
&= \sum_{i=1}^{n} i \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} \\
&= \sum_{i=1}^{n} \frac{n(n-1)!}{(i-1)!(n-i)!} p^i (1-p)^{n-i} \\
&= n \sum_{i=1}^{n} \frac{(n-1)!}{(i-1)!(n-i)!} p^i (1-p)^{n-i} \quad \text{substituting } j = i-1, \text{ we obtain,} \\
&= n \sum_{j=0}^{n-1} \frac{(n-1)!}{j!(n-1-j)!} p^{j+1} (1-p)^{n-1-j} \\
&= np \sum_{j=0}^{n-1} \frac{(n-1)!}{j!(n-1-j)!} p^{j} (1-p)^{n-1-j} \\
&= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} = np(p+1-p)^{n-1} = np(1)^{n-1} = np
\end{aligned}$$

The variance of the distribution can be computed similarly. The result is

$$\mathrm{Var}(X) = np(1-p)$$

If X is a binomial random variable with parameters (n, p), as k goes from 0 to n, $P\{X = k\}$ first increases monotonically and then decreases monotonically, reaching its largest value when k is the largest integer less than or equal to $(n+1)p$.

4.6.2 Computing the Binomial Distribution Function

Suppose that X is binomial with parameters (n, p). The key to computing its distribution function

$$P\{X \leq i\} = \sum_{k=0}^{i} \binom{n}{k} p^k (1-p)^{n-k} \qquad i = 0, 1, \ldots, n$$

is to utilize the following relationship between $P\{X = k+1\}$ and $P\{X = k\}$,

$$P\{X = k+1\} = \frac{p}{1-p} \frac{n-k}{k+1} P\{X = k\}$$

4.7 The Poisson Random Variable

A random variable X, taking one of the values $0, 1, 2, \ldots$ is said to be a *Poisson* random variable with parameter λ if for some $\lambda > 0$,

$$p(i) = P\{X = i\} = e^{-\lambda} \frac{\lambda^i}{i!}$$

The Poisson random variable is used as an approximation of a binomial random variable with parameter (n, p) when n is large and p is small so that np is in moderate size. The parameter of the approximated Poisson random variable is $\lambda = np$. That is,

$$\begin{aligned} P\{X = i\} &= \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i} \\ &= \frac{n!}{(n-i)!i!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\ &= \frac{n(n-1)\cdots(n-i+1)}{n^i} \frac{\lambda^i}{i!} \frac{(1-\lambda/n)^n}{(1-\lambda/n)^i} \end{aligned}$$

Now, for n large and λ moderate,

$$\left(1 - \frac{\lambda}{n}\right)^n \approx e^{-\lambda} \qquad \frac{n(n-1)\cdots(n-i+1)}{n^i} \approx 1 \qquad \left(1 - \frac{\lambda}{n}\right)^i \approx 1$$

Hence, for n large and λ moderate,

$$P\{X = i\} = e^{-\lambda} \frac{\lambda^i}{i!}$$

The mean of the distribution is:

$$\begin{aligned} E[X] &= \sum_{i=0}^{\infty} \frac{ie^{-\lambda}\lambda^i}{i!} \\ &= \lambda \sum_{i=1}^{\infty} \frac{ie^{-\lambda}\lambda^{i-1}}{(i-1)!} \\ &= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \quad \text{by substituting } j = i-1, \\ &= \lambda \qquad \text{since } \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = e^{\lambda} \end{aligned}$$

The variance of the distribution can be computed similarly. The result is

$$\text{Var}(X) = \lambda$$

The expected value and variance of a Poisson random variable are both equal to its parameter λ.

4.7.1 Computing the Poisson Distribution Function

If X is Poisson with parameter λ, then

$$\frac{P\{X = i+1\}}{P\{X = i\}} = \frac{e^{-\lambda}\lambda^{i+1}/(i+1)!}{e^{-\lambda}\lambda^i/(i)!} = \frac{\lambda}{(i+1)}$$

Starting with $P\{X = 0\} = e^{-\lambda}$, we can use the above equation to compute successively

$$\begin{aligned} P\{X = 1\} &= \lambda P\{X = 0\} \\ P\{X = 2\} &= \frac{\lambda}{2} P\{X = 1\} \\ &\vdots \\ P\{X = i+1\} &= \frac{\lambda}{i+1} P\{X = i\} \end{aligned}$$

4.8 Other Discrete Probability Distributions

Geometric Distribution: Suppose we repeat an experiment, independently, until the event E (success) occurs. What is the probability that the first success occurs on the nth try if, on any one try, we have $P(E) = p$?

$$P\{X = n\} = (1-p)^{n-1}p, \quad n = 1, 2, \ldots$$

This is referred to as the *Geometric Distribution.* We can show that,

$$E[X] = \frac{1}{p}, \qquad \text{Var}(X) = \frac{1-p}{p^2}$$

The negative binomial random variable: Suppose we repeat an experiment. Let X be the number of trials to obtain the first r success. We can show that,

$$P(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \quad n = r, r+1, \ldots$$

This is called the *negative binomial Distribution.* The binomial coefficient enters here to account for the total number of possible ways that the $r-1$ times of success can be located among the $n-1$ times of performing the experiment. Note that the last experiment has definitely resulted in a success, and consequently, we only consider the number of ways that the first $r-1$ success are located among the first $n-1$ trials. We have,

$$E[X] = \frac{r}{p}$$

and

$$\text{Var}(X) = r\frac{(1-p)}{p^2}$$

The hypergeometric random variable: Assume that N items are in an urn, numbered $1, 2, \ldots, N$. Suppose that items $1, 2, \ldots, m$ are white and the remaining $N-m$ items are black. Assume further that n items from the N are chosen at random. The probability that, in a draw of n items, we obtain exactly i white items is equal to:

$$P\{X = i\} = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}, \quad i = 0, 1, 2, \ldots, \min(n, m)$$

This is referred to as the *Hypergeometric Distribution.*
The mean and variance of the Hypergeometric Distribution are equal to:

$$E[X] = \frac{nm}{N}, \qquad \text{Var}(X) = \frac{N-n}{N-1} np(1-p), \text{where } p = m/N$$

4.9 Properties of Cumulative Distribution Function

Some Properties of the cumulative distribution function (c.d.f.) F are

1. F is a nondecreasing function; that is, if $a < b$, then $F(a) \leq F(b)$.

2. $\lim_{b\to\infty} F(b) = 1$.

3. $\lim_{b\to-\infty} F(b) = 0$.

4. F is right continuous. That is, for any b and any decreasing sequence $b_n, n \geq 1$, that converges to b, $\lim_{n\to\infty} F(b_n) = F(b)$.

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

4.10 Some Solved Problems

1. A large lot of items is known to contain a fraction θ defective. Let X denote the random variable for the number of items to be inspected to obtain the second defective item. Find the probability distribution and mean of X.

   Solution: If the second defective item is drawn on the ith draw, then the first defective can be in one of the first $(i-1)$ positions. The probability that $X = i$ is then

   $$P(X = i) = (i-1)(1-\theta)^{i-2}\theta^2 \ , \quad i = 2, 3, \cdots .$$

   To see that this is a probability distribution, note the following formulae, obtained by differentiating the first one two times:

   $$\sum_{i=0}^{\infty} x^i = \frac{1}{1-x} \ , \qquad \sum_{i=1}^{\infty} i x^{i-1} = \frac{1}{(1-x)^2} \ , \qquad \sum_{i=2}^{\infty} i(i-1)x^{i-2} = \frac{2}{(1-x)^3}$$

   It follows readily from these relationships that:

   $$\sum_{i=2}^{\infty}(i-1)\theta^2(1-\theta)^{i-2} = 1 \ \text{ and } \ E(X) = \sum_{i=2}^{\infty} iP(X=i) = \sum_{i=2}^{\infty} i(i-1)\theta^2(1-\theta)^{i-2} = \frac{2}{\theta}$$

2. Find the mean and variance of the Poisson random variable Y if it is three times as likely that $Y = 4$ than $Y = 2$.

   Solution:

   Since, by the problem statement,

   $$P(Y = 4) = e^{-\lambda}\frac{\lambda^4}{4!} = 3e^{-\lambda}\frac{\lambda^2}{2!} \quad \rightarrow \quad \lambda = 6$$

   Then $E(Y) = \text{Var}(Y) = \lambda = 6$ (property of the Poisson distribution).

3. The discrete random variable U has a geometric distribution of the form

   $$P(U = j) = a\alpha^j \ , \quad j = 0, 1, 2, \cdots$$

   If $P(U \geq 4) = 1/256$, find $P(U \geq 2)$.

   Solution: Since the given distribution must sum to unity we have $\sum_{j=0}^{\infty} a\alpha^j = 1 = a/(1-\alpha)$ which implies that $a = 1-\alpha$. Now

   $$P(U \geq 4) = \sum_{j=4}^{\infty} a\alpha^j = 1/256 = \frac{a\alpha^4}{1-\alpha} = \frac{1}{256} \quad \Rightarrow \quad \alpha = \frac{1}{4}$$

   Then $P(U \geq 2) = (1-\alpha)\alpha^2/(1-p) = 1/16$.

4. Z is a discrete random variable with probability distribution

$$P(Z = j) = aj2^{-(j-1)} \ , \quad j = 1, 2, \cdots$$

For what value of a is this a distribution?

Solution: Notice that for $|\ x\ |< 1$, $\sum_{j=0}^{\infty} x^j = 1/(1-x)$. Differentiating both sides of this relationship wrt x gives $\sum_{j=1}^{\infty} jx^{j-1} = 1/(1-x)^2$. Applying this to the problem gives:

$$\sum_{j=1}^{\infty} aj2^{-(j-1)} = \frac{a}{\left(1 - \frac{1}{2}\right)^2} = 1$$

which gives $4a = 1$ or $a = 1/4$.

5. The number of trials, X, to obtain the first defective from a large lot of items, has a geometric distribution with variance 2. What is the probability it takes more than 4 draws to obtain the first defective?

Solution: The mean of a geometric distribution is $1/\theta$ and the variance is $(1-\theta)/\theta^2$. If $(1-\theta)/\theta^2 = 2$ then $\theta = 1/2$. The probability $X > 4$ is then

$$P(X > 4) = \sum_{i=5}^{\infty} \theta^{i-1}(1-\theta) = \frac{\theta^4(1-\theta)}{1-\theta} = \theta^4 = \frac{1}{16} \ .$$

6. The number of emissions from a radioactive source, Z, in a one hour period, is known to be a Poisson random variable. If it is known that the probability there are 10 emissions in the one hour period, is exactly the same as the probability there are 12 emissions, what is the probability there are no emissions in the one hour period?

Solution: If $P(X = 10) = \exp(-\lambda)\lambda^{10}/10! = \exp(-\lambda)\lambda^{12}/12!$ then $\lambda^2 = 132$. it follows that

$$P(X = 0) = e^{-\lambda} = e^{-\sqrt{132}} \ .$$

7. In Lotto 649, a player chooses 6 distinct numbers from 1 to 49. This set is compared to 6 other numbers chosen at random by the house and prizes are awarded according to the number matched.

(a) What is the probability that k of the numbers match, $k = 3, 4, 5, 6$. Calculate the answer to 8 significant digits.

(b) What is the probability that all six numbers chosen by the house are from 1 to 9?

(c) What is the probability that 2 of the numbers chosen by the house are from 1 to 9 and one each is from 10 to 19, 20 to 29, 30 to 39 and 40 to 49 respectively?

Solution: (a) Let X be the number of the player's 6 numbers that match the house's 6 numbers. Then X has a hypergeometric distribution and

$$P(X = k) = \frac{\binom{6}{k}\binom{43}{6-k}}{\binom{49}{6}}$$

which is evaluated as:

$$\begin{array}{llll} P(X=3)= & .017650404 & P(X=4)= & .00096861972 \\ P(X=5)= & .000018449899 & P(X=6)= & 7.1511238\times 10^{-8} \end{array}$$

(b) The probability all six numbers chosen by the house are from 1 to 9 is:

$$\frac{\binom{9}{6}}{\binom{49}{6}} = 6.00694403 \times 10^{-6} \ .$$

(c) The required probability is:

$$\frac{\binom{9}{2}\binom{10}{1}^4}{\binom{49}{6}} = 0.025744046 \ .$$

8. In the game of Keno, a player chooses 5 distinct numbers from 1 to 100, inclusive. A machine then chooses 10 numbers, again from 1 to 100. Let X be the number of the player's numbers chosen by the machine. Find the probability distribution of X.

   Solution: This is a simple hypergeometric distribution:

$$P(X=k) = \frac{\binom{10}{k}\binom{90}{5-k}}{\binom{100}{5}} \ .$$

9. Consider a random experiment in which the probability of certain outcome, say A, is equal to p. The experiment is performed for n times. Let X denote the total number of times that A has happened. Compute the average and the standard deviation of X.

   Solution: Consider an experiment having only two possible outcomes, A and $\bar{A}$, which are mutually exclusive. Let the probabilities be $P(A) = p$ and $P(\bar{A}) = 1 - p = q$. The experiment is repeated n times and the probability of A occurring i times is

$$P_i = \binom{n}{i} p^i q^{n-i}$$

where $\binom{n}{i}$ is the binomial coefficient,

$$\binom{n}{i} = \frac{n!}{i!(n-i)!}$$

The binomial coefficient enters here to account for the total number of possible ways to combine n items taken i at a time, with all possible permutations permitted. We let $0! = 1$ by definition.

This is called a Binomial Random Variable. It is straightforward to show that the mean and the variance of a Binomial Random Variable are equal to,

$$\text{mean} = np \qquad \& \qquad \sigma^2 = np(1-p) = m(1-p)$$

To compute the mean, we have,

$$\begin{aligned}
\text{mean} &= \sum_{i=0}^{n} iP_i = \sum_{i=1}^{n} i\binom{n}{i}p^i q^{n-i} \\
&= \sum_{i=1}^{n} i\frac{n!}{i!(n-i)!}p^i q^{n-i} \\
&= \sum_{i=1}^{n} \frac{n(n-1)!}{(i-1)!(n-i)!}p^i q^{n-i} \\
&= n\sum_{i=1}^{n} \frac{(n-1)!}{(i-1)!(n-i)!}p^i q^{n-i} \quad \text{substituting } j=i-1, \text{ we obtain,} \\
&= n\sum_{j=0}^{n-1} \frac{(n-1)!}{j!(n-1-j)!}p^{j+1} q^{n-1-j} \\
&= np\sum_{j=0}^{n-1} \frac{(n-1)!}{j!(n-1-j)!}p^{j} q^{n-1-j} \\
&= np\sum_{j=0}^{n-1} \binom{n-1}{j}p^{j} q^{n-1-j} = np(p+q)^{n-1} = np(1)^{n-1} = np
\end{aligned}$$

The standard deviation can be computed similarly.

10. A person repeatedly shoots at a target and stops as soon as he hits it. The probability of hitting the target is 2/3 each time. The shots are fired independently, and hence may be regarded as a sequence of repeated trials. Determine the probability that the target is hit on the kth attempt.

    Solution: The event in which we are interested happens if the target is missed $k-1$ times and then is hit the kth time. The probability of missing the target with the first shot is $1-2/3=1/3$, and likewise in each of the following shots. Because of the independence, the probability we seek is $(1/3)^{k-1}(2/3)$.

11. Russian roulette is played with a revolver equipped with a rotatable magazine of six shots. The revolver is loaded with one shot. The first duelist, A, rotates the magazine at random, points the revolver at his head and presses the trigger. If, afterwards, he is still alive, he hands the revolver to the other duelist, B, who acts in the same way as A. The players shoot alternately in this manner, until a shot goes off. Determine the probability that A is killed.

    Solution: Let H_i be the event that "a shot goes off at the ith trial". The events H_i are mutually exclusive. The event H_i occurs if there are $i-1$ "failures" and then one "success". Hence, we get,

$$P(H_i) = \left(\frac{5}{6}\right)^{i-1}\frac{1}{6}$$

    The probability we want is given by

$$P \;=\; P(H_1 \cup H_3 \cup H_5 \cup \cdots) = \frac{1}{6}\left[1 + \left(\frac{5}{6}\right)^2 + \left(\frac{5}{6}\right)^4 + \cdots\right]$$

$$= \frac{1}{6} \cdot \frac{1}{1-\left(\frac{5}{6}\right)^2} = \frac{6}{11}$$

Hence the probability that B loses his life is $1 - 6/11 = 5/11$; that is, the second player has a somewhat greater chance of surviving, as might be expected.

12. In a food package there is a small prize which can be of N different types. All types are distributed into the packages at random with the same probability. A person buys n packages in a shop. What is the probability that he obtains a complete collection of presents?

    Solution: Denote the event in question by H and consider its complement $\bar{H}$, which is the event that at least one type is missing. We then have

    $$P(\bar{H}) = P\left(\bigcup_1^N A_i\right),$$

    where A_i="type no. i is missing". We now utilize the same addition formula for the union of some events. It is evident that

    $$P(A_{i_1}A_{i_2}...A_{i_r}) = \left(1 - \frac{r}{N}\right)^n$$

    for the probability that a given package does not contain types nos. $i_1, i_2, ..., i_r$ is $1 - r/N$. Hence we find

    $$1 - P(H) = P(H^c) = N\left(1 - \frac{1}{N}\right)^n - \binom{N}{2}\left(1 - \frac{2}{N}\right)^n + \cdots$$
    $$+(-1)^{N-1}\binom{N}{N}\left(1 - \frac{N}{N}\right)^n$$

13. Suppose we roll a die repeatedly until a 6 occurs, and let N be the number of times we roll the die. Compute the average value and the variance of N.

    Solution: Using $\times$ to denote "not a 6", we have

    $$P(N=1) = P(6) = \frac{1}{6}$$
    $$P(N=2) = P(\times\ 6) = \frac{5}{6} \cdot \frac{1}{6}$$
    $$P(N=3) = P(\times\ \times\ 6) = \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6}$$

    From the first three terms it is easy to see that for $k \geq 1$

    $$P(N=k) = P(\times\text{on k-1 rolls then 6}) = \left(\frac{5}{6}\right)^{k-1}\frac{1}{6}$$

    Generalizing, we see that if we are waiting for an event of probability p, the number of trials needed, N, has the following distribution,

    $$P(N=k) = (1-p)^{k-1}p \text{ for } k = 1, 2, \ldots$$

since $\{N = k\}$ occurs exactly when we have $k - 1$ failures followed by a success. This is called a *Geometric Distribution* and as we will see in the lecture, it has,

$$E(N) = 1/p \quad \& \quad Var(N) = (1-p)/p^2$$

These values can be easily obtained using the second and the third formulas given in your sheet of "Some useful relationships".

14. Suppose that an airplane engine will fail, when in flight, with probability $1 - p$ independently from engine to engine. Suppose that the airplane will make a successful flight if at least 50 per cent of its engines remain operative. For what values of p is a 4-engine plane preferable to a 2-engine plane?

Solution: As each engine is assumed to fail or function independently of what happens with the other engines, it follows that the number of engines remaining operative is a binomial random variable. Hence the probability that a 4-engine plane makes a successful flight is

$$\binom{4}{2}p^2(1-p)^2 + \binom{4}{3}p^3(1-p) + \binom{4}{4}p^4(1-p)^0 = 6p^2(1-p)^2 + 4p^3(1-p) + p^4$$

whereas the corresponding probability for a 2-engine plane is

$$\binom{2}{1}p(1-p) + \binom{2}{2}p^2 = 2p(1-p) + p^2$$

Hence the 4-engine plane is safer if

$$6p(1-p)^2 + 4p^2(1-p) + p^3 \geq 2 - p$$

which simplifies to

$$3p^3 - 8p^2 + 7p - 2 \geq 0 \quad \text{or} \quad (p-1)^2(3p-2) \geq 0$$

which is equivalent to

$$3p - 2 \geq 0 \quad \text{or} \quad p \geq \frac{2}{3}$$

Hence the 4-engine plane is safer when the engine success probability is at least as large as $\frac{2}{3}$, whereas the 2-engine plane is safer when this probability falls below $\frac{2}{3}$.

15. Ten hunters are waiting for ducks to fly by. When a flock of ducks flies overhead, the hunters fire at the same time, but each chooses his target at random, independently of the others. If each hunter independently hits his target with probability p, compute the expected number of ducks that escape unhurt when a flock of size 10 flies overhead.

Solution: Let X_i equal 1 if the ith duck escapes unhurt and 0 otherwise, $i = 1, 2, \ldots, 10$. The expected number of ducks to escape can be expressed as

$$E[X_1 + \cdots + X_{10}] = E[X_1] + \cdots + E[X_{10}]$$

To compute $E[X_i] = P(X_i = 1)$, we note that each of the hunters will, independently, hit the ith duck with probability $p/10$ and so

$$P(X_i = 1) = \left(1 - \frac{p}{10}\right)^{10}$$

Hence

$$E[X] = 10\left(1 - \frac{p}{10}\right)^{10}$$

16. A fair coin is to be tossed until a head comes up for the first time. What is the chance of that happening on an odd-numbered toss?

Solution: Suppose we let $P(k)$ denote the probability that the first head appears on the kth toss. Since the coin was presumed to be fair, $P(1) = \frac{1}{2}$. Furthermore, we would expect half of the coins that showed a tail on the first toss to come up heads on the second, so, intuitively, $P(2) = \frac{1}{4}$. In general, $P(k) = \left(\frac{1}{2}\right)^k$, $k = 1, 2, \ldots$.
Let E be the event "first head appears on an odd-numbered toss." Then

$$\begin{aligned} P(E) &= P(1) + P(3) + P(5) + \cdots \\ &= \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^{2i+1} \\ &= \frac{1}{2}\sum_{i=0}^{\infty} \left(\frac{1}{4}\right)^{i} \end{aligned}$$

Recall the formula for the sum of a geometric series: If $0 < x < 1$,

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$$

Applying that result to $P(E)$ gives

$$\begin{aligned} P(E) &= \frac{1}{2} \cdot \left(\frac{1}{1 - \frac{1}{4}}\right) \\ &= \frac{2}{3} \end{aligned}$$

A similar computation would show that the probability of the first head appearing on an even-numbered toss is $\frac{1}{3}$.

17. Two gamblers, A and B, each choose an integer from 1 to m (inclusive) at random. What is the probability that the two numbers they pick do not differ by more than n?

Solution: It will be easier if we approach this problem via its complement. Let x and y denote the numbers selected by A and B, respectively. The complement has two cases, depending on whether $x < y$ or $x > y$. Let us first suppose that $x < y$. Then, for a given x, the values of y such that

$y-x>n$ are $y=x+n+1$, $y=x+n+2, \ldots$, and $y=m$, altogether, a range of $m-n-x$ choices. Summing over x, we find that the total number of (x,y)-pairs such that $y-x>n$ reduces to the sum of the first $m-n-1$ integers:

$$\sum_{x=1}^{m-n-1}(m-n-x)=\sum_{i=1}^{m-n-1} i=\frac{(m-n-1)(m-n)}{2}$$

By symmetry, the same number of (x,y)-pairs satisfies the second case: $x>y$ and $x-y>n$. Thus the total number of (x,y)-selections such that $|x-y|>n$ is $(m-n-1)(m-n)$.
The sample space S contains m^2 outcomes, all equally likely by assumption. It follows that

$$P(|x-y|\leq n)=1-\frac{(m-n-1)(m-n)}{m^2}$$

18. A secretary is upset about having to stuff envelopes. Handed a box of n letters and n envelopes, he vents his frustration by putting the letters into the envelopes *at random*. How many people, on the average, will receive the correct mail?

    Solution: If X denotes the number of envelopes properly stuffed, what we are seeking is $E(X)$. Let X_i denote a random variable equal to the number of correct letters put into the ith envelope, $i=1,2,\ldots,n$. Then X_i equals 0 or 1, and

    $$f_{X_i}(k)=P(X_i=k)=\begin{cases}\frac{1}{n} & \text{for } k\text{=1}\\ \frac{n-1}{n} & \text{for } k\text{=0}\end{cases}$$

    But $X=X_1+X_2+\cdots+X_n$ and $E(X)=E(X_1)+E(X_2)+\cdots+E(X_n)$.
    Furthermore, each of the X_i's has the same expected value, $1/n$:

    $$E(X_i)=\sum_{k=0}^{1} k\cdot P(X_i=k)=0\cdot\frac{n-1}{n}+1\cdot\frac{1}{n}=\frac{1}{n}$$

    It follows that

    $$E(X)=\sum_{i=1}^{n}E(X_i)=n\left(\frac{1}{n}\right)=1$$

    showing that, *regardless of* n, the expected number of properly stuffed envelopes is 1.

19. The honor count in a (13-card) bridge hand can vary from 0 to 37 according to the formula

    $$\begin{aligned}\text{Honor count} &= 4\cdot\text{number of aces}+3\cdot\text{number of kings}\\ &\quad +2\cdot\text{number of queens}+1\cdot\text{number of jacks}\end{aligned}$$

    What is the expected honor count of North's hand?

    Solution: If X_i, i=1,2,3,4, denotes the honor count for North, South, East, and West, respectively, and if X denotes the analogous sum for the entire deck, we can write

    $$X=X_1+X_2+X_3+X_4$$

    But

    $$X=E(X)=4\cdot 4+3\cdot 4+2\cdot 4+1\cdot 4=40$$

    By symmetry, $E(X_i)=E(X_j)$, $i\neq j$, so it follows that $40=4\cdot E(X_1)$, which implies that 10 is the expected honor count of North's hand.

20. Let $X_1, X_2, \ldots, X_n$ denote a set of n independent observations made on a random variable X having pdf $f_X(x)$. Let $\sigma^2 = E[(X-\mu)^2]$ denote the variance of X. The *sample variance* of the X_i's, denoted by S^2, is defined as

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

where $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Show that $E(S^2) = \sigma^2$.

Solution: We know that if X is a random variable having mean μ and $E(X^2)$ finite, then

$$Var(X) = E(X^2) - \mu^2$$

Hence, we rewrite S^2 in a form that enables us to apply the above equation:

$$\begin{aligned} E(S^2) &= E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\right] \\ &= E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i^2 - 2X_i\overline{X} + \overline{X}^2)\right] \\ &= E\left[\frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - 2\overline{X}\sum_{i=1}^{n} X_i + n\overline{X}^2\right)\right] \\ &= \frac{1}{n-1}\left[\sum_{i=1}^{n} E(X_i^2) - 2E\left(\overline{X}\sum_{i=1}^{n} X_i\right) + nE(\overline{X}^2)\right] \end{aligned}$$

But, we have:

$$E(X_i^2) = \sigma^2 + \mu^2$$

and,

$$E\left(\overline{X}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} E(X_i X_j) = \frac{1}{n}\left[(n^2-n)\mu^2 + n(\sigma^2+\mu^2)\right] = n\mu^2 + \sigma^2$$

and,

$$E(\overline{X}^2) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} E(X_i X_j) = \frac{1}{n^2}\left[(n^2-n)\mu^2 + n(\sigma^2+\mu^2)\right] = \frac{1}{n}(n\mu^2 + \sigma^2) = \mu^2 + \frac{\sigma^2}{n}$$

Therefore:

$$\begin{aligned} E(S^2) &= \frac{1}{n-1}\left[n(\sigma^2+\mu^2) - 2n\mu^2 - 2\sigma^2 + n\mu^2 + \sigma^2\right] \\ &= \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2 \end{aligned}$$

21. A chip is to be drawn at random from each of k urns, each holding n chips numbered 1 through n. What is the probability all k chips will bear the same number?

Solution: If $X_1, X_2, \ldots, X_k$ denote the numbers on the 1st, 2nd,...,kth chips, respectively, we have,

$$P(X_i = \alpha) = \frac{1}{n}, \quad \forall X_i \in \{X_1, \ldots, X_k\}, \forall \alpha \in [1, n]$$

We are looking for the probability that $X_1 = X_2 = \cdots = X_k$. Considering the independence of the events, we obtain,

$$P(X_1 = X_2 = \cdots = X_k = \alpha) = \left(\frac{1}{n}\right)^k, \quad \forall \alpha \in [1, n]$$

Adding up these values for different values of α, we obtain,

$$P(X_1 = X_2 = \cdots = X_k) = \frac{1}{n^{k-1}}$$

22. **Mean of a Hypergeometric Random Variable:** If n balls are randomly selected from an urn containing N white and M black balls, find the expected number of white balls selected.

    Solution: Let X denote the number of white balls selected. It follows that

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

    Hence, assuming $n \leq N$:

    $$E[X] = \frac{\sum_{k=0}^{n} k \binom{N}{k}\binom{M}{n-k}}{\binom{N+M}{n}}$$

    However, we can obtain a simpler expression for $E[X]$ by using the representation

    $$X = X_1 + \cdots + X_n$$

    where

    $$X_i = \begin{cases} 1 & \text{if the } i\text{th ball selected is white} \\ 0 & \text{otherwise} \end{cases}$$

    Since the ith ball selected is equally likely to be any of the $N + M$, we have

    $$E[X_i] = \frac{N}{M + N}$$

    and thus

    $$E[X] = E[X_1] + \cdots + E[X_n] = \frac{nN}{M + N}$$

23. Six balls are tossed independently into three boxes A, B, C. For each ball the probability of going into a specific box is 1/3. Find the probability that box A will contain (a) exactly four balls, (b) at least two balls, (c) at least five balls.

    Solution: Here we have six Bernoulli trials, with success corresponding to a ball in box A, failure to a ball in box B or C. Recall, a sequence of n Bernoulli trials is a sequence of n independent observations each of which may result in exactly one of two possible situations, called "success" or "failure". At each observation the probability of success is p and the probability of failure is $q = 1 - p$.

Thus, referring the our problem we have that n =6, p =1/3, q =2/3, and so the required probabilities are

$$(a) \qquad p(4) = \binom{6}{4}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^2$$

$$(b) \quad 1 - p(0) - p(1) = 1 - \left(\frac{2}{3}\right)^6 - \binom{6}{1}\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^5$$

$$(c) \qquad p(5) + p(6) = \binom{6}{5}\left(\frac{1}{3}\right)^5\left(\frac{2}{3}\right) + \left(\frac{1}{3}\right)^6$$

24. A bubble gum company is printing a special series of baseball cards featuring r of the greatest base-stealers of the past decade. Each player appears on the same number of cards, and the cards are randomly distributed to retail outlets. If a collector buys n ($\geq r$) packets of gum (each containing one card), what is the probability she gets a complete set of the special series?

    Solution: Let A_i be the event the collector has *no* card for player i, $i = 1, 2, \ldots, r$, and define A to be the union, $A = \cup_{i=1}^{r} A_i$. Then

    $$P(\text{collector has at least one card of each player}) = 1 - P(A)$$

    To begin the derivation of $P(A)$, notice that for any value of k, $k = 1, 2, \ldots, r$,

    $$P(A_1 A_2 \cdots A_k) = \left(1 - \frac{k}{r}\right)^n$$

    Therefore, from the general Addition Law,

    $$\begin{aligned} P\left(\bigcup_{i=1}^{r} A_i\right) &= P(A) \\ &= \sum_{i=1}^{r}\left(1-\frac{1}{r}\right)^n - \sum_{i<j}\left(1-\frac{2}{r}\right)^n + \sum_{i<j<k}\left(1-\frac{3}{r}\right)^n - \cdots + (-1)^{r+1}\cdot 0 \\ &= \binom{r}{1}\left(1-\frac{1}{r}\right)^n - \binom{r}{2}\left(1-\frac{2}{r}\right)^n + \binom{r}{3}\left(1-\frac{3}{r}\right)^n \\ &\quad - \cdots + (-1)^{r+1}\cdot 0 \end{aligned}$$

    Or more concisely,

    $$P(A) = \sum_{k=1}^{r} (-1)^{k+1}\binom{r}{k}\left(1 - \frac{k}{r}\right)^n$$

25. An urn contains nine chips, five red and four white. Three are drawn out at random without replacement. Let X denote the number of red chips in the sample. Find $E(X)$, the expected number of red chips selected.

    Solution: We recognize X to be a Hypergeometric random variable, where

    $$P(X = x) = \frac{\binom{5}{x}\binom{4}{3-x}}{\binom{9}{3}}, \quad x = 0, 1, 2, 3$$

Hence,

$$\begin{aligned} E(X) &= \sum_{\text{all x}} x \cdot P(X=x) \\ &= \sum_{i=0}^{3} x \cdot \frac{\binom{5}{x}\binom{4}{3-x}}{\binom{9}{3}} \\ &= (0)\left(\frac{4}{84}\right) + (1)\left(\frac{30}{84}\right) + (2)\left(\frac{40}{84}\right) + (3)\left(\frac{10}{84}\right) \\ &= \frac{5}{3} \end{aligned}$$

26. *The following problem was posed and solved in the eighteenth century by Daniel Bernoulli.* Suppose that a jar contains $2N$ cards, two of them marked 1, two marked 2, two marked 3, and so on. Draw out m cards at random. What is the expected number of pairs that still remain in the jar?

Solution: Define, $i = 1, 2, \ldots, N$

$$X_i = \begin{cases} 1 & \text{if the } i\text{th pair remains in the jar} \\ 0 & \text{otherwise} \end{cases}$$

Now,

$$\begin{aligned} E[X_i] &= P(X_i = 1) \\ &= \frac{\binom{2N-2}{m}}{\binom{2N}{m}} \\ &= \frac{\dfrac{(2N-2)!}{m!(2N-2-m)!}}{\dfrac{(2N)!}{m!(2N-m)!}} = \frac{(2N-m)(2N-m-1)}{(2N)(2N-1)} \end{aligned}$$

Hence the desired result is

$$\begin{aligned} E[X_1 + X_2 + \cdots + X_N] &= E[X_1] + \cdots + E[X_N] \\ &= \frac{(2N-m)(2N-m-1)}{2(2N-1)} \end{aligned}$$

27. Consider a jury trial in which it takes 8 of the 12 jurors to convict; that is, in order for the defendant to be convicted, at least 8 of the jurors must vote him guilty. If we assume that jurors act independently and each makes the right decision with probability θ , what is the probability that the jury renders a correct decision?

Solution: The problem, as stated, is incapable of solution, for there is not yet enough information. For instance, if the defendant is innocent, the probability of the jury's rendering a correct decision is

$$\sum_{i=5}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i}$$

whereas, if he is guilty, the probability of a correct decision is

$$\sum_{i=8}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i}$$

Therefore, if α represents the probability that the defendant is guilty, then, by conditioning on whether or not he is guilty, we obtain that the probability that the jury renders a correct decision is

$$\alpha \sum_{i=8}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i} + (1-\alpha) \sum_{i=5}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i}$$

28. A single unbiased die is tossed independently n times. Let R_1 be the number of 1's obtained, and R_2 the number of 2's. Find $E(R_1 R_2)$.

    Solution: The *indicator* of an event A is a random variable I_A and is defined as follows.

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

    If A_i is the event that the ith toss results in a 1, and B_i the event that the ith toss results in a 2, then

    $$\begin{aligned} R_1 &= I_{A_1} + \cdots + I_{A_n} \\ R_2 &= I_{B_1} + \cdots + I_{B_n} \end{aligned}$$

    Hence

    $$E(R_1 R_2) = \sum_{i,j=1}^{n} E(I_{A_i} I_{B_j})$$

    Now if $i \neq j$, I_{A_i} and I_{B_j} are independent; hence

    $$E(I_{A_i} I_{B_j}) = E(I_{A_i}) E(I_{B_j}) = P(A_i) P(B_j) = \frac{1}{36}$$

    If $i = j$, A_i and B_j are disjoint, since the ith toss cannot simultaneously result in a 1 and a 2. Thus $I_{A_i} I_{B_i} = I_{A_i \cap B_i} = 0$. Thus

    $$E(R_1 R_2) = \frac{n(n-1)}{36}$$

    since there are $n(n-1)$ ordered pairs (i,j) of integers $\in \{1, 2, \ldots, n\}$ such that $i \neq j$.

29. A miner is trapped in a mine containing 3 doors. The first door leads to a tunnel that will take him to safety after 3 hours of travel. The second door leads to a tunnel that will return him to the mine after 5 hours of travel. The third door leads to a tunnel that will return him to the mine after 7 hours. If we assume that the miner is at all times equally likely to choose any one of the doors, what is the expected length of time until he reaches safety?

    Solution: Let X denote the amount of time (in hours) until the miner reaches safety, and let Y denote the door he initially chooses. Now

    $$\begin{aligned} E[X] &= E[X|Y=1]P(Y=1) + E[X|Y=2]P(Y=2) \\ &\quad + E[X|Y=3]P(Y=3) \\ &= \frac{1}{3}(E[X|Y=1] + E[X|Y=2] + E[X|Y=3]) \end{aligned}$$

However,

$$\begin{aligned} E[X|Y=1] &= 3 \\ E[X|Y=2] &= 5+E[X] \\ E[X|Y=3] &= 7+E[X] \end{aligned}$$

To understand why the above equation is correct, consider, for instance, $E[X|Y=2]$ and reason as follows. If the miner chooses the second door, he spends 5 hours in the tunnel and then returns to his cell. But once he returns to his cell the problem is as before; thus his expected additional time until safety is just E[X]. Hence $E[X|Y=2]=5+E[X]$. The argument behind the other equalities in the above equation is similar. Hence

$$E[X]=\frac{1}{3}(3+5+E[X]+7+E[X])$$

or

$$E[X]=15$$

5 Continuous Random Variables

5.1 Introduction

In Chapter 4 we considered discrete random variables, that is, random variables whose set of possible values is either finite or countably infinite. However, there also exist random variables whose set of possible values is uncountable. Two examples would be the time that a train arrives at a specified stop and the lifetime of a transistor.

Definition: We say that X is a *continuous* random variable if there exists a nonnegative function f, defined for all real $x \in (-\infty, \infty)$, having the property that for any set B of real numbers

$$P\{X \in B\} = \int_B f(x)dx$$

The function is called the *probability density function (pdf)* of the random variable X.
Since X must assume some value, f must satisfy

$$P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x)dx = 1$$

Relation of CDF and pdf:

$$P\{X < a\} = P\{X \leq a\} = F(a) = \int_{-\infty}^{a} f(x)dx$$

5.2 Expectation and Variance of Continuous Random Variables

Expected Value: The expected value of the continuous random variable X, which we denote by $E(X)$, is given by

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

Expectation of a Function of a Continuous Random Variable: If X is a continuous random variable with probability density function $f(x)$, then for any real-valued function g,

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

Property: For a nonnegative random variable Y,

$$E(Y) = \int_{0}^{\infty} P\{Y > y\} dy$$

Property: If a and b are constants, then

$$E[aX + b] = aE[X] + b$$

Variance of Continuous Random Variables: The Variance of the continuous random variable X, which we denote by $V(X)$, is given by

$$V(X) = E[(X - \mu)^2] = E(X^2) - \mu^2$$

Where

$$\mu = E[X]$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx$$

5.3 The Uniform Random Variable

Uniform Random Variable: We sat that X is a *uniform random variable* on the interval (α, β) if its probability density function is given by

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \text{if } \alpha < x < \beta \\ 0 & \text{otherwise} \end{cases}$$

Since $F(a) = \int_{-\infty}^{a} f(x)dx$,

$$F(a) = \begin{cases} 0 & a \leq \alpha \\ \dfrac{a - \alpha}{\beta - \alpha} & \text{if } \alpha < a < \beta \\ 1 & a \geq \beta \end{cases}$$

For a uniformly distributed random variable X over (α, β)

$$E[X] = \frac{\alpha + \beta}{2}$$

and

$$\text{Var}(X) = \frac{(\beta - \alpha)^2}{12}$$

5.4 Normal Random Variables

Normal Random Variable: We say that X is a *normal random variable*, or simply that X is *normally distributed*, with parameters μ and σ^2 if the density of X is given by

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

For a normally distributed random variable X with parameters μ and σ^2,

$$E[X] = \mu$$

and

$$\mathrm{Var}(X) = \sigma^2$$

If X has the pdf $N(\mu, \sigma^2)$, then $Y = aX + b$ has the pdf

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}\left[\left(\frac{y-b}{a}\right) - \mu\right]^2\right\} \cdot \frac{1}{|a|}, \quad -\infty < y < \infty \\ &= \frac{1}{\sqrt{2\pi}|a|\sigma} \cdot \exp\left[-\frac{(y - a\mu - b)^2}{2a^2\sigma^2}\right], \quad -\infty < y < \infty \end{aligned}$$

which is the pdf of $N(a\mu + b, a^2\sigma^2)$. Note that a linear transformation of a normal random variable results in a normal random variable.

Standard Normal Random Variable: Standard normal random variable is a normal random variable with parameters $(0, 1)$. That is,

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

It is traditional to denote the cumulative distribution function of a standard normal random variable by $\Phi(x)$. That is,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy$$

For any normal random variable X with parameters (μ, σ^2), $Z = \dfrac{X-\mu}{\sigma}$ is a normal random variable and its cumulative distribution function can be written as

$$\begin{aligned} F_X(a) &= P\{X \leq a\} \\ &= P\left(\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{a-\mu}{\sigma}\right) \end{aligned}$$

5.4.1 The Normal Approximation to the Binomial Distribution

The DeMoivre-Laplace Limit Theorem: If S_n denotes the number of successes that occur when n independent trials, each resulting in a success with probability p, are performed then, for any $a < b$,

$$P\left\{a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right\} \to \Phi(b) - \Phi(a)$$

as $n \to \infty$.

5.5 Exponential Random Variable

Exponential Random Variable: A continuous random variable whose probability density function is given, for some $\lambda > 0$, by

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

is said to be an *exponential* random variable (or, more simply, is said to be exponentially distributed) with parameter λ. The cumulative distribution $F(a)$ of an exponential random variable is given by

$$\begin{aligned} F(a) &= P\{X \leq a\} \\ &= \int_0^a \lambda e^{-\lambda x} dx \\ &= -e^{-\lambda x} \Big|_0^a \\ &= 1 - e^{-\lambda a} \quad a \geq 0 \end{aligned}$$

For an exponential random variable X with parameter λ,

$$E[X] = \frac{1}{\lambda}$$

and

$$\text{Var}(X) = \frac{1}{\lambda^2}$$

Memoryless Random Variable: We say that a nonnegative random variable X is *memoryless* if

$$P\{X > s + t | X > t\} = P\{X > s\} \quad \text{for all } s, t \geq 0$$

Since the above equation is satisfied when X is exponentially distributed, it follows that exponentially distributed random variables are memoryless.

5.5.1 Hazard Rate Functions

Hazard Rate Function: Consider a positive continuous random variable X that we interpret as being the lifetime of some item, having distribution function F and density f. The *hazard rate* (sometimes called the *failure rate* function $\lambda(t)$ of F is defined by

$$\lambda(t) = \frac{f(t)}{\overline{F}(t)} \qquad \overline{F}(t) = 1 - F$$

To interpret $\lambda(t)$, suppose that the item has survived for a time t and we desire the probability that it will not survive for an additional time dt. That is, consider $P\{X \in (t, t+dt)|X > t\}$. Now

$$\begin{aligned} P\{X \in (t, t+dt)|X > t\} &= \frac{P\{X \in (t, t+dt), X > t\}}{P\{X > t\}} \\ &= \frac{P\{X \in (t, t+dt)\}}{P\{X > t\}} \\ &\approx \frac{f(t)}{\overline{F}(t)} dt \end{aligned}$$

Thus, $\lambda(t)$ represents the conditional probability intensity that a t-unit-old item will fail.
For an exponentially distributed random variable, the hazard rate function is constant.

5.6 Other Continuous Distributions

5.6.1 The Gamma Distribution

The Gamma Distribution: A continuous random variable is said to have a gamma distribution with parameters (α, λ), $\lambda > 0$, and $\alpha > 0$ if its density function is given by

$$f(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

where $\Gamma(\alpha)$, called the gamma function, is defined as

$$\Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1} dy$$

The integration by parts of $\Gamma(\alpha)$ yields that

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$$

For integer values of n

$$\Gamma(n) = (n-1)!$$

When α is a positive integer, say $\alpha = n$, the gamma distribution with parameters (α, γ) often arises as the distribution of the amount of time one has to wait until a total of n events has occurred, when the conditions told for Poisson distribution are valid. Let T_n denote the time at which the nth event occurs, and note T_n is less then or equal to t if and only if the number of events that have occurred by time t is at least n. That is, when $N(t)$ equal to the number of events in $[0, t]$,

$$\begin{aligned} P\{T_n \le t\} &= P\{N(t) \ge n\} \\ &= \sum_{j=n}^{\infty} P\{N(t) = j\} \\ &= \sum_{j=n}^{\infty} \frac{e^{-\lambda t}(\lambda t)^j}{j!} \end{aligned}$$

where the final identity follows because the number of events in $[0, t]$ has a Poisson distribution with parameter λt. Differentiation of the above yields that the density function of T_n is as follows:

$$\begin{aligned} f(t) &= \sum_{j=n}^{\infty} \frac{e^{-\lambda t} j(\lambda t)^{j-1}\lambda}{j!} - \sum_{j=n}^{\infty} \frac{\lambda e^{-\lambda t}(\lambda t)^j}{j!} \\ &= \sum_{j=n}^{\infty} \frac{\lambda e^{-\lambda t}(\lambda t)^{j-1}}{(j-1)!} - \sum_{j=n}^{\infty} \frac{\lambda e^{-\lambda t}(\lambda t)^j}{j!} \\ &= \frac{\lambda e^{-\lambda t}(\lambda t)^{n-1}}{(n-1)!} \end{aligned}$$

Note that when $n = 1$, this distribution reduces to the exponential.

For a gamma distributed random variable we have,

$$E[X] = \frac{\alpha}{\lambda}$$

and

$$\mathrm{Var}(X) = \frac{\alpha}{\lambda^2}$$

5.7 The Distribution of a Function of a Random Variable

Theorem: Let X be a continuous random variable having probability density function f_X. Suppose that $g(x)$ is a strictly monotone (increasing or decreasing), differentiable (and thus continuous) function of x. Then the random variable Y defined $Y = g(X)$ has a probability density function given by

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

5.8 Some Solved Problems

1. Consider the probability density $f_X(x) = a\, e^{-b|x|}$, where X is a random variable whose allowable values range from $x = -\infty$ to $x = +\infty$. Find (a) the cumulative distribution function $F_X(x)$, (b) the relationship between a and b, and (c) the probability that the outcome X lies between 1 and 2.

   Solution: (a) The cumulative distribution function is

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\, dt = \int_{-\infty}^{x} a\, e^{-b|t|}\, dt$$
$$= \begin{cases} \frac{a}{b}\, e^{bx} & x \le 0 \\ \frac{1}{2} + \frac{a}{b}\,(1 - e^{-bx}) & x \ge 0 \end{cases}$$

   (b) In order that $f(x)$ be a probability density, it is necessary that

$$\int_{-\infty}^{+\infty} f(x)\, dx = \int_{-\infty}^{+\infty} a\, e^{-b|x|} dx = \frac{2a}{b} = 1$$

   so that $\frac{a}{b} = \frac{1}{2}$.
   (c) The probability that X lies in the range between 1 and 2 is

$$P(1 \le X \le 2) = \frac{b}{2} \int_{1}^{2} e^{-b|x|}\, dx = \frac{1}{2}\,(e^{-b} - e^{-2b})$$

2. A certain retailer for a petroleum product sells a random amount, X, each day. Suppose that X, measured in hundreds of gallons, has the probability density function

$$f_X(x) = \begin{cases} (3/8)x^2 & 0 \le x \le 2 \\ 0 & \text{elsewhere.} \end{cases}$$

   The retailer's profit turns out to be \$5 for each 100 gallons sold (5 cents per gallon) if $X \le 1$ and \$8 per 100 gallons if $X > 1$. Find the retailer's expected profit for any given day.

   Solution: Let $g(X)$ denote the retailer's daily profit. Then,

$$g(X) = \begin{cases} 5X, & 0 \le X \le 1 \\ 8X, & 1 < X \le 2 \end{cases}$$

   We want to find expected profit, and

$$\begin{aligned} E[g(X)] &= \int_{-\infty}^{\infty} g(x) f(x) dx \\ &= \int_{0}^{1} 5x \left[\left(\frac{3}{8}\right) x^2\right] dx + \int_{1}^{2} 8x \left[\left(\frac{3}{8}\right) x^2\right] dx \\ &= \frac{15}{(8)(4)}\, [x^4]_0^1 + \frac{24}{(8)(4)}\, [x^4]_1^2 \\ &= \frac{15}{32}\,(1) + \frac{24}{32}\,(15) \\ &= \frac{(15)(25)}{32} = 11.72 \end{aligned}$$

3. Let X have the probability density function given by

$$f_X(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

   (a) Find the density function of $U = 3X - 1$.

   (b) Find the density function of $U = -4X + 3$.

Solution: We know that for a given pdf, say $f_X(x)$, the probability that X is in a neighborhood of Δx around x is given by $f_X(x)\Delta x$. In this problem, the value of U is uniquely determined by the value of X. Then, the probability that U is in a neighborhood of Δu around u, i.e., $f_U(u)\Delta u$, is equal to the probability that X is in a neighborhood of Δx around $x = h^{-1}(u)$ where $h(x) = 3x - 1$. This means that

$$f_U(u)\Delta u = f_X[h^{-1}(u)]\Delta x \quad \Longrightarrow f_U(u) = f_X[h^{-1}(u)]\,\frac{\Delta x}{\Delta u}.$$

Here, we have,

$$x = h^{-1}(u) = \frac{u+1}{3}$$

and

$$\frac{\Delta x}{\Delta u} = \frac{dx}{du} = \frac{1}{3}$$

Thus,

$$\begin{aligned} f_U(u) &= f_X[h^{-1}(u)]\frac{dx}{du} = 2x\frac{dx}{du} \\ &= 2 \cdot \frac{u+1}{3} \cdot \frac{1}{3} \\ &= \frac{2(u+1)}{9} \qquad -1 < u < 2 \\ &= 0 \quad \text{elsewhere} \end{aligned}$$

The range of u (i.e., the range over which $f_U(u)$ is positive) is simply the interval $0 < x < 1$ transformed to the u-axis by the function $u = 3x - 1$. This results in $u \in [-1, 2]$.

For $u = h(x) = -4x + 3$, we have,

$$x = h^{-1}(u) = \frac{3-u}{4}$$

and

$$\frac{dx}{du} = -\frac{1}{4}$$

As the function $h(x)$ is decreasing in x, we have $\Delta x/\Delta u = -dx/du$ (note that Δx and Δu are always positive but dx and du may be positive or negative) and we can write,

$$\begin{aligned} f_U(u) &= f_X[h^{-1}(u)]\left|\frac{dx}{du}\right| = 2x\left|\frac{dx}{du}\right| \\ &= 2 \cdot \frac{3-u}{4} \cdot \frac{1}{4} \\ &= \frac{3-u}{8} \qquad -1 < u < 3 \\ &= 0 \quad \text{elsewhere} \end{aligned}$$

The range of u is equal to, $u \in [-1, 3]$.

4. Consider a random variable X with the following pdf:

$$f_X(x) = 1 - a|x|, \quad |x| \leq 1/a, \quad f_X(x) = 0, \quad \text{otherwise}$$

   (a) Find the constant a and compute the mean and the standard deviation of X.
   (b) The random variable X is applied to a "full-wave" rectifier whose output-input gain characteristic is $y = b|x|$. Determine the mean and standard deviation of the output random variable.
   (c) The random variable X is applied to a "half-wave" rectifier whose output-input gain characteristic is $y = b|x|, x \geq 0$ and $y = 0, x < 0$. Determine the mean and standard deviation of the output random variable.

Solution: The pdf of X is shown in the following figure.

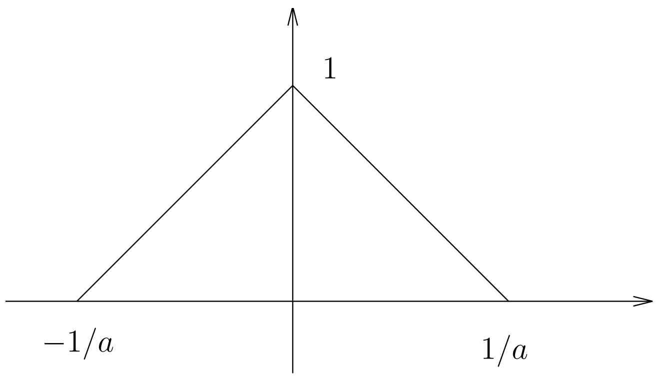


Figure 7: Pdf related to above Problem.

We should have,

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_{-1/a}^{1/a} f_X(x) = 1$$

This results in $a = 1$. We also have:

$$E(X) = 0$$

and,

$$\sigma_X = \sqrt{E(X^2)} = \sqrt{1/6}$$

In parts 2 and 3, we want to compute the average value of a function of a random variable. Given a random variable X, the general formula for the average value of the function $H(X)$ is,

$$E[H(X)] = \int_{-\infty}^{\infty} H(X) f_X(x)dx$$

For the case of full wave rectifier, the function is equal to, $H(X) = b|X|$. This results in,

$$E(Y) = E(b|X|) = b[\int_{-1}^{0} (-x) f_X(x)dx + \int_{0}^{1} x f_X(x)dx] = 2b \int_{0}^{1} x(1-x)dx = \frac{b}{3}$$

and, similarly,

$$E(Y^2) = 2b^2 \int_0^1 x^2(1-x)dx = \frac{b^2}{6}$$

This results in $\sigma_Y = \sqrt{[E(Y^2)] - [E(Y)]^2} = \sqrt{b^2/18}$.

For the case of half wave rectifier, we note that the output is in part continuous and in part discrete (note that all the negative value of X are mapped to zero). This means that the output of the half-rectifier is with probability 1/2 equal to zero, while for values greater than zero, it obeys the pdf of X. This results in,

$$E(Y) = (0 \times 1/2) + \int_0^1 bx(1-x)dx = \frac{b}{6}$$

and

$$E(Y^2) = (0^2 \times 1/2) + \int_0^1 b^2x^2(1-x)dx = \frac{b^2}{12}$$

This results in $\sigma_Y = \sqrt{[E(Y^2)] - [E(Y)]^2} = \sqrt{b^2/18}$.

5. The random variable X of the life length of certain kind of battery (in hundreds of hours) is equal to:

$$f(x) = \begin{cases} \frac{1}{2}e^{-x/2} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

(i) Find the probability that the life of a given battery is less than 200 or greater than 400 hours.
(ii) Find the probability that a battery of this type lasts for 300 hours if we know that it has already been in use for 200 hours.

Solution: (i) Let A denote the event that X is less than 2 and B the event that X is greater than 4. Then, because A and B are mutually exclusive,

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) \\ &= \int_0^2 \frac{1}{2}e^{-x/2}\,dx + \int_4^\infty \frac{1}{2}e^{-x/2}\,dx \\ &= (1 - e^{-1}) + e^{-2} \\ &= 1 - 0.368 + 0.135 \\ &= 0.767 \end{aligned}$$

(ii) We are interested in $P(X > 3|X > 2)$ and, by the definition of conditional probability,

$$P(X > 3|X > 2) = \frac{P(X > 3)}{P(X > 2)}$$

because the intersection of the events $(X > 3)$ and $(X > 2)$ is the event $(X > 3)$. Now

$$\frac{P(X > 3)}{P(X > 2)} = \frac{\int_3^\infty \frac{1}{2}e^{-x/2}\,dx}{\int_2^\infty \frac{1}{2}e^{-x/2}\,dx} = \frac{e^{-3/2}}{e^{-1}} = e^{-1/2} = 0.606$$

6. The failure of a circuit board interrupts work by a computer system until a new board is delivered. Delivery time X is uniformly distributed over the interval 1 to 5 days. The cost C of this failure and

interruption consists of a fixed cost c_0 for the new part and a cost that increases proportional to X^2, i.e,

$$C = c_0 + c_1 X^2$$

(a) Find the probability that the delivery time takes 2 or more days.
(b) Find the expected cost of a single failure, in terms of c_0 and c_1.

Solution: (a) The delivery time X is distributed uniformly from 1 to 5 days, which gives

$$f(x) = \begin{cases} \frac{1}{4} & 1 \leq x \leq 5 \\ 0 & \text{elsewhere} \end{cases}$$

Thus,

$$\begin{aligned} P(X \geq 2) &= \int_2^5 \left(\frac{1}{4}\right) dx \\ &= \frac{1}{4}(5-2) = \frac{3}{4} \end{aligned}$$

(b) We know that

$$E(C) = c_0 + c_1 E(X^2)$$

so it remains to find $E(X^2)$. This could be found directly from the definition or by using the variance and the fact that

$$E(X^2) = V(X) + \mu^2$$

Using the latter approach,

$$\begin{aligned} E(X^2) &= \frac{(b-a)^2}{12} + \left(\frac{a+b}{2}\right)^2 \\ &= \frac{(5-1)^2}{12} + \left(\frac{1+5}{2}\right)^2 = \frac{31}{3} \end{aligned}$$

Thus,

$$E(C) = c_0 + c_1 \left(\frac{31}{3}\right)$$

7. Let X denote the life time (in hundreds of hours) of a certain type of electronic component. These components frequently fail immediately upon insertion into the system. It has been observed that the probability of immediate failure is 1/4. If a component does not fail immediately, the life-length distribution has the exponential density:

$$f(x) = \begin{cases} e^{-x} & x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

Find the distribution function for X and evaluate $P(X > 10)$.

Solution: There is only one discrete point, $X = 0$, and this point has probability 1/4. It follows that X is a mixture of two random variables, X_1 and X_2, where X_1 has a probability of one at the point zero and X_2 has the given exponential density. That is,

$$F_1(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$

and

$$\begin{aligned} F_2(x) &= \int_0^x e^{-y}\, dy \\ &= 1 - e^{-x} \quad x > 0 \end{aligned}$$

Now,

$$F(x) = \left(\frac{1}{4}\right) F_1(x) + \left(\frac{3}{4}\right) F_2(x)$$

Hence,

$$\begin{aligned} P(X > 10) &= 1 - P(X \le 10) \\ &= 1 - F(10) \\ &= 1 - \left[\frac{1}{4} + \left(\frac{3}{4}\right)(1 - e^{-10})\right] \\ &= \left(\frac{3}{4}\right)[1 - (1 - e^{-10})] = \left(\frac{3}{4}\right) e^{-10} \end{aligned}$$

8. Suppose X has density function $f(x)$ for $-1 \le x \le 1$ and 0 otherwise. Find the density function of (a) $Y = |X|$, (b) $Z = X^2$.

   Solution: (a) For $y > 0$, $P(Y \le y) = P(-y \le X \le y) = F(y) - F(-y)$. Differentiating the cdf, we conclude that the pdf of Y is $f(y) + f(-y)$ for $0 < y < 1$ and 0 elsewhere.

   You may also try to use the following relationship:

$$f_Y(y) = \sum_{x_i} \frac{f_X(x_i)}{\left|\frac{dy}{dx}\right|}$$

   (b) For $z > 0$, $P(Z \le z) = P(-\sqrt{z} \le X \le \sqrt{z}) = F(\sqrt{z}) - F(-\sqrt{z})$. Hence the pdf of Z is

$$[f(\sqrt{z}) + f(-\sqrt{z})] \cdot (\frac{1}{2} z^{-1/2}), \text{ for } 0 < z < 1$$

   and 0 otherwise.

9. Suppose X is uniform on (0,1). Find the density function of $Y = X^n$.

   Solution: Suppose X has density f_X and $P(a < X < b) = 1$. Let $Y = r(X)$. Suppose $r : (a, b) \to (\alpha, \beta)$ is continuous and strictly increasing, and let $s : (\alpha, \beta) \to (a, b)$ be the inverse of r. Then, we know that Y has density

$$f_Y(y) = f_X[s(y)]s'(y) \quad \text{for } y \in (\alpha, \beta)$$

   Therefore, since X has density function $f_X(x) = 1$ for $0 < x < 1$, and $r(x) = x^n$ has inverse $s(x) = x^{1/n}$, the theorem gives the density function $f_X(y^{\frac{1}{n}}) \cdot \frac{1}{n} y^{\frac{1}{n}-1} = \frac{1}{n} y^{\frac{1}{n}-1}$ for $0 < y < 1$.

10. Suppose X is uniform on $(0, \pi/2)$ and $Y = \sin X$. Find the density function of Y. The answer is called the **arcsine law** because the distribution function contains the arcsine function.

    Solution: The density function of X is $f_X(x) = 2/\pi$ for $0 < x < \pi/2$ and $r(x) = \sin x$ has the inverse $s(x) = \sin^{-1} x$. Using the same result as in the previous problem, we obtain the density function $f_X[s(y)] \cdot s'(y) = \frac{2}{\pi} \cdot \frac{1}{\sqrt{1 - y^2}}$ for $0 < y < 1$.

11. Suppose X has density function $3x^{-4}$ for $x \geq 1$. (a) Find a function g so that $g(X)$ is uniform on (0,1). (b) Find a function h so that if U is uniform on (0,1), $h(U)$ has density function $3x^{-4}$ for $x \geq 1$.

    Solution: Suppose X has a continuous distribution. Then $Y = F_X(X)$ is uniform on (0,1).

    (a) $P(X \leq x) = \int_1^x 3y^{-4}\,dy = 1 - x^{-3}$ for $x > 1$ and 0 elsewhere. The above statement tells that $Y = g(X) = 1 - X^{-3}$ is uniform on (0,1).

    Suppose U has a uniform distribution on (0,1). Then $Y = F^{-1}(U)$ has distribution function F.

    (b) $F(x)$ has inverse $F^{-1}(x) = (1-x)^{-\frac{1}{3}}$. The above statement says that $F^{-1}(U) = (1-U)^{-\frac{1}{3}}$ has the given density function.

12. A Gaussian distributed random variable X with zero mean and the unit variance is applied to a "full-wave" rectifier whose output-input gain characteristic is $y = |x|/a, \quad a > 0$. Determine the pdf of the output random variable Y.

    The mapping is one-to-one for $x < 0$ and one-to-one for $x > 0$. In both cases, $y > 0$. Using the equation

$$f_Y(y) = \frac{f_X(x)}{|dy/dx|}$$

we have for $x > 0$

$$f_Y(y) = a\left[\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\right]_{x=ay} = \frac{a}{\sqrt{2\pi}}e^{-a^2y^2/2} \qquad y > 0$$

and for $x < 0$

$$f_Y(y) = a\left[\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\right]_{x=-ay} = \frac{a}{\sqrt{2\pi}}e^{-a^2y^2/2} \qquad y > 0$$

Adding these two results, we obtain,

$$f_Y(y) = a\sqrt{\frac{2}{\pi}}e^{-a^2y^2/2}u(y), \quad \text{where } \ u(y) \ \text{ is the unit step function}$$

We could equivalently use the following relationship:

$$f_Y(y) = \sum_k \frac{f_X(x_k)}{|dg/dx_k|}\Bigg|_{x_k=g^{-1}(y)}$$

where $g(x) = |x|/a$.

13. The random variable X of the previous problem is applied to the half-wave rectifier whose output-input characteristic is $y = (x/a)u(x)$. Determine the pdf of the output.

    Solution: For $x > 0$, it can easily be shown that

$$f_y(y) = \frac{a}{\sqrt{2\pi}}e^{-a^2y^2/2}, \qquad y > 0$$

For $x < 0$, however all points of the input are mapped into zero in the output. To conserve probability we must add a contribution of $\int_{-\infty}^0 f_X(x)dx = 1/2$ at the point $y = 0$ so that

$$f_y(y) = \frac{a}{\sqrt{2\pi}}e^{-a^2y^2/2}u(y) + \left(\frac{1}{2}\right)\delta(y)$$

14. Suppose X has density x^{-2} for $x \geq 1$ and $Y = X^{-2}$. Find the pdf of Y.

    Solution: Noting $P(Y \leq x) = P(X \geq x^{-\frac{1}{2}}) = 1 - F(x^{-\frac{1}{2}})$ for $x \leq 1$ leads to the density function of Y by differentiating:

$$\frac{d}{dx}P(Y \leq x) = F'(x^{-\frac{1}{2}})(\frac{1}{2}x^{-3/2}) = \frac{1}{2}x^{-1/2}, \quad for \quad 0 < x \leq 1$$

15. The actual weight of a bag of sugar is assumed to be a normal random variable with mean 202 grams and standard deviation of 3 grams. If a bag weighs less than 199 grams or more than 205 grams it is rejected.
    (i) What is the probability that a bag will be rejected?
    (ii) Given that a bag was rejected, what is the probability it weighs less than 195 grams?
    (iii) If the standard deviation of the filling process is changed to σ, but the mean remains at 202 grams, what is the largest value that σ can have so that the probability a bag is rejected is less than .01?

    Give numerical answers.

    Solution: (i) $P(\text{bag rejected}) = 1 - P(199 < X < 205) = 1 - P(\frac{199-202}{3} < \frac{X-202}{3} < \frac{205-202}{3})$ $= 1 - (\Phi(1) - \Phi(-1)) = 0.84134 - 0.15866 = 0.31732$.

    (ii) Let A =bag rejected, $B = \{X < 195\}$, then $P(B|A) = P(AB)/P(A)$

$$= P(X < 195)/P(A) = \Phi(\frac{195-202}{3})/0.31732 = \Phi(-7/3)/0.31732 = 0.00982/0.31732. = 0.0309.$$

    (iii) $P(\text{bag rejected}) = 1 - [\Phi(3/\sigma) - \Phi(-3/\sigma)] = 2(1 - \Phi(3/\sigma) = 0.01$ or $\Phi(3/\sigma) = 0.995$ and $3/\sigma = 2.575$ and $\sigma = 1.165$.

16. The error in a linear measurement, is assumed to be a normal random variable with mean 0 and variance σ^2, in mm^2.

    (i) What is the largest value of σ allowable if $P(|X| < 2)$ is to be at least 0.90?
    (ii) If $\sigma = 2$ evaluate $P(X > 4 \mid |X| < 5)$.

    Solution: (i) $P(|X| < 2) = \Phi(2/\sigma) - \Phi(-2/\sigma) = 2\Phi(2/\sigma) - 1 = 0.9$ or $\Phi(2/\sigma) = 0.95$ and $2/\sigma = 1.65$ and $\sigma = 1.212$.

    (ii) $P(X > 4 \,||X| < 5) = P(X > 4 \text{ and } |X| < 5)/P(|X| < 5) =$

$$\frac{\Phi(5/2) - \Phi(4/2)}{\Phi(5/2) - \Phi(-5/2)} = 0.0167.$$

17. The projection of a point chosen at random on the circumference of a circle of radius a onto a fixed diameter, has the cdf:

$$F_X(x) = \begin{cases} 1 & x \geq a \\ \frac{1}{2} + \frac{1}{\pi}\arcsin\frac{x}{a} & -a \leq x \leq a \\ 0 & \text{otherwise} \end{cases}$$

(i) Determine the probability that X will be on the interval $(-a/2, a/2)$

(ii) Find the probability density function of X

(iii) Find the mode and the median of the distribution.

Solution: (i)

$$P(-\frac{a}{2} < X < \frac{a}{2}) = F_X(\frac{a}{2}) - F_X(-\frac{a}{2}) = \frac{2}{\pi}\arcsin(\frac{1}{2}) = \frac{1}{3}.$$

The density function is given by

$$f_X(x) = \frac{d}{dx}F_X(x) = \frac{d}{dx}(\frac{1}{2} + \frac{1}{\pi}\arcsin(\frac{x}{a})) = \frac{1}{\pi(a^2 - x^2)^{1/2}}\ , \qquad -a \le x \le a$$

and zero otherwise. The mode of the distribution is the point at which the pdf achieves its maximum value. This distribution has no mode. The median is the point at which $F_X(x) = 1/2$, which is $x = 0$.

18. The signal strength in volts at the input to an antenna, is a random variable with cdf

$$F(x) = 1 - e^{-x^2/a} \quad , x \ge 0, \quad a > 0.$$

(i) Find the probability density function, mean and variance of X.

(ii) If ten independent samples of the signal strength are taken, what is the probability that exactly 7 of them will be greater than 2 volts?

Solution: (i) The probability density function is:

$$\begin{aligned} f_X(x) &= \frac{d}{dx}F_X(x) = \frac{2x}{a}e^{-x^2/a}\ , \quad x \ge 0 \\ E(X) &= \frac{2}{a}\int_0^\infty x^2 e^{-x^2/a}dx = \frac{\sqrt{a\pi}}{2} \\ E(X^2) &= \frac{2}{a}\int_0^\infty x^3 e^{-x^2/a}dx = a \\ Var(X) &= a(1 - \frac{\pi}{4}) \end{aligned}$$

(ii) The probability that one sample is greater than 2 is $p = \exp(-4/a) = 1 - F_X(2)$ and so:

$$P(7 \text{ out of 10 greater than 2}) = \binom{10}{7}p^7(1-p)^3\ .$$

19. The random variable X is normal with mean 81 and variance 16 while Y is normal with mean 85 and variance 4. Which random variable is more likely to be less than 88?

Solution:

$$\begin{aligned} P(X \le 88) = &\ P(\frac{X - 81}{4} \le \frac{7}{4}) = \Phi(\frac{7}{4}) \\ P(Y \le 88) = &\ P(\frac{Y - 85}{2} \le \frac{3}{2}) = \Phi(\frac{3}{2}) \end{aligned}$$

and hence X is more likely to be less than 88.

20. The lifetime of an electronic component is a random variable which has an exponential pdf. If 50% of the components fail before 2,000 hours, what is the average lifetime of the component?

    Solution: The pdf is of the form $\alpha \exp(-\alpha x)$ and $P(X \leq 2000) = 1 - \exp(-2,000\alpha) = 0.5$ Thus $\alpha = -\ln(0.5)/2,000$ and the average lifetime of the component is $1/\alpha = 2,000/\ln(2)$.

21. The probability density function of a random variable X is

$$f_X(x) = ax^2 e^{-kx} \quad , \; k > 0 \; , \quad 0 \leq x < \infty \; .$$

Find:

(i) the coefficient a in terms of k.

(ii) the cdf of X.

(iii) The probability that $0 \leq X \leq 1/k$.

Solution: (i) From the formula for the gamma function it is easily shown that

$$\int_0^{\infty} ax^2 e^{-kx} dx = \frac{2a}{k^3} \; .$$

and $a = k^3/2$.

(ii) The cdf is

$$F_X(x) = \int_0^x \frac{k^3}{2} y^2 e^{-ky} dy = 1 - \frac{k^2x^2 + 2kx + 2}{2} e^{-kx} \; , \quad x \geq 0 \; .$$

(iii)

$$P(0 \leq X \leq \frac{1}{k}) = F_X(\frac{1}{k}) = 1 - \frac{5}{2} e^{-1} \; .$$

6 Jointly Distributed Random Variables

6.1 Joint Distribution Functions

So far, we have only concerned ourselves with probability distributions for single random variables. In this chapter, we deal with probability statements concerning two or more random variables. In order to deal with such probabilities, we define, for any two random variables X and Y, the *joint cumulative probability distribution function* of X and Y by

$$F(a,b) = P\{X \leq a, Y \leq b\} \qquad -\infty < a,b < \infty$$

The distribution of X can be obtained from the joint distribution of X and Y as follows:

$$\begin{aligned} F_X(a) &= P\{X \leq a\} \\ &= P\{X \leq a, Y < \infty\} \\ &= F(a,\infty) \end{aligned}$$

and similarly

$$F_Y(b) = F(\infty, b)$$

The distribution functions F_X and F_Y are sometimes referred to as the *marginal* distributions of X and Y.

In the case that both X and Y are both discrete random variables, it is convenient to define the *joint probability mass function* of X and Y by

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

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

$$p_X(x) = P\{X = x\} = \sum_y p(x,y)$$

and similarly

$$p_Y(y) = P\{Y = y\} = \sum_x p(x,y)$$

Jointly Continuous Random Variables We say that X and Y are *jointly continuous* if there exists a function $f(x,y)$ defined for all real x and y, having the property that for every set C of pairs of real numbers (that is, C is a set in the two dimensional plane)

$$P\{(X,Y) \in C\} = \iint_{(x,y)\in C} f(x,y)\,dx\,dy$$

The function $f(x,y)$ is called the *joint probability density function* of X and Y.

If A and B are any sets of real numbers, then by defining $C = \{(x,y) : x \in A,\ y \in B\}$, we see that

$$P\{X \in A, Y \in B\} = \int_B \int_A f(x,y)\,dx\,dy$$

Because

$$F(a,b) = P\{X \in (-\infty,a], Y \in (-\infty,b]\} = \int_{-\infty}^{b} \int_{-\infty}^{a} f(x,y)\,dx\,dy$$

after differentiating

$$f(a,b) = \frac{\partial^2}{\partial a\,\partial b} F(a,b)$$

wherever the partial derivations are defined.
Another interpretation of the joint density function is as follows:

$$P\{a < X < a + da, b < Y < b + db\} = \int_{b}^{b+db} \int_{a}^{a+da} f(x, y)dxdy \approx f(a, b)dadb$$

when da and db are small and $f(x, y)$ is continuous at a, b. Hence $f(a, b)$ is a measure of how likely it is that the random vector (X, Y) will be near (a, b).
Similar to discrete case we have

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

and

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

We can also define joint probability distributions for n random variables in exactly the same manner as we did for $n = 2$. The joint cumulative probability distribution function $F(a_1, a_2, \ldots, a_n)$ of the n random variables $X_1, X_2, \ldots, X_n$ is defined by

$$F(a_1, a_2, \ldots, a_n) = P\{X_1 \leq a_1, X_2 \leq a_2, \ldots, X_n \leq a_n\}$$

Further, the n random variables are said to be jointly continuous if there exists a function $f(x_1, x_2, \ldots, x_n)$, called the joint probability density function, such that for any set C in n-space

$$P\{(X_1, X_2, \ldots, X_n) \in C\} = \iint \cdots \int_{(x_1, x_2, \ldots, x_n) \in C} f(x_1, \ldots, x_n)\, dx_1\, dx_2\, dx_n$$

6.2 Independent Random Variables

The random variables X and Y are said to be *independent* if for any two sets of real numbers A and B,

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\}$$

In terms of the joint distribution function F of X and Y, we have that X and Y are independent if

$$F(a,b) = F_X(a)F_Y(b) \quad \text{for all } a, b$$

When X and Y are discrete random variables, the condition of independence is equivalent to

$$p(x,y) = p_X(x)p_Y(y) \quad \text{for all } x, y$$

and for continuous case,

$$f(x,y) = f_X(x)f_Y(y) \quad \text{for all } x, y$$

Proposition: The continuous (discrete) random variables X and Y are independent if and only if their joint probability density (mass) function can be expressed as

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \quad -\infty < x < \infty, -\infty < y < \infty$$

Remark: For set of random variables $X_1, \ldots, X_n$ we can show that theses random variables are independent by showing that

X_2	is independent of	X_1
X_3	is independent of	X_1, X_2
X_4	is independent of	X_1, X_2, X_3
	$\cdots$	
	$\cdots$	
X_n	is independent of	$X_1, \ldots, X_{n-1}$

6.3 Sums of Independent Random Variables

Suppose that X and Y are independent, continuous random variables having probability distribution functions f_x and f_y. The cumulative distribution function of $X + Y$ is obtained as follows:

$$\begin{aligned}
F_{X+Y}(a) &= P\{X + Y \le a\} \\
&= \iint_{x+y\le a} f_X(x)f_Y(y)\,dx\,dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_X(x)f_Y(y)\,dx\,dy \\
&= \int_{-\infty}^{\infty} F_X(a-y)f_Y(y)\,dy
\end{aligned}$$

By differentiating, we obtain that the probability density function f_{X+Y} of $X+Y$ is given by

$$\begin{aligned} f_{X+Y}(a) &= \frac{d}{da}\int_{-\infty}^{\infty} F_X(a-y)f_Y(y)\,dx\,dy \\ &= \int_{-\infty}^{\infty} f_X(a-y)f_Y(y)\,dy \end{aligned}$$

Proposition, Sum of Gamma Random Variables: If X and Y are independent gamma random variables with respective parameters (s,λ) and (t,λ), then $X+Y$ is a gamma random variable with parameters $(s+t,\lambda)$.
Proof:

$$\begin{aligned} f_{X+Y}(a) &= \frac{1}{\Gamma(s)\Gamma(t)}\int_0^a \lambda e^{-\lambda(a-y)}[\lambda(a-y)]^{s-1}\lambda e^{-\lambda y}(\lambda y)^{t-1}dy \\ &= Ke^{-\lambda a}\int_o^a (a-y)^{s-1}y^{t-1}dy \\ &= Ke^{-\lambda a}a^{s+t-1}\int_o^1 (1-x)^{s-1}x^{t-1}dx \quad \text{by letting } x=\frac{y}{a} \\ &= Ce^{-\lambda a}a^{s+t-1} \end{aligned}$$

where C is a constant that does not depend on a. But as the above is a density function and thus must integrate to 1, the value of C is determined, and we have

$$f_{X+Y}(a) = \frac{\lambda e^{-\lambda a}(\lambda a)^{s+t-1}}{\Gamma(s+t)}$$

Sum of Square of Standard Normal Random Variables: If $Z_1, Z_2, \ldots, Z_n$ are independent standard normal random variables, then $Y \equiv \sum_{i=1}^{n} Z_i^2$ is said to have the chi-squared (sometimes seen as χ^2) distribution with n degrees of freedom. When $n=1$, $Y=Z_1^2$, we can see that its probability density function is given by

$$\begin{aligned} f_{Z^2}(y) &= \frac{1}{2\sqrt{y}}[f_Z(\sqrt{y})+f_Z(-\sqrt{y})] \\ &= \frac{1}{2\sqrt{y}}\frac{2}{\sqrt{2\pi}}e^{-y/2} \\ &= \frac{\frac{1}{2}e^{-y/2}(y/2)^{1/2-1}}{\sqrt{\pi}} \end{aligned}$$

Noting that $\Gamma(\frac{1}{2})=\sqrt{\pi}$, this is the gamma distribution with parameters $(\frac{1}{2},\frac{1}{2})$. From the above proposition we obtain that the χ^2 distribution with n degree of freedom is just the gamma distribution with parameters $(\frac{n}{2},\frac{1}{2})$ and hence has the probability density function as

$$\begin{aligned} f_{\chi^2}(y) &= \frac{\frac{1}{2}e^{-y/2}(\frac{y}{2})^{n/2-1}}{\Gamma(\frac{n}{2})} \qquad y>0 \\ &= \frac{e^{-y/2}y^{n/2-1}}{2^{n/2}\Gamma(\frac{n}{2})} \qquad y>0 \end{aligned}$$

When n is an even integer, $\Gamma(\frac{n}{2})=[(n/2)-1]!$, and when n is odd, $\Gamma(\frac{n}{2})$ can be obtained from iterating the relationship $\Gamma(t)=(t-1)\Gamma(t-1)$ and then using $\Gamma(\frac{1}{2})=\sqrt{\pi}$.

Proposition, Sum of Normal Random Variables: If $X_i, i = 1, \ldots, n$ are independent random variables that are normally distributed with respective parameters $\mu_i, \sigma_i^2, i = 1, \ldots, n$, then $\sum_{i=1}^{n} X_i$ is normally distributed with parameters $\sum_{i=1}^{n} \mu_i$ and $\sum_{i=1}^{n} \sigma_i^2$.

6.4 Conditional Distributions: Discrete Case

For any two events E and F, the conditional probability of E given F is defined, provided that $P(F) > 0$, by

$$P(E|F) = \frac{P(EF)}{P(F)}$$

Hence, If X and Y are discrete random variables, it is natural to define the conditional probability mass function of X given by $Y = y$, by

$$\begin{aligned} p_{X|Y}(x|y) &= P\{X = x|Y = y\} \\ &= \frac{P\{X = x, Y = y\}}{P\{Y = y\}} \\ &= \frac{p(x, y)}{p_Y(y)} \end{aligned}$$

for all values of y such that $p_Y(y) > 0$. Similarly, the conditional probability distribution function of X given that $Y = y$ is defined, for all y such that $p_Y(y) > 0$, by

$$\begin{aligned} F_{X|Y}(x|y) &= P\{X \leq x|Y = y\} \\ &= \sum_{a \leq x} p_{X|Y}(a|y) \end{aligned}$$

If X and Y are independent, then

$$p_{X|Y}(x|y) = P\{X = x\}$$

6.5 Conditional Distributions: Continuous Case

If X and Y have a joint density function $f(x, y)$, then the conditional probability density function of X, given that $Y = y$, is defined for all values of y such that $f_Y(y) > 0$, by

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

6.6 Joint Probability Distribution Functions of Random Variables

Let X_1 and X_2 be jointly continuous random variables with joint probability density function f_{X_1,X_2}. It is sometimes necessary to obtain the joint distribution of the random variables Y_1 and Y_2, which arise as functions of X_1 and X_2. Specifically, suppose that $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$. Assume that functions g_1 and g_2, satisfy the following conditions:

1. The equations $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$ can be uniquely solved for x_1 and x_2 in terms of y_1 and y_2 with solutions given by, $x_1 = h_1(y_1, y_2)$ and $x_2 = h_2(y_1, y_2)$.
2. The functions g_1 and g_2 have the continuous partial derivatives at all points (x_1, x_2) that are such that the following 2×2 determinant

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

   at all points (x_1, x_2).

Under these two conditions it can be shown that the random variables Y_1 and Y_2 are jointly continuous with joint density function given by

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(x_1, x_2)|J(x_1, x_2)|^{-1}$$

where $x_1 = h_1(y_1, y_2)$ and $x_2 = h_2(y_1, y_2)$.
When the joint density function of n random variables $X_1, X_2, \ldots, X_n$ is given and we want to compute the joint density function of $Y_1, Y_2, \ldots, Y_n$, where $Y_1 = g_1(X_1, \ldots, X_n)$, $Y_2 = g_2(X_1, \ldots, X_n), \ldots,$ $Y_n = g_n(X_1, \ldots, X_n)$ the approach is the same. Namely, we assume that the functions g_i have continuous partial derivatives and that the Jacobian determinant $J(x_1, \ldots, x_n) \neq 0$ at all points $(x_1, \ldots, x_n)$, where

$$J(x_1, \ldots, x_n) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial g_n}{\partial x_1} & \frac{\partial g_n}{\partial x_2} & \cdots & \frac{\partial g_n}{\partial x_n} \end{vmatrix}$$

6.7 Some Solved Problems

1. Let R_1 and R_2 be independent, each with density $f(x) = e^{-x}, x \geq 0; f(x) = 0, x < 0$. Let R_3=max(R_1,R_2). Compute E(R_3).

   Solution:

$$\begin{aligned} E(R_3) &= E[g(R_1,R_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{12}(x,y)\,dx\,dy \\ &= \int_0^{\infty}\int_0^{\infty} \max(x,y) e^{-x}e^{-y}\,dx\,dy \end{aligned}$$

   Now max $(x, y) = x$ if $x \geq y$; max $(x, y) = y$ if $x \leq y$. Thus

$$\begin{aligned} E(R_3) &= \int\int_{x\geq y} xe^{-x}e^{-y}\,dx\,dy + \int\int_{y\geq x} ye^{-x}e^{-y}\,dx\,dy \\ &= \int_{x=0}^{\infty} xe^{-x}\int_{y=0}^{x} e^{-y}\,dy\,dx + \int_{y=0}^{\infty} ye^{-y}\int_{x=0}^{y} e^{-x}\,dx\,dy \end{aligned}$$

   The two integrals are equal, since one may be obtained from the other by interchanging x and y. Thus

$$\begin{aligned} E(R_3) &= 2\int_0^{\infty} xe^{-x}\int_0^{x} e^{-y}\,dy\,dx = 2\int_0^{\infty} xe^{-x}(1-e^{-x})\,dx \\ &= 2\int_0^{\infty} xe^{-x}\,dx - 2\int_0^{\infty}\frac{z}{2}e^{-z}\frac{dz}{2} = \frac{3}{2}\Gamma(2) = \frac{3}{2} \end{aligned}$$

2. We arrive at a bus stop at time $t = 0$. Two buses A and B are in operation. The arrival time R_1 of bus A is uniformly distributed between 0 and t_A minutes, and the arrival time R_2 of bus B is uniformly distributed between 0 and t_B minutes, with $t_A \leq t_B$. The arrival times are independent. Find the probability that bus A will arrive first.

   Solution: We are looking for the probability that $R_1 < R_2$. Since R_1 and R_2 are independent, the conditional density of R_2 given R_1 is

$$\frac{f(x,y)}{f_1(x)} = f_2(y) = \frac{1}{t_B}, \quad 0 \leq y \leq t_B$$

   If bus A arrives at $x, 0 \leq x \leq t_A$, it will be first provided that bus B arrives between x and t_B. This happens with probability $(t_B - x)/t_B$. Thus

$$P\{R_1 < R_2 | R_1 = x\} = 1 - \frac{x}{t_B}, \quad 0 \leq x \leq t_A$$

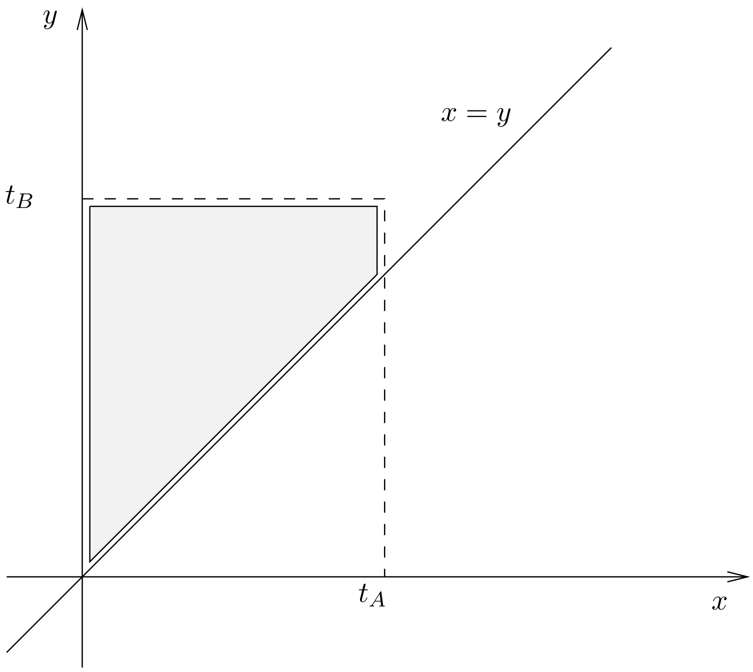
   Hence,

$$\begin{aligned} P\{R_1 < R_2\} &= \int_{-\infty}^{\infty} P\{R_1 < R_2|R_1 = x\} f_1(x)\,dx \\ &= \int_0^{t_A}\left(1 - \frac{x}{t_B}\right)\frac{1}{t_A}\,dx = 1 - \frac{t_A}{2t_B} \end{aligned}$$

   Alternatively, we may simply use the joint density:

$$\begin{aligned} P\{R_1 < R_2\} &= \int\int_{x<y} f(x,y)\,dx\,dy \\ &= \text{the shaded area in figure below, divided by total area } t_A t_B \\ &= 1 - \frac{t_A^2/2}{t_A t_B} = 1 - \frac{t_A}{2t_B} \end{aligned}$$

as before.



3. Suppose that X and Y have joint density $f(x,y) = (3x^2 + 4xy)/2$ when $0 < x, y < 1$. Find the marginal density of X and the conditional density of Y given $X = x$.

   Solution:
   $$f_X(x) = \int_0^1 (3x^2 + 4xy)/2\, dy = \frac{3}{2}x^2 + x, \text{ for } 0 < x < 1$$
   and
   $$f_Y(y|X = x) = \frac{3x^2 + 4xy}{2} / (\frac{3}{2}x^2 + x) = \frac{3x + 4y}{3x + 2}, \text{ for } 0 < y < 1.$$

4. Suppose the joint pdf for the random variables X and Y is given by
   $$f_{X,Y}(x,y) = \begin{cases} c, & 0 \le x \le y \le 1 \\ 0, & \text{otherwise.} \end{cases}$$
   Find the following: (a) The constant c, (b) The marginal pdfs $f_X(x)$ and $f_Y(y)$, and (c) The probability that $X + Y < 1$.

   Solution: (a) We know that the area under the pdf should be equal to one, and in this case,
   $$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx\, dy = \int_0^1 \int_x^1 c\, dy\, dx = 1$$
   Thus
   $$c/2 = 1 \Longrightarrow c = 2$$
   (b)
   $$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$$

so,

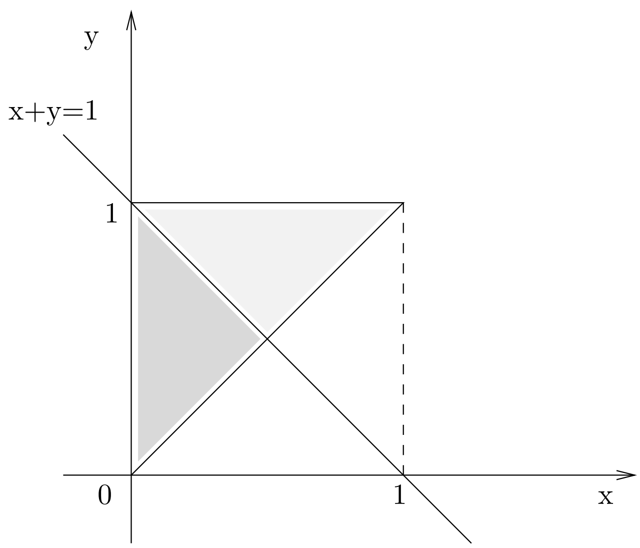
$$f_X(x) = \int_{y=x}^{1} 2\,dy = \begin{cases} 2(1-x), & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Similarly,

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_{x=0}^{y} 2\,dx \\ &= \begin{cases} 2y, & 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

(c) By noticing the following figure, it is clear that:

$$P(X+Y<1) = \frac{1}{2}$$



5. Consider a joint pdf for random variables X and Y defined as

$$f_{X,Y}(x,y) = \begin{cases} cxy, & 0 \le x, y \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Define the event A as $Y > X$. (i) Compute the contact c. (ii) Find the conditional pdf of X and Y, given A. (ii) Find the conditional pdf of Y given A.

Solution: We first must find the value for c in $f_{X,Y}(x,y)$. We have,

$$\int_0^1 \int_0^1 cxy\,dx\,dy = 1$$

And so,

$$c\int_0^1 y(\left.\frac{x^2}{2}\right|_0^1)\,dy = c\int_0^1 \frac{y}{2}\,dy = 1$$

Which results in $c = 4$.
To find $f_{X,Y|A}(x, y|A)$, we first find $p(A)$.

$$P(A) = P(Y > X) = 4\int_0^1 \int_0^y xy\,dx\,dy = \frac{1}{2}$$

The conditional density, given A, is

$$\begin{aligned} f_{X,Y|A}(x,y|A) &= \begin{cases} \frac{f_{X,Y}(x,y)}{1/2}, & (x,y) \in A \equiv 0 \le x < y \le 1 \\ 0, & \text{otherwise.} \end{cases} \\ &= \begin{cases} 8xy, & 0 \le x < y \le 1 \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

The marginal pdf of Y given A, is obtained by integrating $f_{X,Y|A}(x, y|A)$ with respect to x.

$$\begin{aligned} f_{Y|A}(y|A) &= \int_0^y 8xy\,dx = 8y\left.\frac{x^2}{2}\right|_0^y \\ &= \begin{cases} 4y^3, & 0 \le y \le 1 \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

6. Let X and Y be independent and uniformly distributed over the interval $(0, a)$. The density functions are given by

$$\begin{aligned} f_X(x) &= 1/a \quad (0 \le x \le a) \\ f_Y(y) &= 1/a \quad (0 \le y \le a) \end{aligned}$$

and zero otherwise. Find the density function of $Z = X + Y$.

Solution: We know that the PDF of Z is given by

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)\,dx$$

Hence, the integrand is $1/a^2$ if $0 \le x \le a, 0 \le z - x \le a$ and zero otherwise. Two cases depending on the value of z:

(a) For $0 \le z \le a$, we find,

$$f_Z(z) = \int_0^z (1/a)^2\,dx = z/a^2$$

(b) For $a \le z \le 2a$, we find,

$$f_Z(z) = \int_{z-a}^a (1/a)^2\,dx = (2a - z)/a^2$$

7. At a crossing there is a traffic light showing alternately green and red light for a seconds. A car driver who arrives at random has to wait for a time period Z. Find the distribution of Z.

Solution: If the driver arrives during a green period, his waiting time is zero. Since the green and the red lights have the same durations, we have $P(Z = 0) = 1/2$. If the driver arrives during the first $a - z$ seconds of a red period, his waiting time is greater than z.
Hence for $0 < z < a$ we have

$$P(Z > z) = \frac{a - z}{2a} \quad \text{and} \quad P(Z \leq z) = 1 - \frac{a - z}{2a} = \frac{1}{2} + \frac{z}{2a}$$

Thus, the cdf function is:

$$F_Z(z) = \begin{cases} 0 & \text{if } z < 0 \\ 1/2 & \text{if } z = 0 \\ \frac{1}{2} + \frac{z}{2a} & \text{if } 0 < z < a \\ 1 & \text{if } z \geq a \end{cases}$$

8. Let (X, Y, Z) be a random point uniformly selected in the unit sphere. That is, their joint density is $3/4\pi$ when $x^2 + y^2 + z^2 \leq 1$, and 0 otherwise. Find the marginal densities of (i) (X, Y) and (ii) Z.

   Solution: To find the $f_{XY}(x, y)$, we integrate f_{XYZ} with respect to z in the range $[-\sqrt{1 - x^2 - y^2}, \sqrt{1 - x^2 - y^2}]$. This results in,

   $$f_{X,Y}(x, y) = \frac{3}{4\pi} \cdot 2\sqrt{1 - x^2 - y^2}, \text{ for } x^2 + y^2 \leq 1.$$

   To compute $f_Z(z)$, from the previous result we know that,

   $$f_{X,Z}(x, z) = \frac{3}{4\pi} \cdot 2\sqrt{1 - x^2 - z^2}, \text{ for } x^2 + z^2 \leq 1.$$

   We integrate $f_{X,Z}(x, z)$ with respect to x in the range, $[-\sqrt{1 - z^2}, \sqrt{1 - z^2}]$. One way to do this is to apply the change of variable

   $$\frac{x}{\sqrt{1 - z^2}} = \cos\phi$$

   This results in,

   $$f_Z(z) = \frac{3}{4} \cdot (1 - z^2).$$

9. Suppose $X_1, \ldots, X_n$ are independent and all have the cdf $F_X(x)$. Find the cdf of $Y = \max\{X_1, \ldots, X_N\}$, and $Z = \min\{X_1, \ldots, X_N\}$.

   Solution: (a) $F_Y(y) = P(X_1 \leq y, \cdots, X_n \leq y) = F_X(y)^n$.
   (b) $P(Z > z) = P(X_1 > z, \cdots, X_n > z) = (1 - F_X(z))^n$, so $F_Z(z) = 1 - (1 - F_X(z))^n$.

10. Suppose $X_1, \ldots, X_5$ are independent and all have the same distribution which is continuous. Show that $P(X_3 < X_5 < X_1 < X_4 < X_2) = 1/5!$.

    Solution: Let $X_1, \ldots, X_5$ be independent with distribution f. We use the fact that f is continuous to conclude that with probability one all the X's are distinct. This means that for a given set of $X_1, \ldots, X_5$ values, we can have 5! different permutation to rearrange the X values. Noting the independence of the X's, we conclude that these 5! different permutations all have the same probability. As these 5! different permutations are disjoint and cover all the possibilities for the ordering of X's, we conclude that each happen with probability $1/5!$.

11. Suppose X and Y are independent, X is uniform on $(0,1)$ and Y has the cdf $F(y)$. Show that $Z = X + Y$ has the density function $F(z) - F(z-1)$.

    Solution: Using the result of problem 4 with $f_X(x) = 1$, we obtain,

$$f_{X+Y}(z) = \int_0^1 f_Y(z-x)\, dx = F(z) - F(z-1)$$

12. Fill in the rest of the joint distribution given in the following table knowing that (i) $P(Y = 2|X = 0) = 1/4$, and (ii) X and Y are independent.

Y	$X = 0$	3	6
1	?	?	?
2	.1	.05	?

(i) $P(Y = 2|X = 0) = 1/4$ and $P(X = 0, Y = 2) = 0.1$ implies $P(X = 0) = 0.4$ and $P(X = 0, Y = 1) = 0.3$. Using the fact that X and Y are independent, we conclude now from $P(X = 0, Y = 2) = P(X = 0) \times P(Y = 2) = 0.1$ that $P(Y = 2) = 0.25$, and $P(X = 3, Y = 2) = P(X = 3) \times P(Y = 2) = 0.05$ gives that $P(X = 3) = 0.2$ and hence, $P(X = 6) = 1 - 0.4 - 0.2$. Thus the entire distribution is given by,

Y	$X = 0$	3	6
1	.3	.15	.3
2	.1	.05	.1

13. Suppose X and Y have joint density $f(x, y)$. Are X and Y independent if

$$\begin{array}{lll} (a) & f(x,y) & = & xe^{-x(1+y)} \text{ for } x, y \geq 0 \\ (b) & f(x,y) & = & 6xy^2 \text{ when } x, y \geq 0 \text{ and } x + y \leq 1 \\ (c) & f(x,y) & = & 2xy + x \text{ when } 0 < x < 1 \text{ and } 0 < y < 1 \\ (d) & f(x,y) & = & (x+y)^2 - (x-y)^2 \text{ when } 0 < x < 1 \text{ and } 0 < y < 1 \end{array}$$

In each case, $f(x, y) = 0$ otherwise.

Solution: (a) No. Since $f_X(x) = \int_0^\infty xe^{-x(1+y)}\, dy = e^{-x}$ for $x > 0$.

$$\begin{aligned} f_Y(y) &= \int_0^\infty xe^{-x(1+y)}\, dx \\ &= -\left.\frac{xe^{-x(1+y)}}{1+y}\right|_0^\infty + \int_0^\infty \frac{e^{-x(1+y)}}{1+y}\, dx = \frac{1}{(1+y)^2} \end{aligned}$$

for $y > 0$, and $f(x, y) \neq f_X(x)f_Y(y)$.

We could simply reach this conclusion by noticing that $f_{X,Y}(x, y)$ cannot be factored as the product of two functions, one only function of x and the other one only function of y.

(b) No. Since $\{(x, y) : f(x, y) > 0\}$ is not a rectangle. Note that $f_{X,Y}(x, y)$ can be factored as the product of two functions, one only function of x and the other one only function of y. However, this is not sufficient condition for independence.

Theorem: If $f(x, y)$ can be written as $g(x)h(y)$ then there is a constant c so that $f_X(x) = cg(x)$ and $f_Y(y) = h(y)/c$. It follows that $f(x, y) = f_X(x)f_Y(y)$ and hence X and Y are independent. Note that this requires the range of (x, y) to be a rectangle with one side corresponding to the range of x and the other side corresponding to the range of y.

(c) Yes by the above theorem since $f(x, y) = x(2y + 1)$.
(d) Yes by the above theorem since $f(x, y) = 2x \cdot 2y$.

14. Suppose $X_1, \ldots, X_n$ are independent and have distribution function $F(x)$. Find the joint density function of $Y = \max\{X_1, \ldots, X_n\}$ and $Z = \min\{X_1, \ldots, X_n\}$. Find the joint density of Y and Z.

Solution: We have,

$$P(Y \le y, Z \ge z) = P(z \le X_i \le y \text{ for all } i \le n) = (\int_z^y f(x)\, dx)^n = [F(y) - F(z)]^n.$$

and,

$$P(Y \le y, Z \ge z) + P(Y \le y, Z \le z) = P(Y \le y)$$

This is used to compute the joint distribution function $F(y, z)$ as follows:

$$F(y, z) = P(Y \le y, Z \le z) = P(Y \le y) - P(Y \le y, Z \ge z) = F(y)^n - [F(y) - F(z)]^n$$

The corresponding joint density function is equal to,

$$f(y, z) = \frac{\partial^2 F}{\partial y \partial z} = n(n-1) f(y) f(z) \left(\int_z^y f(x)\, dx \right)^{n-2}$$

15. Suppose $X_1, \ldots, X_n$ are independent and uniform on (0,1). Let $X^{(k)}$ be the kth smallest of the X_j. Show that the density of $X^{(k)}$ is given by

$$n \binom{n-1}{k-1} x^{k-1} (1-x)^{n-k}$$

Solution: This formula is easy to understand: There are n values of index j that we can pick such that the corresponding X_j is the kth smallest of the X's. The rest of the formula then gives the probability that exactly $k-1$ of the remaining $n-1$ variables are smaller than x.

Suppose $X_1, \ldots, X_n$ are independent and have density function f. Let $X^{(1)}$ be the smallest of the X_j, $X^{(2)}$ be the second smallest, and so on until $X^{(n)}$ is the largest. It can be shown that their joint density is given by

$$f(x^1, \ldots, x^n) = \begin{cases} n! f(x^1) \cdots f(x^n) & \text{if } x^1 < x^2 \cdots < x^n \quad (*) \\ 0 & \text{otherwise} \end{cases}$$

Hence, the density function of $X^{(k)}$ is obtained by integrating the above joint density with respect to the other variables:

$$\begin{aligned} n! \int_0^{x_k} \int_{x_1}^{x_k} \cdots \int_{x_{k-2}}^{x_k} \int_{x_k}^1 \cdots \int_{x_{n-1}}^1 & dx_n \cdots dx_{k+1}\, dx_{k-1} \cdots dx_2\, dx_1 \\ &= n! \tfrac{1}{(k-1)!} x_k^{k-1} \tfrac{1}{(n-k)!} (1 - x_k)^{n-k} \\ &= n \binom{n-1}{k-1} x_k^{k-1} (1 - x_k)^{n-k} \end{aligned}$$

To generalize the above result we show that if $X_1, \ldots, X_n$ are independent and have density f then the density of $X^{(k)}$ is given by

$$nf(x)\binom{n-1}{k-1}F(x)^{k-1}(1-F(x))^{n-k}$$

The density function of $X^{(k)}$ is obtained by integrating the (*) equation:

$$n! \cdot \int_{-\infty}^{x_k} \cdots \int_{x_{k-2}}^{x_k} \int_{x_k}^{\infty} \cdots \int_{x_{n-1}}^{\infty} \Pi_{i=1}^{n} f(x_i)\, dx_n \cdots dx_{k+1}\, dx_{k-1} \cdots dx_1$$

To evaluate this we prove by induction that

$$\int_a^b \cdots \int_{x_{k+i-1}}^{b} \Pi_{j=1}^{i} f(x_{k+j})\, dx_{k+i} \cdots dx_{k+1} = \frac{1}{i!}(F(b)-F(a))^i \quad (**)$$

This is clear if $i = 1$. If the formula is true for i then the integral for $i+1$ is

$$\begin{aligned} &\textstyle\int_a^b f(x_{k+1})\frac{1}{i!}(F(b)-F(x_{k+1}))^i\, dx_{k+1} \\ &= -\left.\frac{(F(b)-F(x_{k+1}))^{i+1}}{(i+1)!}\right|_a^b = \tfrac{1}{(i+1)!}(F(b)-F(a))^{i+1} \end{aligned}$$

and we have proved (**). Using (**) with $a = -\infty, b = x_k, i = k-1$ and then $a = x_k, b = \infty, i = n-k$ we have

$$\begin{aligned} g(x_k) &= n! f(x_k)\frac{1}{(k-1)!}F(x_k)^{k-1} \cdot \frac{1}{(n-k)!}(1-F(x_k))^{n-k} \\ &= nf(x_k)\binom{n-1}{k-1}F(x_k)^{k-1}(1-F(x_k))^{n-k} \end{aligned}$$

16. Suppose we take a die with 3 on three sides, 2 on two sides, and 1 on one side, roll it n times, and let X_i be the number of times side i appeared. Find the conditional distribution $P(X_2 = k|X_3 = m)$.

    Solution:

    $$\begin{aligned} P(X_3 = m) &= \binom{n}{m}\left(\frac{3}{6}\right)^m\left(1-\frac{3}{6}\right)^{n-m} \\ P(X_2 = k, X_3 = m) &= \frac{n!}{(n-k-m)!k!m!}(3/6)^m(2/6)^k(1/6)^{n-m-k} \\ P(X_2 = k|X_3 = m) &= \binom{n-m}{k}(2/3)^k(1/3)^{n-m-k} \end{aligned}$$

    that is, *binomial*$(n-m, 2/3)$.

17. Suppose $X_1, \ldots, X_m$ are independent and have a geometric distribution with parameter p. Find $P(X_1 = k|X_1 + \cdots X_m = n)$.

Solution: One can show that the sum of i.i.d random variables with geometric distribution has a Pascal distribution. Using this fact we have, $P(X_1 + \cdots + X_m = n) = \binom{n-1}{m-1} p^m (1-p)^{n-m}$, so:

$$\begin{aligned} P(X_1 = k | X_1 + \cdots + X_m = n) &= \frac{P(X_1 = k)P(X_2 + \cdots + X_m = n-k)}{P(X_1 + \cdots + X_m = n)} \\ &= \binom{n-1-k}{m-2} \Big/ \binom{n-1}{m-1} \end{aligned}$$

This is quite intuitive since among $\binom{n-1}{m-1}$ choices for the first $m-1$ successes there are $\binom{n-1-k}{m-2}$ with the first success occurring at the kth trial.

18. Suppose X and Y have joint density $f(x,y) = (1/2)e^{-y}$ when $y \geq 0$ and $-y \leq x \leq y$. Compute $P(X \leq 1 | Y = 3)$.

Solution:

$$\begin{aligned} f_Y(y) &= \int_{-y}^{y} \frac{1}{2} e^{-y}\, dx = y e^{-y} \text{ for } y \geq 0 \\ f_X(x|Y=y) &= (1/2)e^{-y}/ye^{-y} = 1/2y \text{ for } -y \leq x \leq y \\ P(X \leq 1 | Y = 3) &= \int_{-3}^{1} 1/6\, dx = 2/3 \end{aligned}$$

19. Jobs 1 and 2 must be completed before job 3 is begun. If the amount of time each task takes is independent and uniform on (2,4), find the density function for the amount of time T it takes to complete all three jobs.

Solution: Let T_i be the time required for job i, $S = \max\{T_1, T_2\}$, and $T = S + T_3$ be the total time. $P(S \leq s) = (s-2)^2/4$ when $2 \leq s \leq 4$ so S has density function $(s-2)/2$, when $2 \leq s \leq 4$. Since S and T_3 are independent, the pdf of T is the convolution of pdf's for S and T_3. So, we get:

$$f_T(t) = \begin{cases} (u-4)^2/8 & \text{if } 4 < u < 6 \\ \frac{1}{2}\left[1 - \frac{(u-6)^2}{4}\right] & \text{if } 6 < u < 8 \end{cases}$$

20. Suppose X has density function $x/2$ for $0 < x < 2$ and 0 otherwise. Find the density function of $Y = X(2-X)$.

Solution:

$$\begin{aligned} P(Y \geq y) &= P(X^2 - 2X + 1 \leq 1 - y) \\ &= P(|X-1| \leq \sqrt{1-y}) \\ &= F(1 + \sqrt{1-y}) - F(1 - \sqrt{1-y}) \end{aligned}$$

Differentiating we see that the density function of Y is (for $0 \leq y \leq 1$)

$$\frac{f(1-\sqrt{1-y}) + f(1+\sqrt{1-y})}{2\sqrt{1-y}} = 1/(2\sqrt{1-y})$$

21. Suppose X_1 and X_2 have joint density

$$f(x_1, x_2) = \begin{cases} 1 & \text{for } \ 0 < x_1, x_2 < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the density of $Y_1 = X_1/X_2$ and $Y_2 = X_1 X_2$.

Solution: The map $r(x_1, x_2) = (x_1/x_2, x_1 x_2)$ has the inverse

$$s(y_1, y_2) = \left[(y_1 y_2)^{1/2}, (y_2/y_1)^{1/2}\right]$$

The elements of the Jacobian are equal to: $D_{11} = \frac{1}{2} y_1^{-1/2} y_2^{1/2}, D_{12} = \frac{1}{2} y_1^{1/2} y_2^{-1/2}, D_{21} = -\frac{1}{2} y_1^{-3/2} y_2^{1/2}, D_{22} = \frac{1}{2} y_1^{-1/2} y_2^{-1/2}$ and the Jacobian is $D = \frac{1}{2} y_1^{-1}$. Thus

$$f_{Y_1,Y_2}(y_1, y_2) = 1/2y_1 \quad \text{for } 0 < y_1 y_2 < 1, 0 < y_2 < y_1$$

22. Let

$$\begin{aligned} f_{XY}(x, y) &= 1/\pi \quad \text{if } 0 \le x^2 + y^2 \le 1 \\ &= 0 \quad \text{elsewhere} \end{aligned}$$

Find the pdf of $r^2 = x^2 + y^2$ and $\theta = \tan^{-1} y/x$. Solve the same problem for the following joint pdf:

$$f_{XY}(x, y) = Ae^{-(x^2+y^2)}$$

We solve the second part of the problem which is more general. Consider,

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \tag{4}$$

where $\sigma^2 = 1/2$ and $A = 1/\pi$. For the polar coordinates we have,

$$\begin{aligned} R &= \sqrt{X^2 + Y^2} \\ \Theta &= \arctan(Y/X) \end{aligned} \tag{5}$$

This results in,

$$\mathbf{J}(x, y) = \begin{bmatrix} \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \\ \frac{-y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{bmatrix} \tag{6}$$

We have,

$$|\det[\mathbf{J}(x, y)]| = \frac{1}{\sqrt{x^2 + y^2}} = \frac{1}{r} \tag{7}$$

The set of equations,

$$\begin{cases} r = \sqrt{x^2 + y^2} \\ \theta = \arctan(y/x) \end{cases} \tag{8}$$

has only one solution given by,

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases} \tag{9}$$

Substituting these results, we obtain,

$$f_{R,\Theta}(r,\theta) = r f_{X,Y}(r\cos\theta, r\sin\theta) = \frac{r}{2\pi\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right) \tag{10}$$

The marginal pdf's for R, Θ are equal to,

$$\begin{aligned} f_\Theta(\theta) &= \int_0^\infty f_{R,\Theta}(r,\theta)dr \\ &= \frac{1}{2\pi}\int_0^\infty \frac{r}{\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right)dr \\ &= \frac{-1}{2\pi}\left[\exp\left(-\frac{r^2}{2\sigma^2}\right)\right]_0^\infty \\ &= \frac{1}{2\pi} \quad \Longrightarrow \quad \Theta \quad \text{has a uniform distribution in } [0,2\pi] \end{aligned} \tag{11}$$

and

$$f_R(r) = \int_0^{2\pi} f_{R,\Theta}(r,\theta)d\theta = \frac{r}{\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right), \quad r \ge 0 \tag{12}$$

Then,

$$f_R(r) = \begin{cases} \frac{r}{\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right), & r \ge 0 \\ 0 & r < 0 \end{cases} \tag{13}$$

This is known as the *Raleigh probability density function*. We also note that

$$f_{R,\Theta}(r,\theta) = f_R(r) f_\Theta(\theta). \tag{14}$$

This means that R and Θ are independent of each other.

23. Let

$$\begin{aligned} f_{XY}(x,y) &= e^{-(x+y)}, \quad x,y \ge 0 \\ &= 0 \quad \text{elsewhere} \end{aligned}$$

Find the pdf of $Z = X + Y$.

Solution: Noticing that X and Y are independent, the pdf of Z is the convolution of pdf's for X and Y. This results:

$$\begin{aligned} f_Z(z) &= ze^{-z}, \quad z \ge 0 \\ &= 0 \quad \text{elsewhere} \end{aligned}$$

24. Suppose a point (X,Y) is chosen at random in the unit circle $x^2 + y^2 \le 1$. Find the marginal density function of X.

Solution: The bivariate density function is uniform over the unit circle and hence has height $1/\pi$. The marginal density function for X is then

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}dy = \frac{2}{\pi}\sqrt{1-x^2}\ , \qquad -1 < x < 1\ .$$

25. Suppose X and Y have the joint density function $f(x, y)$. Are X and Y independent if:
   (i) $f(x, y) = xe^{-x(1+y)}$, $x, y > 0$.
   (ii) $f(x, y) = 6xy^2$ when $x, y \geq 0$ and $x + y < 1$.
   (iii) $f(x, y) = (x + y)^2 - (x - y)^2$ when $0 \leq x, y \leq 1$.

   Solution: (i)

$$f_X(x) = \int_0^\infty xe^{-x}e^{-xy}dy = e^{-x} \ , \quad x \geq 0$$

   and

$$f_Y(y) = \int_0^\infty xe^{-x(1+y)}dx = \frac{1}{(1+y)^2} \ , \quad y \geq 0$$

   and since $f(x, y) \neq f_X(x)f_Y(y)$ X and Y are not independent.
   (ii)

$$f_X(x) = \int_0^{1-x} 6xy^2dy = 2x(1-x)^3 \ , \quad 0 \leq x \leq 1$$

   and

$$f_Y(y) = \int_0^{1-y} 6xy^2dx = 3y^2(1-y)^2 \ , \quad 0 \leq y \leq 1$$

   and clearly X and Y are not independent.
   (iii) Notice that $f(x, y) = 4xy$ and so $f_X(x) = 2x$, $0 \leq x \leq 1$ and $f_Y(y) = 2y$, $0 \leq y \leq 1$ and X and Y are independent.

26. The bivariate random variable (X, Y) has the joint pdf:

$$f(x, y) = \begin{cases} 4x(1-y) & 0 \leq x, y \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

   (i) find the marginal density functions of X and Y and determine whether or not they are independent.
   (ii) Determine the probability that $Y > X^2$.

   Solution: (i) By straight forward computation the marginals are determined as

$$\begin{aligned} f_X(x) &= \int_0^1 4x(1-y)dy = 2x \ , \quad 0 \leq x \leq 1 \\ f_Y(y) &= \int_0^1 4x(1-y)dx = 2(1-y) \ , \quad 0 \leq y \leq 1 \end{aligned}$$

   and X and Y are independent.
   (ii)

$$P(Y > X^2) = \int_0^1 \left\{ \int_{x^2}^1 4x(1-y)dy \right\} dx = 1/3 \ .$$

27. (X, Y) is a bivariate random variable which is uniformly distributed over the triangle $x, y \geq 0$, $x+y \leq 1$. (i) Find the marginal pdf, mean and variance of X. (ii) Find the correlation coefficient ρ_{XY}.

   Solution: (i) Notice that the problem is symmetric in X and Y. The marginal of X is $f_X(x) = 2(1-x)$, $0 \leq x \leq 1$. The mean and variance of X are, respectively:

$$E(X) = \int_0^1 x \cdot 2(1-x)dx = 1/3 \ , \quad E(X^2) = \int_0^1 x^2 \cdot 2(1-x)dx = 1/6 \ \text{ and } \ V(X) = 1/18.$$

(ii) To compute the correlation coefficient of X and Y we need:

$$E(XY) = 2\int_0^1 x\left\{\int_0^{1-x} y dy\right\} dx = 1/12$$

and thus

$$\rho_{XY} = \frac{\frac{1}{12} - \frac{1}{3}\cdot\frac{1}{3}}{\sqrt{\frac{1}{18}\cdot\frac{1}{18}}} = -\frac{1}{2} \ .$$

28. Suppose that X and Y have the joint probability density function

$$f_{XY} = \begin{cases} (2m+6)x^m y & 0 < y < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Show that this is a probability density function and find $E(Y|X)$ and $V(Y|X)$.

Solution: The marginal density of X is

$$f_X(x) = \int_0^x (2m+6)x^m y dy = (m+3)x^{m+2} \ , \quad 0 \le x \le 1$$

and notice that

$$\int_0^1 (m+3)x^{m+2} dx = 1$$

and hence is a pdf. The conditional pdf is

$$f_{Y|X}(y|x) = \frac{2(m+3)x^m y}{(m+3)x^{m+2}} = \frac{2y}{x^2} \ , \quad 0 \le y \le x$$

and

$$E(Y|X=x) = \int_0^x y\cdot\frac{2y}{x^2} dy = \frac{2}{3}x \ , \quad E(Y^2|X=x) = \int_0^x y^2\cdot\frac{2y}{x^2} dy = \frac{1}{2}x^2$$

and $V(Y|X) = E(Y^2|X=x) - E^2(Y|X=x) = x^2/18$.

29. If X is a binomial random variable with parameters n and p and Y is a binomial random variable with parameters m and p, X and Y independent, and $Z = X+Y$, find $E(X|Z)$.

Solution:

$$\begin{aligned} P(X=j|Z=X+Y=l) &= \tfrac{P(X=j \text{ and } Y=l-j)}{P(X+Y=l)} = \tfrac{P(X=j)P(Y=l-j)}{P(X+Y=l)} \\ &= \tfrac{\binom{n}{j}\binom{m}{l-j}}{\binom{n+m}{l}} \end{aligned}$$

which is a hypergeometric distribution. The mean of this distribution is in the text and is $E(X|Z) = nl/(n+m)$.

30. If (X,Y) is a bivariate normal random variable with pdf

$$f_{XY}(x,y) = \frac{1}{\pi\sqrt{3}}\exp\{-\frac{2}{3}(x^2 - xy + y^2)\}$$

(i) Find the probability that $P(Y > 3|X = 2)$ and compare this to $P(Y > 3)$.

(ii) Find $E(Y|X)$.

Solution: (i) As with the previous problem, the five parameters of the bivariate density are found as

$$\mu_X = \mu_Y = 0 \quad \sigma_X = \sigma_Y = 1 \quad \rho_{XY} = 1/2 \ .$$

Consequently the conditional density function $f_{Y|X}(y|x) \sim N(\frac{1}{2}x, \frac{3}{4})$. Then

$$P(Y > 3|X = 2) = \int_3^\infty \frac{e^{-(y-1)^2/(3/2)}}{\sqrt{2\pi}\sqrt{3/4}} dy = 1 - \Phi(\frac{4}{\sqrt{3}}) \ .$$

Since $Y \sim N(0, 1)$,

$$P(Y > 3) = \int_3^\infty \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy = 1 - \Phi(3) \ .$$

(ii) $E(Y|X) = X/2$.

31. The time in minutes taken by person A to complete a certain task is assumed to be a random variable with an exponential pdf with parameter α_1. The time taken by person B is independent of that for person A and has an exponential pdf with parameter α_2 i.e.

$$f_X(x) = \alpha_1 e^{-\alpha_1 x} \ , \quad x \geq 0, \qquad f_Y(y) = \alpha_2 e^{-\alpha_2 y} \ , \quad y \geq 0$$

(i) What is the probability it takes A longer to complete the task than B?

(ii) If $\alpha_1 = \alpha_2$ what is the probability the tasks are finished within two minutes of each other?

Solution: (i) Let X be the time taken by person A and Y that taken by person B.

$$P(X > Y) = \int_0^\infty \left\{ \int_0^x \alpha_2 e^{-\alpha_2 y} dy \right\} \alpha_1 e^{-\alpha_1 x} dx$$

which can be computed to be

$$P(X > Y) = \frac{\alpha_2}{\alpha_1 + \alpha_2} \ .$$

(ii) It is desired to compute $P(|X - Y| < 2)$. Let $\alpha_1 = \alpha_2 = \alpha$. Consider

$$P(Y < X - 2) = \int_2^\infty \left\{ \int_0^{x-2} \alpha e^{-\alpha y} dy \right\} \alpha e^{-\alpha x} dx$$

which is readily evaluated to $e^{-2\alpha}/2$. The required probability is then $P(|X - Y| < 2) = 1 - 2(e^{-2\alpha}/2) = 1 - e^{-2\alpha}$.

32. Let X and Y be independent random variables, each with an exponential density function with parameter α (i.e. $f_X(x) = \alpha e^{-\alpha x} \ , \ x \geq 0$). Find the density function of $Z = X/(X + Y)$.

Solution:

$$\begin{aligned} F_Z(z) &= P(\tfrac{X}{X+Y} \leq z) = P(X \leq z(X+Y)) = P(X \leq \tfrac{z}{1-z} Y) \\ &= \int_0^\infty \left\{ \int_{(1-z)x/z}^\infty \alpha e^{-\alpha y} dy \right\} \alpha e^{-\alpha x} dx = \int_0^\infty e^{-\alpha(1-z)x/z} \alpha e^{-\alpha x} dx \\ &= \int_0^\infty \alpha e^{-\alpha x/z} dx = z \end{aligned}$$

and hence $f_Z(z) = 1 \ , \quad 0 \le z \le 1$.

33. Let X, Y and Z be independent and identically distributed random variables, each with an exponential pdf with parameter α. Find the probability density function of $W = X + Y + Z$.

    Solution: Let $U = X + Y$ then

    $$f_U(u) = \int_0^u f_X(u-y) f_Y(y) dy = \int_0^u \alpha e^{-\alpha(u-y)} \alpha e^{-\alpha(y)} dy = \alpha^2 u e^{-\alpha u} \ , \quad 0 \le u < \infty$$

    The process is repeated to find the density of W:

    $$f_W(w) = \int_0^w \alpha e^{-\alpha(w-v)} \alpha^2 v e^{-\alpha(v)} dv = \frac{\alpha^3}{2} w^2 e^{-\alpha w} \ , \quad w \ge 0$$

34. Let X and Y be independent random variables with density functions:

    $$f_X(x) = \alpha e^{-\alpha x} \ , \quad x \ge 0 \quad f_Y(y) = \alpha^n \frac{y^{n-1}}{(n-1)!} e^{-\alpha y} \ , \quad y \ge 0$$

    for n a positive integer. (i) Find the pdf of $U = X + Y$. (ii) Find the pdf of $W = X/(X+Y)$.

    Solution: We use the same technique as the previous two problems:

    $$f_U(u) = \int_0^u \alpha e^{-\alpha(u-y)} \frac{\alpha^n}{(n-1)!} y^{n-1} e^{-\alpha(y)} dy = \frac{\alpha^{n+1}}{n!} u^n e^{-\alpha u} \ .$$

    Notice that this implies that if $Z = X_1 + \cdots X_n$, where the X_i are i.i.d., each with a density $\alpha \exp(-\alpha x)$, then the pdf of Z is:

    $$f_Z(z) = \alpha^n \frac{z^{n-1}}{(n-1)!} e^{-\alpha z} \quad z \ge 0 \ .$$

35. Let X and Y be independent random variables with the pdf's:

    $$f_X(x) = e^{-x} \ , \quad x \ge 0, \quad f_Y(y) = e^{-y} \ , \quad y \ge 0 \ .$$

    Find the pdf of the random variable $Z = (X-Y)/(X+Y)$ and specify clearly its region of definition.

    Solution: Notice that Z takes values in the range $[-1, 1]$.

    $$\begin{aligned} F_Z(z) &= P(\tfrac{X-Y}{X+Y} \le z) = P((X-Y) \le z(X+Y)) = P((\tfrac{1-z}{1+z})X \le Y) \\ &= \int_0^\infty \left\{ \int_{\frac{1-z}{1+z}x}^\infty e^{-y} \right\} e^{-x} dx \\ &= \int_0^\infty e^{-\frac{1-z}{1+z}x} \cdot e^{-x} dx = \tfrac{1+z}{2} \ , \quad -1 \le z \le 1 \ . \end{aligned}$$

    Hence $f_Z(z) = 1/2 \ , \quad -1 \le z \le 1$ and zero elsewhere.

36. The random variables X and Y have the joint probability density function

$$f_{XY}(x,y) = xe^{-x(1+y)} \ , \quad 0 \le x, y < \infty \ .$$

Find the pdf of $Z = XY$.

Solution:

$$\begin{aligned} F_Z(z) &= P(Z \le z) = P(XY \le z) \\ &= \int_0^\infty \left\{ \int_0^{z/x} xe^{-x(1+y)} dy \right\} dx \\ &= \int_0^\infty xe^{-x} \left\{ -\tfrac{1}{x} e^{-xy} |_0^{z/x} \right\} dx = 1 - e^{-z} \end{aligned}$$

and consequently $f_Z(z) = dF_Z(z)/dz = e^{-z} \ , \quad z \ge 0$.

37. Let Y be a random variable with pdf $f_X(x) = e^{-x} \ , \quad x \ge 0$. Find the pdf of the random variable $Y = 1 - e^{-X} \ , X \ge 0$. Show that, in general, if X has the pdf $f_X(x)$ and cdf $F_X(x)$ then $Y = F_X(X)$ is uniformly distributed on $(0, 1)$.

Solution: Notice that Y takes values in the range $[0, 1]$ and $-ln(1 - y) > 0$.

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(1 - e^{-X} \le y) = P(X \le -ln(1 - y)) \\ &= \int_0^{-ln(1-y)} e^{-x} dx = 1 - e^{ln(1-y)} = y \end{aligned}$$

and consequently $f_Y(y) = dF_Y(y)/dy = 1 \ , \quad 0 \le y \le 1$.

In the general case, note that we are using the cdf as a montonic function, and that again Y takes values in $[0, 1]$:

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(F_X(X) \le y) = P(X \le F_X^{-1}(y)) \\ &= F_X(F_X^{-1}(y)) = y \ , \quad 0 \le y \le 1 \end{aligned}$$

where $F_X^{-1}(\cdot)$ is the inverse function of $F_X(\cdot)$. The result that Y is a uniformly distributed random variable on $[0, 1]$ follows.

38. Let $X, \ Y$ be independent normal random variables, each with zero mean and variance σ^2. Show that $Z = X + Y$ has a normal density $N(0, 2\sigma^2)$.

Solution: As in previous problems:

$$\begin{aligned} F_Z(z) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-(z-x)^2/(2\sigma^2)} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/(2\sigma^2)} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}(2x^2 - 2zx + z^2)} dx \\ &= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} e^{-\frac{2}{2\sigma^2}\{(x - z/2)^2 + z^2/4\}} dx \\ &= \frac{1}{\sqrt{4\pi}\sigma} e^{-z^2/4\sigma^2} \sim N(0, 2\sigma^2) \ . \end{aligned}$$

7 Properties of Expectation

7.1 Introduction

In this chapter we develop and exploit additional properties of expected values.

7.2 Expectation of Sums of Random Variables

Suppose that X and Y are random variables and g is a function of two variables. Then we have the following result.
Proposition: If X and Y have a joint probability mass function $p(x, y)$, then

$$E[g(X,Y)] = \sum_y \sum_x g(x,y)p(x,y)$$

If X and Y have a joint probability density function $f(x, y)$, then

$$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy$$

For an important application of this proposition, suppose that $E[X]$ and $E[Y]$ are both finite and let $g(X, Y) = X + Y$. Then, in the continuous case,

$$\begin{aligned} E[X+Y] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x+y)f(x,y)dxdy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xf(x,y)dxdy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yf(x,y)dxdy \\ &= \int_{-\infty}^{\infty} xf_X(x)dx + \int_{-\infty}^{\infty} yf_Y(y)dy \\ &= E[X] + E[Y] \end{aligned}$$

The same result holds in general.
Using this result and induction shows that, if $E[X_i]$ is finite for all $i = 1, \ldots, n$, then

$$E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n]$$

7.3 Covariance, Variance of Sums, and Correlations

Proposition: If X and Y are independent, then for any functions h and g,

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

Definition: The *covariance* between X and Y, denoted by $\mathrm{Cov}(X,Y)$, is defined by

$$\mathrm{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$$

Just as the expected value and the variance of a single random variable give us information about this random variable, so does the covariance between two random variables give us information about the relationship between the random variables.

Upon expanding the preceding definition, we see that

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y]$$

Note that if X and Y are independent then using the preceding proposition, it follows that $\mathrm{Cov}(X,Y) = 0$. However, the converse is not true.

Proposition:
(i) $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$

(ii) $\text{Cov}(X, X) = \text{Var}(X)$

(iii) $\text{Cov}(aX, Y) = a\text{Cov}(X, Y)$

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

It follows from parts (ii) and (iv) of the preceding proposition, upon taking $Y_j = X_j,\ j = 1, \ldots, n$, that

$$\begin{aligned}
\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) &= \mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j) \\
&= \sum_{i=1}^{n} \mathrm{Var}(X_i) \ + \ \sum_{i=1}^{n}\sum_{j\neq i} \mathrm{Cov}(X_i, X_j)
\end{aligned}$$

Since each pair of indices, $i, j, i \neq j$ appears twice in the double summation, the above is equivalent to the following:

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

If $X_1, \ldots, X_n$ are pairwise independent, the preceding equation reduces to

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

Definition: The *correlation* of two random variables X and Y, denoted by $\rho(X,Y)$, is defined, as long as $\mathrm{Var}(X)\mathrm{Var}(Y)$ is positive, by

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

It can be shown that

$$-1 \leq \rho(X,Y) \leq 1$$

In fact, $\rho(X,Y)=1$ implies that $Y=a+bX$, where $b=\sigma_X/\sigma_Y>0$ and $\rho(X,Y)=-1$ implies that $Y=a+bX$, where $b=-\sigma_X/\sigma_Y<0$. The reverse is also true, if $Y=a+bX$, then $\rho(X,Y)$ is either +1 or -1, depending on the sign of b.
The correlation coefficient is a measure of the degree of linearity between X and Y. A value of $\rho(X,Y)$ near +1 or -1 indicates a high degree of linearity between X and Y, whereas a value near 0 indicates a lack of such linearity. A positive value of $\rho(X,Y)$ indicates that Y tends to increase when X does, whereas a negative value indicates that Y tends to decrease when X increases. If $\rho(X,Y)=0$, then X and Y are said to be *uncorrelated*.

7.4 Conditional Expectation

7.4.1 definitions

We saw that for two jointly discrete random variables X and Y, given that $Y=y$, the conditional probability mass function is defined by

$$p_{X|Y}(x|y)=P\{X=x|Y=y\}=\frac{p(x,y)}{p_Y(y)}$$

So, the conditional expectation of X, given that $Y=y$, for all values of y such that $p_Y(y)>0$ by

$$\begin{aligned} E[X|Y=y] &= \sum_x xp\{X=x|Y=y\} \\ &= \sum_x xp_{X|Y}(x|y) \end{aligned}$$

Similarly for the case of continuous random variables, we have

$$E[X|Y=y]=\int_{-\infty}^{\infty} xf_{X|Y}(x|y)dx$$

where

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}$$

for all y such that $f_Y(y)>0$.

Similar to ordinary expectations we have

$$E[g(X)|Y=y] = \begin{cases} \sum_x g(x) p_{X|Y}(x|y) & \text{in the discrete case} \\ \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y) dx & \text{in the continuous case} \end{cases}$$

and

$$E\left[\sum_{i=1}^{n} X_i | Y=y\right] = \sum_{i=1}^{n} E[X_i|Y=y]$$

7.4.2 Computing Expectations by Conditioning

Let us denote by $E[X|Y]$ that function of a random variable Y whose value at $Y=y$ is $E[X|Y=y]$. Note that $E[X|Y]$ is itself a random variable. An extremely important property of conditioning property of conditional expectation is given by the following proposition.

Proposition:

$$E[X] = E[E[X|Y]]$$

One way to understand this equation is to interpret it as follows: To calculate $E[X]$, we may take a weighted average of the conditional expected value of X, given that $Y=y$, each of the terms $E[X|Y=y]$ being weighted by the probability of the event on which it is conditioned.

7.4.3 Computing probabilities by Conditioning

We can also use conditioning to find probabilities. To see this, let E denote an arbitrary event and define the indicator random variable X by

$$X = \begin{cases} 1 & \text{if } E \text{ occurs} \\ 0 & \text{if } E \text{ does not occur} \end{cases}$$

It follows from the definition of X that

$$E[X] = P(E)$$
$$E[X|Y=y] = P(E|Y=y) \quad \text{for any random variable } Y$$

Therefore,

$$\begin{aligned} P(E) &= \sum_y P(E|Y=y)P(Y=y) \quad \text{if } Y \text{ is discrete} \\ &= \int_{-\infty}^{\infty} P(E|Y=y)f_Y(y)dy \quad \text{if } Y \text{ is continuous} \end{aligned}$$

Note that if Y is a discrete random variable taking on one of the values $y_1, \ldots, y_n$, then, by defining the events $F_i, i = 1, \ldots, n$ by $F_i = \{Y = y_i\}$, this equation reduces to the familiar equation

$$P(E) = \sum_{i=1}^{n} P(E|F_i)P(F_i)$$

7.4.4 Conditional Variance

We can define the conditional variance of X given that $Y = y$ by

$$\text{Var}(X|Y) \equiv E[(X - E[X|Y])^2|Y]$$

and after simplification

$$\text{Var}(X|Y) = E[X^2|Y] - (E[X|Y])^2$$

There is a very useful relationship between $\text{Var}(X)$, the unconditional variance of X, and $\text{Var}(X|Y)$, the unconditional variance of X given Y.

$$\begin{aligned} E[\text{Var}(X|Y)] &= E[E[X^2|Y]] - E[(E[X|Y])^2] \\ &= E[X^2] - E[(E[X|Y])^2] \end{aligned}$$

Also, as $E[X] = E[E[X|Y]]$, we have

$$\text{Var}(E[X|Y]) = E[(E[X|Y])^2] - (E[X])^2$$

By adding these two equations we obtain the following proposition.
Proposition:

$$\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}(E[X|Y])$$

7.5 Conditional Expectation and Prediction

Sometimes a situation arises where the value of a random variable X is observed and then, based on the observed value, an attempt is made to predict the value of a second random variable Y. We would like to choose a function g so that $g(X)$ tends to be close to Y. One possible criterion for closeness is to choose g so as to minimize $E[(Y - g(X))^2]$. The best possible predictor of Y is $g(X) = E[Y|X]$.

Proposition:

$$E[(Y - g(X))^2] \geq E[(Y - E[Y|X])^2]$$

Linear Predictor: Sometimes the joint probability distribution function of X and Y is not completely known. If we know the means and variances of X and Y and correlation between them we can at least determine the best linear predictor of Y with respect to X. To obtain the best linear prediction of Y in respect to X, we need to choose a and b so as to minimize $E[(Y - (a + bX))^2]$.
Minimizing this equation over a and b, yields

$$b = \frac{E[XY] - E[X]E[Y]}{E[X^2] - (E[X])^2} = \frac{\text{Cov}(X, Y)}{\sigma_X^2} = \rho \frac{\sigma_Y}{\sigma_X}$$

$$a = E[Y] - bE[X] = E[Y] - \frac{\rho \sigma_Y E[X]}{\sigma_X}$$

7.6 Moment Generating Function

The *moment generating function* $M(t)$ of a random variable X is defined for all real values of t by

$$M(t) = E[e^{tX}] = \begin{cases} \sum_x e^{tx} p(x) & \text{if } X \text{ is discrete with mass function } p(x) \\ \int_{-\infty}^{\infty} e^{tx} f(x) dx & \text{if } X \text{ is continuous with density } f(x) \end{cases}$$

We call $M(t)$ the moment generating function because all of the moments of X can be obtained by successively differentiating $M(t)$ and then evaluating the result at $t = 0$. For example,

$$M'(0) = E[X]$$

and

$$M''(0) = E[X^2]$$

In general the nth derivative of $M(t)$ is given by

$$M^n(0) = E[X^n] \qquad n \geq 1$$

The moment generating function of some random variables are as follows.

Moment Generating Functions of Some Random Variables

If X is a binomial random variable with parameters n and p, then

$$M(t) = (pe^t + 1 - p)^n$$

If X is a Poisson random variable with parameter λ, then

$$M(t) = \exp\{\lambda(e^t - 1)\}$$

If X is an exponential random variable with parameter λ, then

$$M(t) = \frac{\lambda}{\lambda - t} \qquad \text{for t} < \lambda$$

If X is a standard normal random variable with parameters 0 and 1, then

$$M(t) = e^{t^2/2}$$

If X is a normal random variable with parameters μ and σ^2, then

$$M(t) = \exp\left\{\frac{\sigma^2 t^2}{2} + \mu t\right\}$$

Moment Generating function of the Sum of Independent Random Variables: An important property of moment generating function is the moment generating function of the sum of independent random variables equals to the product of the individual moment generating functions. If X and Y are independent, then

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

7.6.1 Joint Moment Generating Functions

It is also possible to define the joint moment generating function of two or more random variables.

$$M(t_1, \ldots, t_n) = E[e^{t_1X_1+\cdots+t_nX_n}]$$

The individual moment generating functions can be obtained from $M(t_1, \ldots, t_n)$ by letting all but one of the t_j be 0.

$$M_{X_i}(t) = M(0, \ldots, 0, t, 0, \ldots, 0)$$

where the t is in the ith place.
It can be proved that $M(t_1, \ldots, t_n)$ uniquely determines the joint distribution of $X_1, \ldots, X_n$.

Joint Moment Generating function of Independent Random Variables: If the n random variables are independent, then

$$M(t_1, \ldots, t_n) = M_{X_1}(t_1)\cdots M_{X_n}(t_n)$$

On the other hand, if this equation is satisfied, then the n random variables are independent.

7.7 Some Solved Problems

1. Let X and Y be two independent standard normal random variables (i.e. $N(0,1)$). Find: (i) $E(|X|)$ (ii) $E(X+Y)$ (iii) $V(X-Y)$ (iv) $E(\sqrt{X^2+Y^2})$.

   Solution: (i)

$$E(|X|) = 2\int_0^{\infty} x\frac{e^{-x^2/2}}{\sqrt{2\pi}}dx = \sqrt{\frac{2}{\pi}}$$

   where an elementary property of the normal integral has been used.

   (ii)

$$E(X+Y) = E(X) + E(Y) = 0$$

   (iii) Since $E(X-Y) = 0$ then

$$V(X-Y) = \int_{-\infty}^{\infty} (x-y)^2 \frac{e^{-(x^2+y^2)/2}}{2\pi} dxdy = 1+1 = 2$$

   (iv)

$$\begin{aligned} E(\sqrt{X^2+Y^2}) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x^2+y^2)^{1/2}\frac{e^{-(x^2+y^2)/2}}{2\pi}dxdy \\ &= \int_0^{\infty}\int_0^{2\pi} r^2 \frac{e^{-r^2/2}}{2\pi} drd\theta = \sqrt{\frac{\pi}{2}} \ . \end{aligned}$$

2. Suppose X and Y have the joint density function e^{-y} for $0 < x < y < \infty$. Find the least mean squares line of Y on X and $E(Y|X)$.

   Solution:

$$f_X(x) = \int_x^{\infty} e^{-y}dy = e^{-x} \ , \quad 0 \le x < \infty$$

   and

$$f_{Y|X}(y|x) = e^{-(y-x)} \ , \quad x < y < \infty$$

   and

$$E(Y|X=x) = \int_x^{\infty} ye^{-(y-x)}dy = x+1$$

   Any time the conditional expectation is linear it is also the least squares line. However, for completeness, we compute the least squares line as well.

$$\begin{aligned} E(X) &= \int_0^{\infty} xe^{-x}dx = 1 \quad E(X^2) = 2 \quad V(X) = 1 \\ E(Y) &= \int_0^{\infty} y \cdot ye^{-y}dy = 2 \quad E(Y^2) = 6 \quad V(Y) = 2 \end{aligned}$$

   and

$$E(XY) = \int_0^{\infty}\left\{\int_0^y xdx\right\} ye^{-y}dy = 3$$

   and thus

$$\rho_{XY} = \frac{3-1\cdot 2}{\sqrt{1}\sqrt{2}} = \frac{1}{\sqrt{2}}$$

   and thus the least mean squares line is $x+1$, as before.

3. Let (X, Y) be a bivariate random variable such that $E(X) = 0$, $E(Y) = 3$, $V(X) = 3$, $V(Y) = 5$ and $\rho_{XY} = 1/2$. Find the mean and variance of the random variable $Z = 2X - 3Y$.

   Solution: In general it is easy to show (do it!) that

   $$V(aX + bY) = a^2 V(X) + b^2 V(Y) + 2ab\rho_{XY}\sigma_X\sigma_Y \ .$$

   Using this formula it follows that

   $$\mu_Z = 2\mu_X - 3\mu_Y = -9 \quad \text{and} \quad V(Z) = 57 - 6\sqrt{15} \ .$$

4. It is known that the vertical and horizontal errors of a computer plotter, in standardized units, form a bivariate normal random variable with joint pdf

   $$f_{XY}(x, y) = \frac{3}{4\pi\sqrt{8}} \exp\{-\frac{9}{16}(x^2 - \frac{xy}{3} - \frac{4x}{3} + \frac{y^2}{4} - \frac{2y}{3} + \frac{4}{3})\}$$

   (i) Find $E(Y|X)$.

   (ii) For what value of the constant a will the random variables X and $Z = X + aY$ be independent?

   (iii) Find the probability that $|X - Y| < 1$. (Note: Sums of normal random variables are normal.)

   Solution: By comparing the form of the given bivariate normal density with that of the standard one, we can set up equations to be solved for the five parameters. The solutions are:

   $$\mu_X = 2/3 \ , \quad \mu_Y = 4/3 \ , \quad \sigma_X = 1 \ , \quad \sigma_Y = 2 \ , \quad \rho_{XY} = 1/3$$

   and these may be verified by substituting into the general form.
   (i) $E(Y|X) = \mu_Y + \rho_{XY}\frac{\sigma_Y}{\sigma_X}(x - \mu_X) = \frac{2}{3}X + \frac{8}{9}$.
   (ii) The value of a such that $E(XZ) = E(X)E(Z)$ is sought. By direct computation:

   $$E(XZ) = E(X)E(Z) = E(X(X + aY)) = E(X)(E(X) + aE(Y))$$

   or

   $$V(X) + a\rho_{XY}\sigma_X\sigma_Y = 0$$

   and from the parameters found, this implies that $a = -3/2$.
   (iii) From the parameters already found, let $Z = X - Y$ and note that

   $$\mu_Z = -2/3 \quad \text{and} \quad V(Z) = V(X - Y) = 11/3$$

   Thus $P(|X - Y| < 1) = P(|Z| < 1) = \Phi(\frac{5}{\sqrt{33}}) - \Phi(-\frac{1}{\sqrt{33}})$.

8 Limit Theorems

8.1 Introduction

The most important theoretical results in probability theory are limit theorems. Of these, the most important are those that are classified either under the heading laws of large numbers or under the heading central limit theorems. Usually, theorems are considered to be laws of large numbers if they are concerned with stating conditions under which the average of a sequence of random variables converges (in some sense) to the expected average. On the other hand, central limit theorems are concerned with determining conditions under which the sum of a large number of random variables has a probability distribution that is approximately normal.

8.2 Chebyshev's (Tchebychev) Inequality and the Weak Law of Large Numbers

Proposition, Markov's Inequality: If X is a random variable that takes only nonnegative values, then for any value $a > 0$,

$$P\{X \geq a\} \leq \frac{E[X]}{a}$$

Proposition, Chebyshev's Inequality: If X is a random variable with finite mean μ and variance σ^2, then for any value $k > 0$,

$$P\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$

The importance of Markov's and Chebyshev's inequalities is that they enable us to derive bounds on the probabilities when only the mean, or both the mean and the variance, of the probability distribution are known. Of course, if the actual distributions were known, then the desired probabilities could be exactly computed and we would not need to resort to bounds. Because Chebyshev's inequalities is valid for all distributions of the random variable X, we cannot expect the bound to be very close to the actual probability.
Proposition: If $\mathrm{Var}(X) = 0$, then

$$P\{X = E[X]\} = 1$$

In other words, the only random variables having variances equal to 0 are those that are constant with probability 1.

Theorem: The Weak Law of Large Numbers Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having finite mean $E[X_i] = \mu$. Then, for any $\epsilon > 0$,

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right\} \to 0 \qquad \text{as} \qquad n \to \infty$$

8.3 The Central Limit Theorem

The central limit theorem is one of the most remarkable results in probability theory. It states that the sum of a large number of independent random variables has a distribution that is approximately normal. Hence it not only provides simple method for computing approximate probabilities for sums of independent random variables, but it also helps explain the remarkable fact that the empirical frequencies of so many natural populations exhibit bell-shaped (normal) curves.
Theorem, The Central Limit Theorem: Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having μ and variance σ^2. Then, the distribution of

$$\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

tends to the standard normal as $n \to \infty$. That is, for $-\infty < a < \infty$,

$$P\left\{\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\right\} \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-x^2/2}dx \quad \text{as} \quad n \to \infty$$

Remark. Although central limit theorem only states that for each a,

$$P\left\{\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}\leq a\right\}\to\Phi(a)$$

it can be shown that the convergence is uniform in a. [We say that $f_n(a)\to f(a)$ uniformly in a, if for each $\epsilon>0$, there exists an N such that $|f_n(a)-f(a)|<\epsilon$ for all a whenever $n\geq N$.]

Theorem, Central Limit Theorem for Independent Random Variables:
Let $X_1, X_2, \ldots$ be a sequence of independent random variables having respective means and variances $\mu_i=E[X_i], \sigma_i^2=\text{Var}(X_i)$. If (a) the X_i are uniformly bounded; that is, if for some M, $P\{|X_i|<M\}=1$ for all i, and (b) $\sum_{i=1}^{\infty}\sigma_i^2=\infty$, then

$$P\left\{\frac{\sum_{i=1}^{n}(X_i-\mu_i)}{\sqrt{\sum_{i=1}^{n}\sigma_i^2}}\leq a\right\}\to\Phi(a)\quad\text{as}\quad n\to\infty$$

8.4 The Strong Law of Large Numbers

The *strong law of large numbers* states that the average of a sequence of independent random variables having a common distribution will, with probability 1, converge to the mean of that distribution.
Theorem, The Strong Law of Large Numbers: Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having a finite mean $\mu=E[X_i]$. Then, with probability 1,

$$\frac{X_1+X_2\cdots+X_n}{n}\to\mu\quad\text{as}\quad n\to\infty$$

As an application of the strong law of large numbers, suppose that a sequence of independent trials of some experiment is performed. Let E be a fixed event of the experiment and denote by $P(E)$ the probability that E occurs on any particular trial. Letting

$$X_i=\begin{cases}1 & \text{if } E \text{ occurs on th } i\text{th trial}\\ 0 & \text{if } E \text{ does not occur on th } i\text{th trial}\end{cases}$$

we have by the strong law of large numbers that with probability 1,

$$\frac{X_1+X_2\cdots+X_n}{n}\to E[X]=P(E)$$

<u>The difference between the Weak and the Strong Law of Large Numbers:</u> The weak law of large numbers states that for any specified large value n^*, $(X_1 + X_2 \cdots + X_{n^*})/n^*$ is likely to be near μ. However, it does not say that $(X_1 + X_2 \cdots + X_n)/n$ is bound to stay near μ for all values of n larger than n^*. Thus it leaves open the possibility that large values of $|(X_1 + X_2 \cdots + X_n)/n - \mu|$ can occur infinitely often (though at infrequent intervals). The strong law of large numbers shows that this cannot occur.

9 Stochastic Processes

Definitions

Stochastic Process: A *stochastic process* $X(t)$ consists of an experiment with a probability measure $P[\cdot]$ defined on a sample space S and a function that assigns a time function $x(t, s)$ to each outcome s in the sample space of the experiment.

Sample Function: A *sample function* $x(t, s)$ is the time function associated with the outcome s of an experiment.

Ensemble: The *ensemble* of a stochastic process is the set of all possible time functions that can result from an experiment.

9.1 Stochastic Process Examples

Example: Starting on January 1, we measure the noontime temperature at Network Airport every day for one year. This experiment generates a sequence of temperatures $C(1), C(2), \ldots, C(365)$. With respect to the kinds of averages of the stochastic processes, people make frequent reference to both ensemble averages such as "the average noontime temperature for February 19," and time averages, such as "average noontime temperatures for 1923."

9.2 Types of Stochastic Processes

Discrete Value and Continuous Value Processes: $X(t)$ is a *discrete value process* if the set of all possible value of $X(t)$ at all times t is a countable set S_X; otherwise $X(t)$ is a *continuous value process.*

Discrete Time and Continuous Time Process: The stochastic process $X(t)$ is a *discrete time process* if $X(t)$ is defined only for a set of time instants, $t_n = nT$, where T is a constant and n is an integer; otherwise $X(t)$ is a *continuous time process.*

So according to these definitions there are four types of stochastic processes:

- discrete time, discrete value process;
- discrete time, continuous value process;
- continuous time, discrete value process; and

- continuous time, continuous value process.

Random Sequence: A random sequence X_n is an ordered sequence of random variables $X_0, X_1, \ldots$

9.3 Independent, Identically Distributed Random Sequences

An independent identically distributed (iid) random sequence is a random sequence, X_n, in which $\ldots, X_{-2}, X_{-1}, X_0, X_1, X_2, \ldots$ are iid random variables. An iid random sequence occurs whenever we perform independent trials of an experiment at a constant rate.

9.4 Random Variables from Random Processes

Suppose we observe a stochastic process at a particular time instant t_1. In this case, each time we perform the experiment, we observe a sample function $x(t,s)$ and that sample function specifies the value of $x(t_1,s)$. Therefore, each $x(t_1,s)$ is a sample value of random variable. We use the notation $X(t_1)$ for this random variable. Like any other random variable , it has either a PDF $f_{X(t_1)}(x)$ or a PMF $p_{X(t_1)}(x)$.
Bernoulli Process: A Bernoulli process X_n with success probability p is an iid random sequence in which each X_n is a Bernoulli random variable such that $P\{X_n = 1\} = p = 1 - P\{X_n = 0\}$.

9.5 The Poisson Process

Counting Process: A stochastic process $N(t)$ is a counting process if for every sample function, $n(t,s) = 0$ for $t < 0$ and $n(t,s)$ is integer valued and nondecreasing with time.
We can think of $n(t)$ as counting the number of customers that arrive at a system during the interval $(0,t]$. We can use a Bernoulli process $X_1, X_2, \ldots$ to derive a simple counting process. In particular, consider a small time step of size Δ such that there is an arrival in the interval $(n\Delta, (n+1)\Delta]$ if and only if $X_n = 1$. For an arbitrary constant $\lambda > 0$, we can choose Δ small enough to ensure $\lambda\Delta < 1$. In this case, we choose the success probability of X_n to be $\lambda\Delta$. This implies that the number of arrivals N_m by time $T = m\Delta$ has the binomial PMF

$$p_{N_m}(n) = \begin{cases} \binom{m}{n} (\lambda T/m)^n (1 - \lambda T/m)^{m-n} & n = 0, 1, \ldots, m \\ 0 & \text{otherwise} \end{cases}$$

When $m \to \infty$, or equivalently as $\Delta \to 0$, the PMF of N_m becomes a Poisson random variable $N(T)$ with PMF

$$p_{N(T)}(n) = \begin{cases} (\lambda T)^n e^{-\lambda T}/n! & n = 0, 1, \ldots \\ 0 & \text{otherwise} \end{cases}$$

Poisson Process: A counting process $N(t)$ is a *Poisson process* of rate λ if

- The Number of arrivals in any interval $(t_0, t_1]$, $N(t_1) - N(t_0)$, is a Poisson random variable with expected value $\lambda(t_1 - t_0)$.
- For any pair of nonoverlapping intervals $(t_0, t_1]$ and $(t'_0, t'_1]$, the number of arrivals in each interval, $N(t_1) - N(t_0)$ and $N(t'_1) - N(t'_0)$ respectively, are independent random variables.

By the definition of a Poisson random variable, $M = N(t_1) - N(t_0)$ has PMF

$$p_M(m) = \begin{cases} \frac{[\lambda(t_1 - t_0)]^m}{m!} e^{-\lambda(t_1 - t_0)} & m = 0, 1, \ldots \\ 0 & \text{otherwise} \end{cases}$$

Theorem: For a Poisson process $N(t)$ of rate λ, the joint PMF of $N(t_1), \ldots, N(t_k), t_1 < t_2 < \ldots < t_k$, is

$$p_{N(t_1),\ldots,N(t_k)}(n_1, \ldots, n_k) = \begin{cases} \frac{\alpha_1^{n_1} e^{-\alpha_1}}{n_1!} \frac{\alpha_2^{n_2 - n_1} e^{-\alpha_2}}{(n_2 - n_1)!} \cdots \frac{\alpha_k^{n_k - n_{k-1}} e^{-\alpha_k}}{(n_k - n_{k-1})!} & 0 \leq n_1 \leq \cdots \leq n_k \\ 0 & \text{otherwise} \end{cases}$$

where $\alpha_i = \lambda(t_i - t_{i-1})$.
The independence of intervals property of the Poisson process must hold even for very small intervals. For example, the number of arrivals in $(t, t + \delta]$ must be independent of the arrival process over $[0, t]$ no matter how small we choose $\delta > 0$. Essentially, the probability of an arrival during any instant is independent of the past history of the process. In this sense, the Poisson process is memoryless.
This memoryless property can also be seen when we examine the times between arrivals. The random time between arrival $n - 1$ and arrival n is called the *nth interarrival time.* In addition, we call the time, X_1, of the first arrival the first interarrival time even though there is no previous arrival.
Theorem: For a Poisson process of rate λ, the interarrival times $X_1, X_2, \ldots$ are an iid random sequence with the exponential PDF

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

The memoryless property of the Poisson process can also be seen in the exponential interarrival times. Since $P\{X_n > x\} = e^{-\lambda x}$, the conditional probability that $X_n - x' > x$ given $X_n > x'$, is

$$P\{X_n - x' > x | X_n > x'\} = \frac{P\{X_n > x' + x, X_n > x'\}}{P\{X_n > x'\}} = e^{-\lambda x}$$

Theorem: A counting process with independent exponential interarrivals $X_1, X_2, \ldots$ with mean $E[X_i] = 1/\lambda$ is a Poisson process of rate λ

9.6 The Brownian Motion Process

The Poisson process is an example of a continuous time, discrete value stochastic process. Now we will examine Brownian motion, a continuous time, continuous value stochastic process.

Brownian Motion Process: A *Brownian motion process* $X(t)$ has the property that $X(0) = 0$ and for $\tau > 0, X(t+\tau) - X(t)$ is a Gaussian random variable with mean 0 and variance $\alpha\tau$ that is independent of $X(t')$ for all $t' \leq t$.

For a Brownian motion, we can view $X(t)$ as the position of a particle on a line. For a small time increment δ,

$$X(t+\delta) = X(t) + [X(t+\delta) - X(t)]$$

Although this expansion may seem trivial, by the definition of Brownian motion, the increment $Y_\delta = X(t+\delta) - X(t)$, is independent of $X(t)$ and is a Gaussian with mean zero and variance $\alpha\delta$. This property of the Brownian motion is called *independent increments*. Thus after a time step δ, the particle's position has moved by an amount Y_δ that is independent of the previous position $X(t)$.

The PDf of Y_δ is

$$P_{Y_\delta}(y) = \frac{1}{\sqrt{2\pi\alpha\delta}} e^{-\frac{y^2}{2\alpha\delta}}, -\infty < y < \infty$$

Theorem: For the Brownian motion process $X(t)$, the joint PDF of $X(t_1), \ldots, X(t_k)$ is

$$f_{X(t_1),\ldots,X(t_k)}(x_1, \ldots, x_k) = \prod_{n=1}^{k} \frac{1}{\sqrt{2\pi\alpha(t_n - t_{n-1})}} e^{-(x_n - x_{n-1})^2/[2\alpha(t_n - t_{n-1})]}$$

9.7 Expected Value and Correlation

The Expected Value of a Process: The *expected value* of a stochastic process $X(t)$ is the deterministic function

$$\mu_X(t) = E[X(t)]$$

Autocovariance: The *autocovariance* function of the stochastic process $X(t)$ is

$$C_X(t, \tau) = \text{Cov}[X(t), X(t + \tau)]$$

For $\tau = 0$, the definition implies $C_X(t, 0) = \text{Var}[X(t)]$.

Autocorrelation: The *autocorrelation* function of the stochastic process $X(t)$ is

$$R_X(t, \tau) = E[X(t)X(t + \tau)]$$

Theorem: The autocorrelation and autocovariance functions of a process $X(t)$ satisfy

$$C_X(t, \tau) = R_X(t, \tau) - \mu_X(t)\mu_X(t + \tau)$$

Autocovariance and Autocorrelation of a Random Sequence: The autocovariance function of the random X_n is

$$C_X[m, k] = \text{Cov}[X_m, X_{m+k}]$$

where m and k are integers. The autocorrelation function of the random sequence X_n is

$$R_X[m, k] = E[X_m X_{m+k}]$$

9.8 Stationary Processes

Stationary Process: A stochastic process $X(t)$ is *stationary* if and only if for all sets of time instants $t_1, \ldots, t_m$ and any time difference τ,

$$f_{X(t_1),\ldots,X(t_m)}(x_1, \ldots, x_m) = f_{X(t_1+\tau),\ldots,X(t_m+\tau)}(x_1, \ldots, x_m)$$

Stationary Random Sequence: A random sequence X_n is stationary if and only if for any finite sets of time instants $n_1, \ldots, n_m$ and any time difference k,

$$f_{X_{n_1},\ldots,X_{n_m}}(x_1,\ldots,x_m) = f_{X_{n_1+k},\ldots,X_{n_m+k}}(x_1,\ldots,x_m)$$

Theorem: For a stationary process $X(t)$, the expected value, the autocorrelation, and the autocovariance satisfy for all t:

$$\begin{aligned} \mu_X(t) &= \mu_X \\ R_X(t,\tau) &= R_X(0,\tau) = R_X(\tau) \\ C_X(t,\tau) &= R_X(\tau) - \mu_X^2 = C_X(\tau) \end{aligned}$$

Theorem: For a stationary random sequence X_n, the expected value, the autocorrelation, and the autocovariance satisfy for all m:

$$\begin{aligned} E[X_m] &= \mu_X \\ R_X[m,k] &= R_X[0,k] = R_X[k] \\ C_X[m,k] &= R_X[k] - \mu_X^2 = C_X[k] \end{aligned}$$

9.9 Wide Sense Stationary Random Processes

Wide Sense Stationary: $X(t)$ is a *wide sense stationary process* if and only if for all t,

$$\begin{aligned} E[X(t)] &= \mu_X \\ R_X(t,\tau) &= R_X(0,\tau) = R_X(\tau) \end{aligned}$$

X_n is a wide sense stationary random sequence if and only if for all n

$$\begin{aligned} E[X_n] &= \mu_X \\ R_X[n,k] &= R_X[0,k] = R_X[k] \end{aligned}$$

Theorem: For a wide sense stationary process $X(t)$, the autocorrelation function $R_X(\tau)$ satisfies

$$\begin{aligned} R_X(0) &\geq 0 \\ R_X(\tau) &= R_X(-\tau) \\ |R_X(\tau)| &\leq R_X(0) \end{aligned}$$

Theorem: If X_n is a wide sense stationary random sequence, the autocorrelation function $R_X[k]$ satisfies

$$\begin{aligned} R_X[0] &\geq 0 \\ R_X[k] &= R_X[-k] \\ |R_X[k]| &\leq R_X[0] \end{aligned}$$

Average Power: The average power of a wide sense stationary process $X(t)$ is

$$R_X(0) = E[X^2(t)]$$

10 Random Signal Processing

In this chapter, electrical signals are represented as sample functions of wide sense stationary random processes. We use this representation to describe the effects of linear filters.

10.1 Linear Filtering of a Random Process

We consider a linear time-invariant filter with impulse response $h(t)$. If the input is a deterministic signal $v(t)$, the output, $w(t)$, is the convolution,

$$w(t) = \int_{-\infty}^{\infty} h(u)v(t-u)du$$

Definition, Fourier Transform: Functions $g(t)$ and $G(f)$ are called a *Fourier Transform* pair if

$$G(f) = \int_{-\infty}^{\infty} g(t)e^{-j2\pi ft}dt \qquad g(t) = \int_{-\infty}^{\infty} G(f)e^{j2\pi ft}df$$

Definition, The Linear Time Invariant Filter Output Process: $X(t)$ is the input to a linear time invariant filter with impulse response $h(t)$, and $Y(t)$ is the output if all inputs to the filter are sample functions $X(t)$ and the outputs are sample functions of $Y(t)$. $Y(t)$ is related to the $X(t)$ by the convolution integral

$$Y(t) = \int_{-\infty}^{\infty} h(u)X(t-u)du = \int_{-\infty}^{\infty} h(t-u)X(t)du$$

Theorem: If the input to a linear time invariant filter with impulse response $h(t)$ is a wide sense stationary process $X(t)$, the output is a wide sense stationary process $Y(t)$ with mean value

$$\mu_Y = \mu_X \int_{-\infty}^{\infty} h(t)dt = \mu_X H(0)$$

and autocorrelation function

$$R_Y(\tau) = \int_{-\infty}^{\infty} h(u) \int_{-\infty}^{\infty} h(v)R_X(\tau+u-v)dvdu$$

10.2 Power Spectral Density

Definition, Power Spectral Density: For a wide sense stationary random process $X(t)$, the autocorrelation $R_X(\tau)$ and the power spectral density $S_X(f)$ are the Fourier transform pair

$$S_X(f) = \int_{-\infty}^{\infty} R_X(\tau)e^{-j2\pi f\tau}d\tau \qquad R_X(\tau) = \int_{-\infty}^{\infty} S_X(f)e^{j2\pi f\tau}df$$

Theorem: For a wide sense stationary random process $X(t)$, the power spectral density $S_X(f)$ has the following properties

$$\begin{aligned} &\text{(a)} \quad E[X^2(t)] = R_X(0) = \int_{-\infty}^{\infty} S_X(f)df \\ &\text{(b)} \quad S_X(-f) = S_X(f) \end{aligned}$$

Theorem: When a wide sense stationary stochastic process $X(t)$ is the input to a linear time invariant filter with frequency response $H(f)$, the power spectral density of the output $Y(t)$ is

$$S_Y(f) = |H(f)|^2 S_X(f)$$

Example: Suppose that $H(f)$ is a narrow, ideal bandpass filter of bandwidth B centered at frequency f_0. That is,

$$H(f) = \begin{cases} 1 & |f \pm f_0| \le B/2 \\ 0 & \text{otherwise} \end{cases}$$

We saw that $S_Y(f) = |H(f)|^2 S_X(f)$ and the average power of $Y(t)$ satisfies

$$E[Y^2(t)] = \int_{-\infty}^{\infty} S_Y(f)df = \int_{-f_0-B/2}^{-f_0+B/2} S_X(f)df + \int_{f_0-B/2}^{f_0+B/2} S_X(f)df$$

Since $S_X(f) = S_X(-f)$, when B is very small, we have

$$E[Y^2(t)] \approx 2BS_X(f_0)$$

Thus we see that $S_X(f_0)$ characterize the power per unit frequency of $X(t)$ at frequencies near f_0.

Theorem: For a wide sense stationary stochastic process $X(t)$, the power spectral density $S_X(f) \ge 0$ for all f.

10.3 Cross Correlations

Definition, Independent Processes: Stochastic processes $X(t)$ and $Y(t)$ are *independent* if for any collection of time samples, $t_1, \ldots, t_n$ and $t'_1, \ldots, t'_m$,

$$\begin{aligned} &f_{X(t_1),\ldots,X(t_n),Y(t'_1),\ldots,Y(t'_m)}(x_1, \ldots, x_n, y_1, \ldots, y_m) \\ &= f_{X(t_1),\ldots,X(t_n)}(x_1, \ldots, x_n) f_{Y(t'_1),\ldots,Y(t'_m)}(y_1, \ldots, y_m) \end{aligned}$$

Definition, Cross Correlation Function: The *cross correlation* of random processes $X(t)$ and $Y(t)$ is

$$R_{XY}(t, \tau) = E[X(t)Y(t + \tau)]$$

Definition, Jointly Wide sense Stationary Processes: The random processes $X(t)$ and $Y(t)$ are *jointly wide sense stationary* if $X(t)$ and $Y(t)$ are each wide sense stationary, and the cross correlation satisfies

$$R_{XY}(t, t + \tau) = R_{XY}(\tau)$$

Theorem: If $X(t)$ and $Y(t)$ are jointly wide sense stationary, then

$$R_{XY}(\tau) = R_{YX}(-\tau)$$

Definition, Cross Spectral Density: For jointly wide sense stationary random processes $X(t)$ and $Y(t)$, the Fourier transform of the cross correlation yields the *cross spectral density*

$$S_{XY}(f) = \int_{-\infty}^{\infty} R_{XY}(\tau) e^{-j2\pi f \tau} d\tau$$

Theorem: When a wide sense stationary process $X(t)$ is the input to a linear time invariant filter $h(t)$, the input-output cross correlation is

$$R_{XY}(t, t+\tau) = R_{XY}(\tau) = \int_{-\infty}^{\infty} h(u) R_X(\tau - u) du$$

Theorem: When a wide sense stationary process $X(t)$ is the input to a linear time invariant filter, the input $X(t)$ and the output $Y(t)$ are jointly wide sense stationary.

Theorem: When a wide sense stationary process $X(t)$ is the input to a linear time invariant filter $h(t)$, the autocorrelation of the output $Y(t)$ is

$$R_Y(\tau) = \int_{-\infty}^{\infty} h(-w) R_{XY}(\tau - w) dw$$

Theorem: Let $X(t)$ be a wide sense stationary input to a linear time invariant filter $H(f)$. The input $X(t)$ and output $Y(t)$ satisfy

$$S_{XY}(f) = H(f)S_X(f) \qquad S_Y(f) = H^*(f)S_{XY}(f)$$

10.4 Gaussian Processes

Definition, Gaussian Process: $X(t)$ is a Gaussian random process if the joint PDF of $X(t_1), \ldots, X(t_k)$ has the multivariate Gaussian density

$$f_{X(t_1),\ldots,X(t_k)} = \frac{1}{(2\pi)^{k/2}|\mathbf{C}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu_X)^T \mathbf{C}^{-1}(\mathbf{x}-\mu_X)}$$

Theorem: If $X(t)$ is a wide sense stationary Gaussian process, then $X(t)$ is a stationary Gaussian process.

Theorem: $X(t)$ is a Gaussian random process if $Y = \int_0^T g(t)X(t)dt$ is a Gaussian random variable for every $g(t)$ such that $E[Y^2] < \infty$.

Theorem: Passing a stationary Gaussian process $X(t)$ through a linear filter $h(t)$ yields as the output a Gaussian random process $Y(t)$ with mean as

$$\mu_Y = \mu_X \int_{-\infty}^{\infty} h(t)dt = \mu_X H(0)$$

and autocorrelation function

$$R_Y(\tau) = \int_{-\infty}^{\infty} h(u) \int_{-\infty}^{\infty} h(v) R_X(\tau + u - v) dv du$$

10.5 White Gaussian Noise Processes

In electrical engineering, Gaussian processes appear as models of noise voltages in resistors and models of receiver noise in communication systems. Noise is an unpredictable waveform that we model as a stationary Gaussian random process $W(t)$. Noise has no DC component, so that

$$E[W(t_1)] = \mu_W = 0$$

We assume that for any collection of distinct time instants $t_1, \ldots, t_k, W(t_1), \ldots, W(t_k)$ is a set of independent random variables. So, for $\tau \neq 0$,

$$R_W(\tau) = E[W(t)W(t+\tau)] = E[W(t)]E[W(t+\tau)] = 0$$

On the other hand with these assumptions,

$$S_W(f) = \int_{-\infty}^{\infty} R_W(\tau) e^{-j2\pi f\tau} d\tau$$

is a constant for all f. The constant is 0 unless $R_W(\tau) = \frac{N_0}{2}\delta(\tau)$. Therefore N_0 is the power per unit bandwidth of the white Gaussian noise stochastic process. This model does not conform to any signal that can be observed physically. Note that the average noise power is

$$E[W^2(t)] = R_W(0) = \int_{-\infty}^{\infty} \frac{N_0}{2} df = \infty$$

Passing the white noise process through a filter $h(t)$ yields a noise process

$$Y(t) = \int_0^t h(t-\tau) W(\tau) d\tau$$

Unlike the white process $W(t)$, the noise process $Y(t)$ does have finite average power.

11 Renewal Processes and Markov Chains

11.1 Continuous Time Markov Chains

Definition, Continuous Time Markov Chain: A continuous time Markov chain $\{X(t)|t \geq 0\}$ is a continuous time, discrete value random process such that for an infinitesimal time step of size Δ,

$$P[X(t+\Delta) = j|X(t) = i] = q_{ij}\Delta$$
$$P[X(t+\Delta) = i|X(t) = i] = 1 - \sum_{j \neq i} q_{ij}\Delta$$

Note that the above model assumes that only a single transition can occur in the small time Δ. The continuous time Markov chain is closely related to the Poisson process.

11.2 Birth-Death Processes and Queueing Systems

Definition, Birth-Death Process: A continuous time Markov chain is a *birth-death process* if the transition rates satisfy $q_{ij} = 0$ for $|i - j| > 1$.
Birth-Death processes earn their name because the state can represent the number in population. A transition from i to $i+1$ is a birth since the population increases by one. A transition from i to $i-1$ is a death in the population.
Queueing systems are often modelled as birth-death process in which the population consists of the customers in the system. For a Markov chain that represent a queue, we make use of some new terminology and notation. Specifically, the transition rate $q_{i,i-1}$ is denoted by μ_i and is called the *service rate* in state i since the transition from i to $i-1$ occurs only if a customer completes service and leaves the system. Similarly, $\lambda_i = q_{i,i+1}$ is called the arrival rate in state i since a transition from state i to state $i+1$ corresponds to the arrival of a customer.
Theorem: For a birth-death queue with arrivals rates λ_i and service rates μ_i, the stationary probabilities p_i satisfy

$$p_{i-1}\lambda_{i-1} = p_i\mu_i \qquad \sum_{i=0}^{\infty} p_i = 1$$

Theorem: For a birth-death queue with arrivals rates λ_i and service rates μ_i, let $\rho_i = \lambda_i/\mu_{i+1}$. The limiting state probabilities, if they exist, are

$$p_i = \frac{\prod_{j=1}^{i-1} \rho_j}{1 + \sum_{k=1}^{\infty} \prod_{j=1}^{k-1} \rho_j}$$