

Chapter 1 - Introduction to Probability

Several concepts of probability have evolved over the centuries. We discuss three different approaches:

- 1) The classical approach.
- 2) The relative frequency approach.
- 3) The axiomatic approach.

The Classical Approach

An experiment is performed, and outcomes are observed, some of which are equated to (or associated with) an event E . The probability of event E is computed *a priori* by counting the number of ways N_E that event E can occur and forming the ratio N_E/N , where N is the number of all possible outcomes. An important requirement here is that the experiment has equally likely outcomes. The classical approach suffers from several significant problems. However, in those problems where 1) it is impractical to determine the outcome probabilities by empirical methods, and 2) equally likely outcomes occur the classical approach is useful.

An example of the classical approach can be constructed by considering balls in an urn. Suppose there are 2 black and 3 white balls in an urn. Furthermore, suppose a ball is drawn at random from the urn. By the classical approach, the probability of obtaining a white ball is 3/5.

The Relative Frequency Approach

The relative frequency approach to defining the probability of an event E is to perform an experiment n times. The number of times that E appears is denoted by n_E . Then, the probability of E is given by

$$\mathbf{P}[E] = \lim_{n \rightarrow \infty} \frac{n_E}{n}. \quad (1-1)$$

There are some difficulties with this approach (*e.g.* you cannot do the experiment an infinite number of times). Despite the problems with this notion of probability, the relative

frequency concept is essential in applying the probability theory to the physical world.

Axiomatic Approach

The Axiomatic Approach is followed in most modern textbooks on probability. It is based on a branch of mathematics known as *measure theory*. The axiomatic approach has the notion of a *probability space* as its main component. Basically, a probability space consists of 1) a sample space, denoted as \mathcal{S} , 2) a collection of events, denoted as \mathcal{F} , and 3) a probability measure, denoted by \mathbf{P} . Without discussing the *measure-theoretic* aspects, this is the approach employed in this course. Before discussing $(\mathcal{S}, \mathcal{F}, \mathbf{P})$, we must review some elementary set theory.

Elementary Set Theory

A *set* is a collection of objects. These objects are called *elements* of the set. Usually, upper case, bold face, letters in italics font are used to denote sets (*i.e.*, \mathbf{A} , \mathbf{B} , \mathbf{C} , ...). Lower case letters in italics font are used to denote set elements (*i.e.*, a , b , c ...). The notation $a \in \mathbf{A}$ ($a \notin \mathbf{A}$) denotes that a is (is not) an element of \mathbf{A} . All sets are considered to be subsets of some *universal set*, denoted here as \mathcal{S} .

Set \mathbf{A} is a *subset* of set \mathbf{B} , denoted as $\mathbf{A} \subset \mathbf{B}$, if all elements in \mathbf{A} are elements of \mathbf{B} . The *empty* or *null* set is the set that contains no elements. It's denoted as $\{\emptyset\}$.

Transitivity Property

If $\mathbf{U} \subset \mathbf{B}$ and $\mathbf{B} \subset \mathbf{A}$ then $\mathbf{U} \subset \mathbf{A}$, a result known as the *Transitivity Property*.

Set Equality

$\mathbf{B} = \mathbf{A}$ is equivalent to the requirements that $\mathbf{B} \subset \mathbf{A}$ **and** $\mathbf{A} \subset \mathbf{B}$. Often, two sets are shown to be equal by showing that this requirement holds.

Unions

The *union* of sets \mathbf{A} and \mathbf{B} is a set containing all of the elements of \mathbf{A} plus all of the elements of \mathbf{B} (and no other elements). The union is denoted as $\mathbf{A} \cup \mathbf{B}$.

Union is *commutative*: $\mathbf{A} \cup \mathbf{B} = \mathbf{B} \cup \mathbf{A}$.

Union is *associative*: $(\mathbf{A} \cup \mathbf{B}) \cup \mathbf{C} = \mathbf{A} \cup (\mathbf{B} \cup \mathbf{C})$

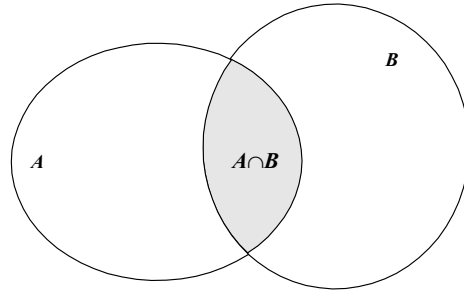


Figure 1-1: The intersection of sets A and B.

Intersection

The *intersection* of sets A and B is a set consisting of all elements common to both A and B . It is denoted as $A \cap B$. Figure 1-1 illustrates the concept of intersection.

Intersection is *commutative*: $A \cap B = B \cap A$.

Intersection is *associative*: $(A \cap B) \cap C = A \cap (B \cap C)$.

Intersection is *distributive* over unions: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Sets A and B are said to be *mutually exclusive* or *disjoint* if they have no common elements so that $A \cap B = \{\emptyset\}$.

Set Complementation

The *complement* of A is denoted as \bar{A} , and it is the set consisting of all elements of the universal set that are not in A . Note that $A \cup \bar{A} = S$, and $A \cap \bar{A} = \{\emptyset\}$.

Set Difference

The difference $A - B$ denotes a set consisting of all elements in A that are not in B . Often, $A - B$ is called the *complement of B relative to A*.

Shaded Area is $A - B$

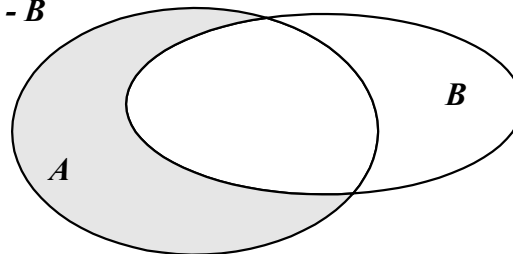


Figure 1-2: The difference between sets A and B.

De Morgan's Laws

$$\begin{aligned}
 1) \quad \overline{A \cup B} &= \overline{A} \cap \overline{B} \\
 2) \quad \overline{A \cap B} &= \overline{A} \cup \overline{B}
 \end{aligned}
 \tag{1-2}$$

More generally, if in a set identity, we replace all sets by their complements, all unions by intersections, and all intersections by unions, the identity is preserved. For example, apply this rule to the set identity $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ to obtain the result $\overline{A} \cup (\overline{B} \cap \overline{C}) = (\overline{A} \cup \overline{B}) \cap (\overline{A} \cup \overline{C})$.

Infinite Unions/Intersection of Sets

An infinite union of sets can be used to formulate questions concerning whether or not some specified item belongs to **one or more** sets that are part of an infinite collection of sets. Let A_i , $1 \leq i < \infty$, be a collection of sets. The union of the A_i is written/defined as

$$\bigcup_{i=1}^{\infty} A_i \equiv \{\omega : \omega \in A_n \text{ for some } n, 1 \leq n < \infty\}.
 \tag{1-3}$$

Equivalently, $\omega \in \bigcup_{i=1}^{\infty} A_i$ if, and only if, there is **at least** one integer n for which $\omega \in A_n$. Of course, ω may be in more than one set belonging to the infinite collection.

An infinite intersection of sets can be used to formulate questions concerning whether or not some specified item belongs to **all** sets belonging to an infinite collection of sets. Let A_i , $1 \leq i < \infty$, be a collection of sets. The intersection of the A_i is written/defined as

$$\bigcap_{i=1}^{\infty} A_i \equiv \{\omega : \omega \in A_n \text{ for all } n, 1 \leq n < \infty\} = \{\omega : \omega \in A_n, 1 \leq n < \infty\}.
 \tag{1-4}$$

Equivalently, $\omega \in \bigcap_{i=1}^{\infty} A_i$ if, and only if, $\omega \in A_n$ for **all** n .

σ - Algebra of Sets

Consider an arbitrary set \mathcal{S} of objects. In general, there are many ways to form subsets of \mathcal{S} . Set \mathcal{F} is said to be a σ -algebra of subsets of \mathcal{S} (often, we just say σ -algebra; the phrase “of subsets of \mathcal{S} ” is understood) if

- 1) \mathcal{F} is a set of subsets of \mathcal{S} ,
- 2) If $A \in \mathcal{F}$ then $\bar{A} \in \mathcal{F}$ (i.e. \mathcal{F} is closed under complementation)
- 3) $\{\emptyset\} \in \mathcal{F}$ and $\mathcal{S} \in \mathcal{F}$, and
- 4) If $A_i \in \mathcal{F}$, $1 \leq i < \infty$, then

$\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ (i.e. \mathcal{F} is closed under countable unions).

These four properties can be used to show that if $A_i \in \mathcal{F}$, $1 \leq i < \infty$, then

$$\bigcap_{i=1}^{\infty} A_i \in \mathcal{F},$$

that is, \mathcal{F} is closed under countable intersections. Some examples of σ -algebras follow.

Example 1-1: For $\mathcal{S} = \{H, T\}$ then $\mathcal{F} = \{\{\emptyset\}, \{H, T\}, \{H\}, \{T\}\}$ is a σ -algebra.

Example 1-2: All possible subsets of \mathcal{S} constitute a σ -algebra. This is the largest σ -algebra that can be formed from subsets of \mathcal{S} . We define $\mathcal{F}_L \equiv \{\text{all possible subsets of } \mathcal{S}\}$ to be the σ -algebra comprised of all possible subsets of \mathcal{S} . Often, \mathcal{F}_L is called the *Power Set* for \mathcal{S} .

Example 1-3: $\{\{\emptyset\}, \mathcal{S}\}$ is the smallest σ -algebra that can be formed from subsets of \mathcal{S} .

Intersection and Unions of σ -algebras

The intersection of σ -algebras is a σ -algebra, a conclusion that follows directly from the basic definition.

Example 1-4: Let \mathcal{F}_1 and \mathcal{F}_2 be σ -algebras of subsets of \mathcal{S} . Then the intersection $\mathcal{F}_1 \cap \mathcal{F}_2$ is a σ -algebra. More generally, let I denote an index set and $\mathcal{F}_k, k \in I$, be a collection of σ -algebras. Then the intersection

$$\bigcap_{k \in I} \mathcal{F}_k$$

is a σ -algebra.

Example 1-5: Let A be a non-empty proper subset of \mathcal{S} . Then $\sigma(A) \equiv \{ \{\emptyset\}, \mathcal{S}, A, \bar{A} \}$ is the smallest σ -algebra that contains A . Here, \bar{A} denotes the complement of A .

In general, a union of σ -algebras is not a σ -algebra. A simple counter example that establishes this can be constructed by using Example 1-5.

Example 1-5 can be generalized to produce the smallest σ -algebra that contains n sets A_1, \dots, A_n . (In Examples 1-5 and 1-6, can you spot a general construction method for generating the smallest σ -algebra that contains a given finite collection \mathcal{C} of sets?).

Example 1-6: Let A_1 and A_2 be non-empty proper subsets of \mathcal{S} . We construct $\sigma(\{A_1, A_2\})$, the smallest σ -algebra that contains the sets A_1 and A_2 . Consider the “disjoint pieces” $A_1 \cap A_2, \bar{A}_1 \cap A_2, A_1 \cap \bar{A}_2$ and $\bar{A}_1 \cap \bar{A}_2$ of \mathcal{S} that are illustrated by Figure 1-3. Let \mathcal{G} denote all possible unions (i.e., all unions taken k at a time, $0 \leq k \leq 4$) of these “disjoint pieces”. Note that \mathcal{G} is a σ -

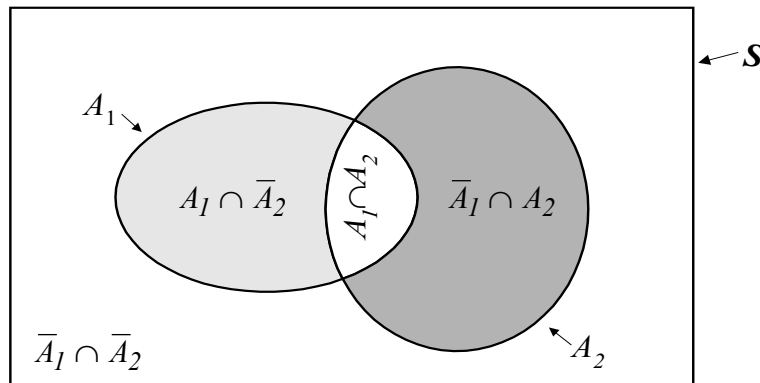


Figure 1-3: “Disjoint pieces” $A_1 \cap A_2, A_1 \cap \bar{A}_2, \bar{A}_1 \cap A_2, \bar{A}_1 \cap \bar{A}_2$ of \mathcal{S} .

algebra that contains A_1 and A_2 . Furthermore, if \mathcal{F} is any σ -algebra that contains A_1 and A_2 , then $\mathcal{G} \subset \mathcal{F}$. Therefore, \mathcal{G} is in the intersection of all σ -algebras that contain A_1 and A_2 so that $\sigma(\{A_1, A_2\}) = \mathcal{G}$. There are four “disjoint pieces”, and each “disjoint piece” may, or may not, be in a given set of \mathcal{G} . Hence, $\sigma(\{A_1, A_2\}) = \mathcal{G}$ will contain $2^4 = 16$ sets, assuming that all of the “disjoint pieces” are non-empty.

This construction technique can be generalized easily to n events A_1, \dots, A_n . The minimal σ -algebra $\sigma(\{A_1, \dots, A_n\})$ consists of the collection \mathcal{G} of all possible unions of sets of the form $C_1 \cap C_2 \cap \dots \cap C_n$, where each C_k is either A_k or \bar{A}_k (these correspond to the “disjoint pieces” used above). Example 1-5 corresponds to the case $n = 1$, and Example 1-6 for the case $n = 2$. Note that \mathcal{G} is a σ -algebra that contains A_1, \dots, A_n . Furthermore, \mathcal{G} must be in every σ -algebra that contains A_1, \dots, A_n . Hence, $\sigma(\{A_1, \dots, A_n\}) = \mathcal{G}$. Now, there are 2^n “disjoint pieces” of the form $C_1 \cap C_2 \cap \dots \cap C_n$, and each “disjoint piece” may, or may not, be in any given set in \mathcal{G} . Hence, $\sigma(\{A_1, \dots, A_n\}) = \mathcal{G}$, contains 2^{2^n} sets, assuming that each “disjoint piece” $C_1 \cap C_2 \cap \dots \cap C_n$ is non-empty (otherwise, there are fewer than 2^{2^n} sets in $\sigma(\{A_1, \dots, A_n\})$).

Example 1-7: Let $A_1, A_2, \dots, A_n, \dots$ be a disjoint partition of \mathcal{S} . By this we mean that

$$\mathcal{S} = \bigcup_{k=1}^{\infty} A_k, \quad A_j \cap A_k = \{\emptyset\} \text{ if } j \neq k.$$

By Example 1-6, note that it is possible to represent each element of $\sigma(\{A_1, A_2, \dots, A_n, \dots\})$ as a union of sets taken from the collection $\{A_1, A_2, \dots, A_n, \dots\}$.

σ -Algebra Generated by Collection of Subsets

Let \mathcal{C} be any non-empty collection of subsets of \mathcal{S} . In general, \mathcal{C} is not a σ -algebra. Every subset of \mathcal{C} is in the σ -algebra $\mathcal{F}_{\mathcal{L}} \equiv \{\text{all possible subsets of } \mathcal{S}\}$; we say \mathcal{C} is in σ -algebra $\mathcal{F}_{\mathcal{L}}$ and write $\mathcal{C} \subset \mathcal{F}_{\mathcal{L}}$. In general, there may be many σ -algebras that contain \mathcal{C} . The intersection of all σ -algebras that contain \mathcal{C} is the *smallest* σ -algebra that contains \mathcal{C} , and this smallest σ -algebra is denoted as $\sigma(\mathcal{C})$. We say that $\sigma(\mathcal{C})$ is the σ -algebra *generated* by \mathcal{C} . Examples 1-5

and 1-6 illustrate the construction of $\sigma(\{A_1, A_2, \dots, A_n\})$.

Example 1-8: Let $S = \mathcal{R}$, the real line. Let \mathcal{C} be the set consisting of all open intervals of \mathcal{R} (\mathcal{C} contains all intervals (a, b) , $a \in \mathcal{R}$ and $b \in \mathcal{R}$). \mathcal{C} is *not* a σ -algebra of $S = \mathcal{R}$. To see this, consider the identity

$$[-1, 1] = \bigcap_{n=1}^{\infty} \left(-1 - \frac{1}{n}, 1 + \frac{1}{n}\right).$$

Hence, the collection of open intervals of \mathcal{R} is *not* closed under countable intersections; the collection of all open intervals of \mathcal{R} is *not* a σ -algebra of $S = \mathcal{R}$.

Example 1-9: As in the previous example, let $S = \mathcal{R}$, the real line, and \mathcal{C} be the set consisting of all open intervals of \mathcal{R} . While \mathcal{C} is not a σ -algebra itself (see previous example), the smallest σ -algebra that contains \mathcal{C} (*i.e.*, the σ -algebra *generated* by \mathcal{C}) is called the *Borel σ -algebra*, and it is denoted as \mathcal{B} in what follows. σ -algebra \mathcal{B} plays an important role in the theory of probability. Obviously, \mathcal{B} contain the open intervals, but it contains much more. For example, all half open intervals $(a, b]$ are in \mathcal{B} since

$$(a, b] = \bigcap_{n=1}^{\infty} (a, b + 1/n) \in \mathcal{B}. \quad (1-5)$$

Using similar reasoning, it is easy to show that all closed intervals $[a, b]$ are in \mathcal{B} .

Example 1-10: Let \mathcal{F} be a σ -algebra of subsets of S . Suppose that $B \in \mathcal{F}$. It is easy to show that

$$\mathcal{G} \equiv [A \cap B : A \in \mathcal{F}] \quad (1-6)$$

is a σ -algebra of subsets of B . To accomplish this, first show that \mathcal{G} contains $\{\emptyset\}$ and B . Next, show that if $A \in \mathcal{G}$ then $B - A \in \mathcal{G}$ (*i.e.*, the complement of A *relative to* B must be in \mathcal{G}).

Finally, show that \mathcal{G} is closed under countable intersections; that is, if $A_k \in \mathcal{G}$, $1 \leq k < \infty$, then

$$\bigcap_{k=1}^{\infty} A_k \in \mathcal{G}. \quad (1-7)$$

Often, \mathcal{G} is called the *trace of \mathcal{F} on B*. Note that \mathcal{G} **will not** be a σ -algebra of subsets of \mathcal{S} . However, $\mathcal{G} \subset \mathcal{F}$. \mathcal{G} is an example of a *sub σ -algebra* of \mathcal{F} .

Probability Space (\mathcal{S} , \mathcal{F} , \mathbf{P})

A *probability space* (\mathcal{S} , \mathcal{F} , \mathbf{P}) consists of a *sample space* \mathcal{S} , a set \mathcal{F} of permissible events and a *probability measure* \mathbf{P} . These three quantities are described in what follows.

Sample Space \mathcal{S}

The *sample space* \mathcal{S} is the set of elementary outcomes of an experiment. For example, the experiment of tossing a single coin has the sample space $\mathcal{S} = \{ \text{heads, tails} \}$, a simple, countable set. The experiment of measuring a random voltage might use $\mathcal{S} = \{v : -\infty < v < \infty\}$, an infinite, non-countable set.

Set \mathcal{F} of permissible Events

In what follows, the collection (*i.e.*, set) of permissible events is denoted as \mathcal{F} . The collection of events \mathcal{F} **must** be a *σ -algebra* of subsets of \mathcal{S} , but **not every** *σ -algebra* can be the collection of events. For a *σ -algebra* to qualify as a set of permissible events, it must be possible to assign probabilities (this is the job of \mathbf{P} discussed next) to the sets/events in the *σ -algebra* without violating the *axioms of probability* discussed below.

Probability Measure

A *probability* must be assigned to every event (element of \mathcal{F}). To accomplish this, we use a set function \mathbf{P} that maps events in \mathcal{F} into $[0,1]$ ($\mathbf{P} : \mathcal{F} \rightarrow [0,1]$). *Probability measure* \mathbf{P} must satisfy

$$1) 0 \leq \mathbf{P}(A) \leq 1 \text{ for every } A \in \mathcal{F}$$

2) If $A_i \in \mathcal{F}$, $1 \leq i < \infty$, is any countable, *mutually exclusive* sequence (i.e., $A_i \cap A_j = \{\emptyset\}$ for $i \neq j$) of events then

$$\mathbf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbf{P}(A_i) \quad (\text{we say that } \mathbf{P} \text{ must be } \textit{countably additive}) \quad (1-8)$$

$$3) \mathbf{P}(\mathcal{S}) = 1 \quad (1-9)$$

Conditions 1) through 3) are called the *Axioms of Probability*. Now, we can say that the permissible set of events \mathcal{F} can be *any* σ -algebra for which there exists a \mathbf{P} that satisfies the *Axioms of Probability*. As it turns out, for some sample spaces, there are some σ -algebras that cannot serve as a set of permissible events because there is no corresponding \mathbf{P} function that satisfies the *Axioms of Probability*.

In many problems, it might be desirable to let all subsets of \mathcal{S} be events. That is, it might be desirable to let "everything" be an event (i.e., let the set of events be the σ -algebra $\mathcal{F}_{\mathcal{L}}$ that is discussed in Example 1-2 above). However, **in general**, it is not possible to do this because a \mathbf{P} function may not exist that satisfies the *Axioms of Probability*.

A **special case** deserves to be mentioned. *If* \mathcal{S} is *countable* (i.e. there exists a 1-1 correspondence between the elements of \mathcal{S} and the integers), then \mathcal{F} can be taken as the σ -algebra consisting of all possible subsets of \mathcal{S} (i.e., let $\mathcal{F} = \mathcal{F}_{\mathcal{L}}$, the largest σ -algebra). That is, if \mathcal{S} is countable, it is possible to assign probabilities (without violating the *Axioms*) to the elements of $\mathcal{F}_{\mathcal{L}}$ in this case. But, in the more general case where \mathcal{S} is not countable, to avoid violating the *Axioms of Probability*, there may be subsets of \mathcal{S} that cannot be events; these subsets must be excluded from \mathcal{F} .

As it turns out, if \mathcal{S} is the real line (a non-countable sample space), then the σ -algebra $\mathcal{F}_{\mathcal{L}}$ of all possible sets (of real numbers) contains "too many" sets. In this case, it is not possible to obtain a \mathbf{P} that satisfies the *Axioms of Probability*, and $\mathcal{F}_{\mathcal{L}}$ cannot serve as the set of events.

Instead, for \mathcal{S} equal to the real line, the Borel σ -algebra \mathcal{B} , discussed in Example 1-9, is usually chosen (it is very common to do this in applications). It is possible to assign probabilities to Borel sets without violating the Axioms of Probability. The Borel sets are assigned probabilities as shown by the following example.

Example 1-11: Many important applications employ a probability space $(\mathcal{S}, \mathcal{F}, \mathbf{P})$ where \mathcal{S} is the set \mathcal{R} of real numbers, and $\mathcal{F} = \mathcal{B}$ is the Borel σ -algebra (see Example 1-9). The probability measure \mathbf{P} is defined in terms of a *density function* $f(x)$. Density $f(x)$ can be any integrable function that satisfies

$$1) f(x) \geq 0 \text{ for all } x \in \mathcal{S} = \mathcal{R},$$

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

Then, probability measure \mathbf{P} is defined by

$$\mathbf{P}(B) = \int_B f(x)dx, \quad B \in \mathcal{F} = \mathcal{B}. \quad (1-10)$$

As defined above, the notion of \mathcal{F} , the set of possible events, is abstract. However, in *most* applications, one encounters only a few general types of \mathcal{F} . In *most* applications, \mathcal{S} is either countable, the real line $\mathcal{R} = (-\infty, \infty)$, or an interval (*i.e.*, $\mathcal{S} = (a,b)$). These cases are discussed briefly in what follows.

Many applications involve countable sample spaces. For *most* of these cases, \mathcal{F} is taken as \mathcal{F}_L , the set of all possible subsets of \mathcal{S} . To events in \mathcal{F}_L , probabilities are assigned in an application-specific, intuitive manner.

On the other hand, many applications use the real line $\mathcal{S} = \mathcal{R} = (-\infty, \infty)$, a non-countable set. For these cases, it is very common to use an \mathcal{F} and \mathbf{P} as discussed in Example 1-11. An

identical approach is used when \mathcal{S} is an interval of the real line.

Implications of the *Axiom of Probabilities*

A number of conclusions can be reached from considering the above-listed *Axioms of Probability*.

1) The probability of the impossible event $[\emptyset]$ is zero.

Proof: Note that $A \cap [\emptyset] = [\emptyset]$ and $A \cup [\emptyset] = A$. Hence, $\mathbf{P}(A) = \mathbf{P}(A \cup [\emptyset]) = \mathbf{P}(A) + \mathbf{P}([\emptyset])$.

Conclude from this that $\mathbf{P}([\emptyset]) = 0$.

2) For any event A we have $\mathbf{P}(A) = 1 - \mathbf{P}(\bar{A})$.

Proof: $A \cup \bar{A} = \mathcal{S}$ and $A \cap \bar{A} = [\emptyset]$. Hence,

$$1 = \mathbf{P}(\mathcal{S}) = \mathbf{P}(A \cup \bar{A}) = \mathbf{P}(A) + \mathbf{P}(\bar{A}),$$

a result that leads to the conclusion that

$$\mathbf{P}(A) = 1 - \mathbf{P}(\bar{A}). \quad (1-11)$$

3) For any events A and B we have

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B). \quad (1-12)$$

Proof: The two identities

$$A \cup B = A \cup ((B \cap A) \cup (B \cap \bar{A})) = (A \cup (B \cap A)) \cup (B \cap \bar{A}) = A \cup (B \cap \bar{A})$$

$$B = B \cap (A \cup \bar{A}) = (A \cap B) \cup (B \cap \bar{A})$$

lead to

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B \cap \bar{A})$$

$$\mathbf{P}(B) = \mathbf{P}(A \cap B) + \mathbf{P}(B \cap \bar{A}).$$

Subtract these last two expressions to obtain the desired results

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

Note that this result is generalized easily to the case of three or more events.

Conditional Probability

The *conditional probability* of an event A , assuming that event M has occurred, is

$$\mathbf{P}(A|M) = \frac{\mathbf{P}(A \cap M)}{\mathbf{P}(M)}, \quad (1-13)$$

where it is assumed that $\mathbf{P}(M) \neq 0$. Note that $\mathbf{P}(A \cap M) = \mathbf{P}(A|M)\mathbf{P}(M)$, a useful identity.

Consider two special cases. The first case is $M \subset A$ so that $A \cap M = M$. For $M \subset A$, we have

$$\mathbf{P}(A|M) = \frac{\mathbf{P}(A \cap M)}{\mathbf{P}(M)} = \frac{\mathbf{P}(M)}{\mathbf{P}(M)} = 1. \quad (1-14)$$

Next, consider the special case $A \subset M$ so that $\mathbf{P}(M|A) = 1$. For this case, we have

$$\begin{aligned} \mathbf{P}(A|M) &= \frac{\mathbf{P}(A \cap M)}{\mathbf{P}(M)} = \frac{\mathbf{P}(M|A)}{\mathbf{P}(M)} \mathbf{P}(A) \\ &= \frac{1}{\mathbf{P}(M)} \mathbf{P}(A) \\ &\geq \mathbf{P}(A), \end{aligned} \quad (1-15)$$

an intuitive result.

Example 1-12: In the fair die experiment, the outcomes are f_1, f_2, \dots, f_6 , the six faces of the die. Let $A = \{f_2\}$, the event "a two occurs", and $M = \{f_2, f_4, f_6\}$, the event "an even outcome occurs". Then we have $P(A) = 1/6$, $P(M) = 1/2$ and $P(A \cap M) = P(A)$ so that

$$P(\{f_2\} | \text{"even"}) = \frac{1/6}{1/2} = 1/3 > P(f_2) = 1/6.$$

Example 1-13: A box contains three white balls w_1, w_2, w_3 and two red balls r_1 and r_2 . We remove at random and without replacement two balls in succession. What is the probability that the first removed ball is white and the second is red?

$$P[\{\text{first ball is white}\}] = 3/5$$

$$P[\{\text{second is red}\} | \{\text{first ball is white}\}] = 1/2$$

$$\begin{aligned} P[\{\text{first ball is white}\} \cap \{\text{second ball is red}\}] &= P[\{\text{second is red}\} | \{\text{first ball is white}\}] P[\{\text{first ball is white}\}] \\ &= (1/2)(3/5) = 3/10 \end{aligned}$$

Theorem 1-1 (Total Probability Theorem – Discrete Version): Let $[A_1, A_2, \dots, A_n]$ be a *partition* of S . That is,

$$\bigcup_{i=1}^n A_i = S \text{ and } A_i \cap A_j = \{\emptyset\} \text{ for } i \neq j. \quad (1-16)$$

Let B be an arbitrary event. Then

$$P[B] = P[B | A_1] P[A_1] + P[B | A_2] P[A_2] + \dots + P[B | A_n] P[A_n]. \quad (1-17)$$

Proof: First, note the set identity

$$B = B \cap S = B \cap (A_1 \cup A_2 \cup \dots \cup A_n) = (B \cap A_1) \cup (B \cap A_2) \cup \dots \cup (B \cap A_n)$$

For $i \neq j$, $B \cap A_i$ and $B \cap A_j$ are mutually exclusive. Hence, we have

$$\begin{aligned} P[B] &= P[B \cap A_1] + P[B \cap A_2] + \dots + P[B \cap A_n] \\ &= P[B | A_1] P[A_1] + P[B | A_2] P[A_2] + \dots + P[B | A_n] P[A_n]. \end{aligned} \tag{1-18}$$

This result is known as the *Total Probability Theorem, Discrete Version*.

Example 1-14: Let $[A_1, A_2, A_3]$ be a partition of S . Consider the identity

$$P[B] = P[B | A_1] P[A_1] + P[B | A_2] P[A_2] + P[B | A_3] P[A_3].$$

This equation is illustrated by Figure 1-4.

Bayes Theorem

Let $[A_1, A_2, \dots, A_n]$ be a partition of S . Since

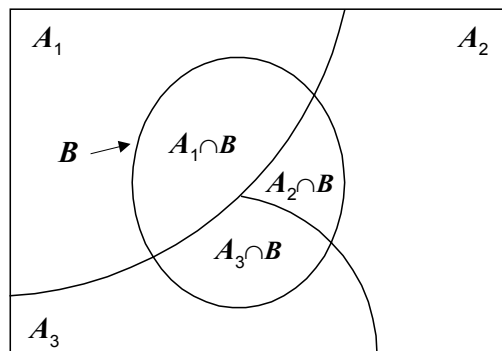


Figure 1-4: Example that illustrates the Total Probability Theorem.

$$P(A_i|B) = P(B|A_i) \frac{P(A_i)}{P(B)}$$

$$P[B] = P[B|A_1]P[A_1] + P[B|A_2]P[A_2] + \dots + P[B|A_n]P[A_n],$$

we have

Theorem 1-2 (Bayes Theorem):

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P[B|A_1]P[A_1] + P[B|A_2]P[A_2] + \dots + P[B|A_n]P[A_n]}. \quad (1-19)$$

The $P[A_i]$ are called *apriori* probabilities, and the $P[A_i|B]$ are called *aposteriori probabilities*.

Bayes theorem provides a method for incorporating experimental observations into the characterization of an event. Both $P[A_i]$ and $P[A_i|B]$ characterize events A_i , $1 \leq i \leq n$; however, $P[A_i|B]$ may be a better (more definitive) characterization, especially if B is an event related to the A_i . For example, consider events $A_1 = [\text{snow today}]$, $A_2 = [\text{no snow today}]$ and $T = [\text{today's temperature is above } 70^\circ\text{F}]$. Given the occurrence of T , one would expect $P[A_1|T]$ and $P[A_2|T]$ to more definitively characterize “snow today” than does $P[A_1]$ and $P[A_2]$.

Example 1-15: We have four boxes. Box #1 contains 2000 components, 5% defective. Box #2 contains 500 components, 40% defective. Boxes #3 and #4 contain 1000 components each, 10% defective in both boxes. At random, we select one box and remove one component.

a) What is the probability that this component is defective? From the theorem of total probability we have

$$\begin{aligned}
 P[\text{Component is defective}] &= \sum_{i=1}^4 P[\text{defective} | \text{Box } i] P[\text{Box } i] \\
 &= (.05)(.25) + (.4)(.25) + (.1)(.25) + (.1)(.25) \\
 &= .1625
 \end{aligned}$$

b) We examine a component and find that it is defective. What is the probability that it came from Box #2? By Bayes Law, we have

$$\begin{aligned}
 P(\text{Box\#2} | \text{Defective}) &= \frac{P(\text{Defective} | \text{Box \#2}) P(\text{Box\#2})}{P[\text{Defective} | \text{Box \#1}] P[\text{Box\#1}] + \dots + P[\text{Defective} | \text{Box \#4}] P[\text{Box\#4}]} \\
 &= \frac{(.4)(.25)}{.1625} \\
 &= .615
 \end{aligned}$$

Independence

Events A and B are said to be *independent* if

$$P[A \cap B] = P[A] P[B]. \quad (1-20)$$

If A and B are independent, then

$$P[A | B] = \frac{P[A \cap B]}{P[B]} = \frac{P[A]P[B]}{P[B]} = P[A]. \quad (1-21)$$

Three events A_1 , A_2 , and A_3 are independent if

- 1) $P[A_i \cap A_j] = P[A_i] P[A_j]$ for $i \neq j$
- 2) $P[A_1 \cap A_2 \cap A_3] = P[A_1] P[A_2] P[A_3]$.

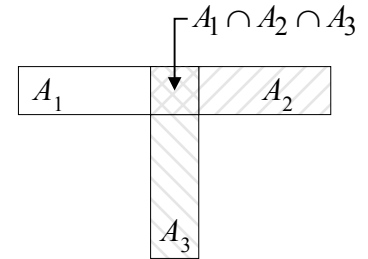
Be careful! Condition 1) may hold, and condition 2) may not. Likewise, Condition 2) may hold, and condition 1) may not. **Both** are required for the three events to be independent.

Example 1-16: Suppose

$$\mathbf{P}[A_1] = \mathbf{P}[A_2] = \mathbf{P}[A_3] = 1/5$$

$$\mathbf{P}[A_1 \cap A_2] = \mathbf{P}[A_1 \cap A_3] = \mathbf{P}[A_2 \cap A_3] = \mathbf{P}[A_1 \cap A_2 \cap A_3] = p$$

If $p = 1/25$, then $\mathbf{P}[A_i \cap A_j] = \mathbf{P}[A_i] \mathbf{P}[A_j]$ for $i \neq j$ holds, so that requirement 1) holds. However, $\mathbf{P}[A_1 \cap A_2 \cap A_3] \neq \mathbf{P}[A_1] \mathbf{P}[A_2] \mathbf{P}[A_3]$, so that requirement 2) fails. On the other hand, if $p = 1/125$, then $\mathbf{P}[A_1 \cap A_2 \cap A_3] = \mathbf{P}[A_1] \mathbf{P}[A_2] \mathbf{P}[A_3]$, and requirement 2) holds. But, $\mathbf{P}[A_i \cap A_j] \neq \mathbf{P}[A_i] \mathbf{P}[A_j]$ for $i \neq j$, so that requirement 1) fails.



More generally, the independence of n events can be defined inductively. We say that n events A_1, A_2, \dots, A_n are independent if

- 1) All combinations of $k, k < n$, events are independent, **and**
- 2) $\mathbf{P}[A_1 \cap A_2 \cap \dots \cap A_n] = \mathbf{P}[A_1] \mathbf{P}[A_2] \dots \mathbf{P}[A_n]$.

Starting from $n = 2$, we can use this requirement to generalize independence to an arbitrary, but finite, number of events.

Cartesian Product of Sets

The *cartesian product* of sets A_1 and A_2 is denoted as $A_1 \times A_2$, and it is a set whose elements are all ordered pairs (a_1, a_2) , where $a_1 \in A_1$ and $a_2 \in A_2$. That is,

$$A_1 \times A_2 \equiv \{(a_1, a_2) : a_1 \in A_1, a_2 \in A_2\}. \quad (1-22)$$

Example 1-17: Let $A = \{h, t\}$ and $B = \{u, v\}$ so that $A \times B = \{hu, hv, tu, tv\}$

Clearly, the notion of cartesian product can be extended to the product of $n, n > 2$, sets.

Generalized Rectangles

Suppose $A \subset S_1$ and $B \subset S_2$. Then, the cartesian product $A \times B$ can be represented as $A \times B = (A \times S_2) \cap (S_1 \times B)$, a result illustrated by Figure 1-5 (however, $A \times B$ need not be one

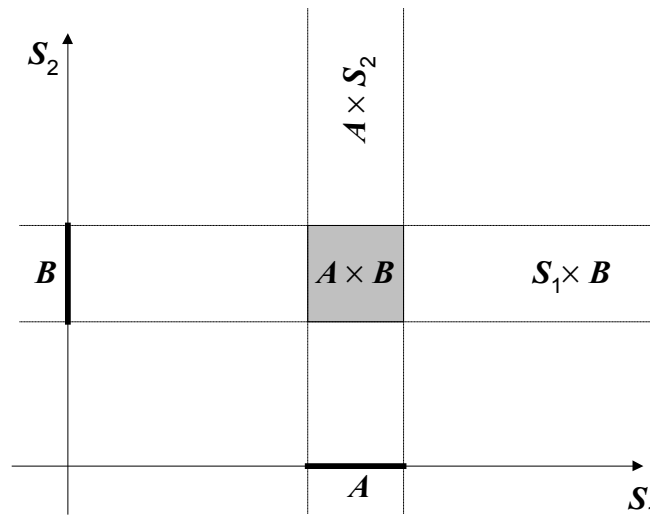


Figure 1-5: Cartesian product as intersection of generalized rectangle.

contiguous piece as depicted by the figure). Because of this geometric interpretation, the sets $A \times B$, $A \subset S_1$ and $B \subset S_2$, are referred to as *generalized rectangles*.

Combined Experiments - Product Spaces

Consider the experiments 1) rolling a fair die, with probability space $(S_1, \mathcal{F}_1, P_1)$, and 2) tossing a fair coin, with probability space $(S_2, \mathcal{F}_2, P_2)$. Suppose both experiments are performed. What is the probability that we get "two" on the die and "heads" on the coin. To solve this problem, we combine the two experiments into a single experiment described by $(S_C, \mathcal{F}_C, P_C)$, known as a *product experiment* or *product space*. The *product sample space* S_C , *product σ -algebra* \mathcal{F}_C , and *product probability measure* P_C are discussed in what follows.

Product Sample Space S_C

To combine the two sample spaces, we take $S_C = S_1 \times S_2$. Sample space S_C is defined as

$$S_C \equiv S_1 \times S_2 \equiv [(\omega_1, \omega_2) : \omega_1 \in S_1, \omega_2 \in S_2]. \quad (1-23)$$

S_C consists of all possible pairs (ω_1, ω_2) of elementary outcomes, $\omega_1 \in S_1$ and $\omega_2 \in S_2$.

Product σ -Algebra \mathcal{F}_C

Set \mathcal{F}_C of combined events must contain all possible products $A \times B$, where $A \in \mathcal{F}_1$ and B

$\in \mathcal{F}_2$. Also, all possible unions and intersections of these products must be included in \mathcal{F}_C . Basically, we take \mathcal{F}_C to be the σ -algebra *generated* by the collection of generalized rectangles $\{A \times B, \text{ where } A \in \mathcal{F}_1 \text{ and } B \in \mathcal{F}_2\}$; see the discussion after Example 1-7. Equivalently, consider σ -algebras that contain the collection $A \times B$, where $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$ are arbitrary events; then, \mathcal{F}_C is the intersection of all such σ -algebras.

Product Probability Measure P_C

The product probability measure $P_C : \mathcal{F}_C \rightarrow [0, 1]$ is the probability measure for the combined experiment. We would like to express P_C in terms of P_1 and P_2 . However, without detailed knowledge of experiment dependencies (*i.e.*, how does an outcome of the first experiment affect possible outcomes of the second experiment), it is not possible to determine P_C , in general.

We can specify P_C for a limited set of events in \mathcal{F}_C . For $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$, the quantities $A \times S_2$ and $S_1 \times B$ are events in \mathcal{F}_C . It is natural to assign the probabilities

$$P_C[A \times S_2] = P_1[A] \tag{1-24}$$

$$P_C[S_1 \times B] = P_2[B].$$

There is an important special case where it is possible to completely specify P_C in terms of P_1 and P_2 . Assume that A and B are independent so that $A \times S_2$ and $S_1 \times B$ are independent events in \mathcal{F}_C . Compute the probability of the combined outcome $A \times B$ as

$$P_C[A \times B] = P_C[(A \times S_2) \cap (S_1 \times B)] = P_C[A \times S_2] P_C[S_1 \times B] = P_1[A] P_2[B].$$

Hence, when $A \times S_2$ and $S_1 \times B$ are independent for all $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$, the probability measure for the combined experiment can be determined from the probability measures of the individual experiments. However, if $A \times S_2$ and $S_1 \times B$ are dependent for some A and B then it is not possible to write P_C for the combined experiment without knowing how $A \in \mathcal{F}_1$ and $B \in$

\mathcal{F}_2 are related.

We have shown how to naturally combine two experiments into one experiment. Clearly, this process can be extended to combine n separate experiments into a single experiment.

Counting Subsets of Size k

If a set has n distinct elements, then the total number of its subsets consisting of k elements each is

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (1-25)$$

Order within the subsets of k elements is not important. For example, the subset $\{a, b\}$ is the same as $\{b, a\}$.

Example 1-18: Form the

$$\binom{3}{2} = \frac{3!}{2!1!} = 3$$

subsets of size two from the set $\{a_1, a_2, a_3\}$. These three subsets are (a_1, a_2) , (a_1, a_3) and (a_2, a_3) .

Bernoulli Trials - Repeated Trials

Consider the experiment $(\mathcal{S}, \mathcal{F}, \mathbf{P})$ and the event $A \in \mathcal{F}$. Suppose

$$\mathbf{P}(A) = p$$

$$\mathbf{P}(\bar{A}) = 1 - p = q.$$

We conduct n *independent trials* of this experiment to obtain the combined experiment with sample space $\mathcal{S}_C = \mathcal{S} \times \mathcal{S} \times \dots \times \mathcal{S}$, a cartesian product of n copies of \mathcal{S} . For the combined experiment, the set of events \mathcal{F}_C and the probability measure \mathbf{P}_C are obtained as described previously (\mathbf{P}_C is easy to obtain because the trials are independent). The probability that A

occurs k times (in any order) is

$$\mathbf{P}[A \text{ Occurs } k \text{ Times in } n \text{ Independent Trials}] = \binom{n}{k} p^k (1-p)^{n-k}. \quad (1-26)$$

In (1-26), it is important to remember that the *order* in which A and \bar{A} occur is not important.

Proof

The n independent repetitions are known as *Bernoulli Trials*. The event $\{A \text{ Occurs } k \text{ Times In a Specific Order}\}$ is the cartesian product $A_1 \times A_2 \times \dots \times A_n$, where k of the A_i are A and $n - k$ are \bar{A} , and a specified ordering is given. The probability of this specific event is

$$\mathbf{P}_1[A_1] \mathbf{P}_1[A_2] \cdots \mathbf{P}_1[A_n] = p^k (1-p)^{n-k}.$$

Equivalently,

$$\mathbf{P}[A \text{ Occurs } k \text{ Times In a Specific Order}] = p^k (1-p)^{n-k}.$$

Now, the event $\{A \text{ Occurs } k \text{ Times In Any Order}\}$ is the union of the $\binom{n}{k}$ *mutually exclusive, equally likely*, events of the type $\{A \text{ Occurs } k \text{ Times In a Specific Order}\}$. Hence, we have

$$\mathbf{P}[A \text{ Occurs } k \text{ Times in } n \text{ Independent Trials}] = \binom{n}{k} p^k (1-p)^{n-k},$$

as claimed.

Often, we are interested in the probability of A occurring *at least* k_1 times, but *no more* than k_2 times, in n independent trials. The probability that event A occurs *at least* k_1 times, but *no more* than k_2 times, is given as

$$\mathbf{P}[\mathcal{A} \text{ occurs between } k_1 \text{ and } k_2 \text{ times}] = \sum_{k=k_1}^{k_2} \mathbf{P}[\mathcal{A} \text{ occurs } k \text{ times}] = \sum_{k=k_1}^{k_2} \binom{n}{k} p^k q^{n-k}. \quad (1-27)$$

Example 1-19: A factory produces items, 1% of which are bad. Suppose that a random sample of 100 of these items is drawn from a large consignment. Calculate the probability that the sample contains no defective items. Let X denote the number of bad items in the sample of 100 items. Then, X is distributed binomially with parameters $n = 100$ and $p = .01$. Hence, we can compute

$$\mathbf{P}[X = 0] = \mathbf{P}[\text{No bad items in sample of 100 items}]$$

$$\begin{aligned} &= \binom{100}{0} (.01)^0 (1 - .01)^{100} \\ &= .366 \end{aligned}$$

Gaussian Function

The function

$$g(x) \equiv \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}x^2\right] \quad (1-28)$$

is known as the *Gaussian function*, see Figure 1-6. The Gaussian function can be used to define

$$G(x) \equiv \int_{-\infty}^x g(y) dy, \quad (1-29)$$

a tabulated function (also, G is an intrinsic function in MatLab, Matcad and other mathematical software). It is obvious that $G(-\infty) = 0$; it is known that $G(\infty) = 1$. We will use tables of G to evaluate integrals of the form

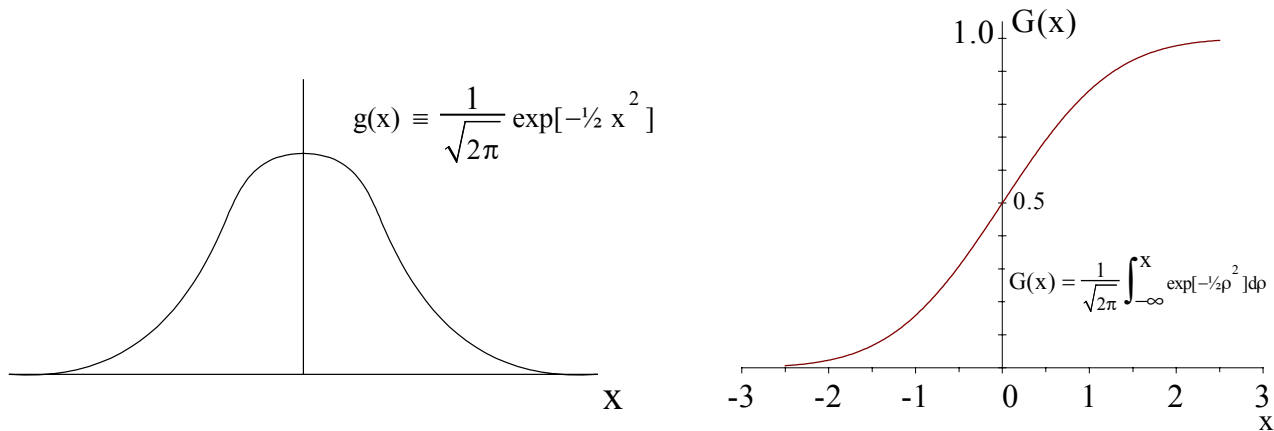


Figure 1-6: The Gaussian function

$$\begin{aligned} \frac{1}{\sigma} \int_{x_1}^{x_2} g\left(\frac{x-\eta}{\sigma}\right) dx &= \frac{1}{\sqrt{2\pi}\sigma} \int_{x_1}^{x_2} \exp\left[-\frac{(x-\eta)^2}{2\sigma^2}\right] dx = \frac{1}{\sqrt{2\pi}} \int_{(x_1-\eta)/\sigma}^{(x_2-\eta)/\sigma} \exp\left[-\frac{y^2}{2}\right] dy \\ &= G\left(\frac{x_2-\eta}{\sigma}\right) - G\left(\frac{x_1-\eta}{\sigma}\right) \end{aligned} \quad (1-30)$$

where $-\infty < \eta < \infty$ and $0 < \sigma < \infty$ are known numbers. Due to symmetry in the g function, it is easy to see that

$$G(-x) = 1 - G(x). \quad (1-31)$$

Many tables contain values of $G(x)$ for $x \geq 0$ only. Using these tables and (1-31), we can determine $G(x)$ for negative x .

The Gaussian function G is related to the *error function*. For $x \geq 0$,

$$G(x) = \frac{1}{2} + \text{erf}(x), \quad (1-32)$$

where

$$\operatorname{erf}(x) \equiv \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{1}{2}u^2} du, \quad x \geq 0, \quad (1-33)$$

is the well-known (and tabulated in mathematical handbooks) *error function*.

Example 1-20: Consider again Example 1-11 with a Gaussian density function. Here, the sample space \mathcal{S} consists of the whole real line \mathcal{R} . For \mathcal{F} , we use the Borel σ -algebra \mathcal{B} discussed in Examples 1-9 and 1-11. Finally, for $B \in \mathcal{B}$, we use the probability measure

$$\mathbf{P}(B) \equiv \frac{1}{\sqrt{2\pi}} \int_B e^{-\frac{1}{2}u^2} du, \quad B \in \mathcal{B}. \quad (1-34)$$

Formally denoted as $(\mathcal{R}, \mathcal{B}, \mathbf{P})$, this probability space is used in many applications.

DeMoivre-Laplace Theorem

The Bernoulli trials formula is not practical for large n since $n!$ cannot be calculated easily. We seek ways to approximate the result. One such approximation is given by

Theorem 1-3 (DeMoivre-Laplace Theorem): Let $q = 1 - p$. If $npq \gg 1$, then

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} \exp\left[-\frac{(k - np)^2}{2npq}\right] = \frac{1}{\sqrt{npq}} g\left(\frac{k - np}{\sqrt{npq}}\right) \quad (1-35)$$

for k in an \sqrt{npq} neighborhood of np (i.e., $|k - np| < \sqrt{npq}$).

This theorem is used as follows. Suppose our experiment has only two outcomes, “success” (i.e., event A) and “failure” (i.e., event \bar{A}). On any trial, the probability of a “success” is p , and the probability of “failure” is $1 - p$. Suppose we conduct n *independent* trials of the experiment. Let S_n denote the number of “successes” that occur in n independent trials. Clearly, $0 \leq S_n \leq n$. The DeMoivre-Laplace theorem states that for $npq \gg 1$ we have

$$\mathbf{P}[S_n = k] = \binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} \exp\left[-\frac{1}{2} \left\{ \frac{(k - np)}{\sqrt{npq}} \right\}^2\right] = \frac{1}{\sqrt{npq}} g\left(\frac{k - np}{\sqrt{npq}}\right) \quad (1-36)$$

for $|k - np| < \sqrt{npq}$.

Example 1-21: A fair coin is tossed 1000 times. Find the probability that heads will show *exactly* 500 times. For this problem, $p = q = .5$ and $k - np = 0$. Hence,

$$\mathbf{P}[\text{exactly 500 heads}] \approx \frac{1}{10\sqrt{5\pi}} = .0252$$

Now, we want to approximate the probability of obtaining between k_1 and k_2 occurrences of event A . By the DeMoivre-Laplace theorem

$$\mathbf{P}[k_1 \leq k \leq k_2] = \sum_{k=k_1}^{k_2} \binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{npq}} \sum_{k=k_1}^{k_2} g\left(\frac{k - np}{\sqrt{npq}}\right) \quad (1-37)$$

assuming that $|k_1 - np| < \sqrt{npq}$ and $|k_2 - np| < \sqrt{npq}$. If \sqrt{npq} is large, then $g([k - np] / \sqrt{npq})$ changes slowly for $k_1 \leq k \leq k_2$, and

$$\begin{aligned} \mathbf{P}[k_1 \leq k \leq k_2] &\approx \frac{1}{\sqrt{npq}} \sum_{k=k_1}^{k_2} g\left(\frac{k - np}{\sqrt{npq}}\right) \approx \frac{1}{\sqrt{npq}} \int_{k_1}^{k_2} g\left(\frac{x - np}{\sqrt{npq}}\right) dx \\ &= G\left(\frac{k_2 - np}{\sqrt{npq}}\right) - G\left(\frac{k_1 - np}{\sqrt{npq}}\right), \end{aligned} \quad (1-38)$$

as illustrated by Figure 1-7.

Example 1-22: A fair coin is tossed 10,000 times. Approximate a numerical value for $\mathbf{P}[4950 \leq \# \text{Heads} \leq 5050]$. Since $k_1 = 4950$ and $k_2 = 5050$, we have

$$\frac{k_2 - np}{\sqrt{npq}} = 1 \quad \text{and} \quad \frac{k_1 - np}{\sqrt{npq}} = -1.$$

Application of (1-38) leads to the answer

$$P[4950 \leq \text{\#Heads} \leq 5050] \approx G(1) - G(-1) = G(1) - [1 - G(1)] = .6826$$

From what is given above, we **might** conclude that approximations (1-37) and (1-38) **require** the restrictions $|k_1 - np| < \sqrt{npq}$ and $|k_2 - np| < \sqrt{npq}$. However, these restrictions **may not** be necessary in a given application. In (1-37), terms near the beginning (*i.e.*, $k = k_1$) and end (*i.e.*, $k = k_2$) of the sum generally contain the most error (assuming $k_1 < np < k_2$). However, for large enough n , the total sum of these errors (*i.e.*, the total error) is small compared to the entire sum of all terms (*i.e.*, the answer). That is, in (1-37) with large n , the “highly accurate” terms (those close to $k = np$) have a sum that dominates the sum of all terms (*i.e.*, the entire sum (1-37)), so the error in “the tails” (the terms near $k = k_1$ and $k = k_2$ contain the most error) becomes less significant as n becomes large. In fact, in the limit as $n \rightarrow \infty$, Approximation (1-38) becomes exact; no restrictions are required on $|k_1 - np|/\sqrt{npq}$ and $|k_2 - np|/\sqrt{npq}$.

Theorem 1-4 (How DeMoivre-Laplace is stated most often): As above, let S_n denote the number of “successes” in n independent trials. Define a “centered” and “normalized” version of

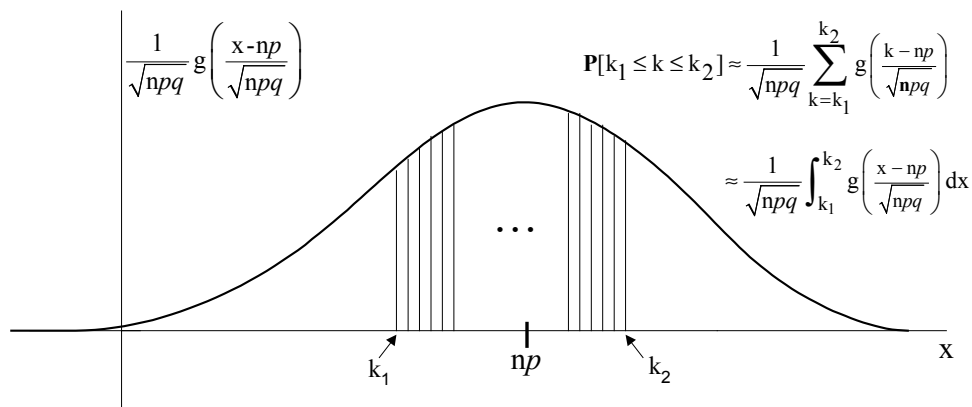


Figure 1-7: Gaussian approximation to the Binomial density function.

S_n as

$$\tilde{S}_n \equiv \frac{S_n - np}{\sqrt{npq}}.$$

Now, denote x_1 and x_2 as arbitrary real numbers. Then, we can say

$$\lim_{n \rightarrow \infty} \mathbf{P}[x_1 \leq \tilde{S}_n \leq x_2] = \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} \exp(-x^2/2) dx = G(x_2) - G(x_1) \quad (1-39)$$

Proof: The proof is based on Stirling's approximation for $n!$, and it can be found in many books. For example, see one of

[1] E. Parzen, *Modern Probability Theory and Its Applications*, John Wiley, 1960.

[2] Y.A. Rozanov, *Probability Theory: A Concise Course*, Dover, 1969.

[3] A. Papoulis, S. Pillai, *Probability, Random Variables and Stochastic Processes*, **Fourth Edition**, McGraw Hill, 2002 (proof not in editions I through III).

Example 1-23: An order of 10^4 parts is received. The probability that a part is defective equals $1/10$. What is the probability that the total number of defective parts does not exceed 1100?

$$\mathbf{P}[\# \text{defective parts} \leq 1100] = \sum_{k=0}^{1100} \mathbf{P}[k \text{ defective parts}] = \sum_{k=0}^{1100} \binom{10^4}{k} (.1)^k (.9)^{10^4 - k}$$

Since np is large

$$\begin{aligned}
 \mathbf{P}[\#\text{defective parts} \leq 1100] &\approx G\left(\frac{1100-1000}{\sqrt{900}}\right) - G\left(\frac{0-1000}{\sqrt{900}}\right) \\
 &\approx G(10/3) \text{ since } G(10/3) \gg G(-100/3) \approx 0 \\
 &= .99936
 \end{aligned}$$

In this example, we used the approximation

$$F(k) = \sum_{i=0}^k \binom{n}{i} p^i q^{n-i} \approx G\left(\frac{k-np}{\sqrt{npq}}\right) - G\left(\frac{-np}{\sqrt{npq}}\right) \approx G\left(\frac{k-np}{\sqrt{npq}}\right) \quad (1-40)$$

which can be used when $np \gg 1$ so that $G(-np/\sqrt{npq}) \approx 0$. The sum of the terms from $k = 900$ to $k = 1100$ equals .99872. Note that the terms from $k = 0$ to 900 do not amount to much!

Example 1-24: Figures 1-8 and 1-9 illustrate the De-Moivre Laplace theorem. The first of these figures depicts, as a solid line plot, $G(\{x - np\}/\sqrt{npq})$ for $n = 25$, $p = q = 1/2$. As a sequence of dots, values are depicted of the Binomial function $F(k)$, for the case $n = 25$, $p = q =$

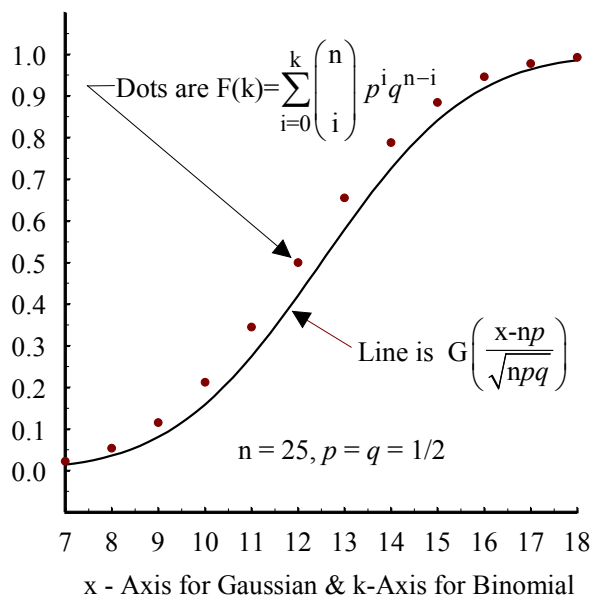


Figure 1-8: Gaussian and Binomial distribution functions for $n = 25$, $p = q = 1/2$.

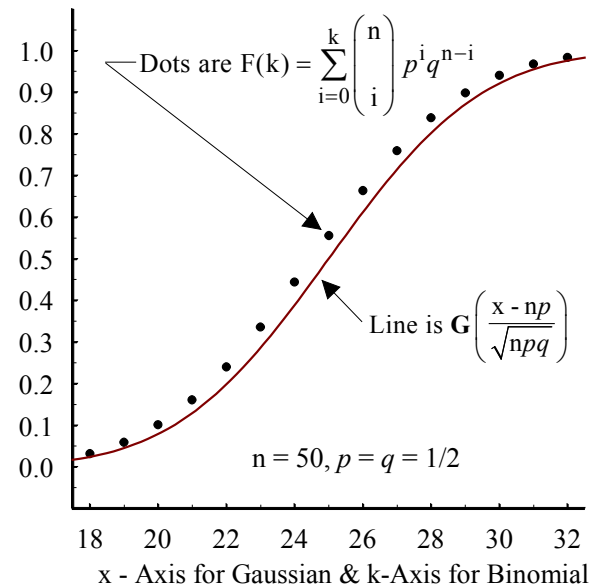


Figure 1-9: Gaussian and Binomial distribution functions for $n = 50$, $p = q = 1/2$.

1/2 (given by the sum in (1-40)) is displayed on Fig. 1-8. In a similar manner, Figure 1-9 illustrates the case for $n = 50$ and $p = q = 1/2$.

Law of Large Numbers (Weak Law)

Suppose we perform n independent trials of an experiment. The probability of event A occurring on any trial is p . We should expect that the number k of occurrences of A is about np so that k/n is near p . In fact, the *Law of Large Numbers* (weak version) says that k/n is close to p in the sense that, for any $\varepsilon > 0$, the probability that $|k/n - p| \leq \varepsilon$ tends to 1 as $n \rightarrow \infty$. This result is given by the following theorem.

Theorem 1-5 (Law of Large Numbers - Weak Version): For all $\varepsilon > 0$, we have

$$\lim_{n \rightarrow \infty} \mathbf{P} \left[\left| \frac{k}{n} - p \right| \leq \varepsilon \right] = 1. \quad (1-41)$$

That is, as n becomes larger, it becomes more probable to find k/n near p .

Proof: Note that

$$\left| k/n - p \right| \leq \varepsilon \Rightarrow -\varepsilon \leq \frac{k}{n} - p \leq \varepsilon \Rightarrow n(p - \varepsilon) \leq k \leq n(p + \varepsilon),$$

so that

$$\begin{aligned} \mathbf{P} \left[\left| \frac{k}{n} - p \right| \leq \varepsilon \right] &= \mathbf{P} (n(p - \varepsilon) \leq k \leq n(p + \varepsilon)) = \sum_{k=n(p-\varepsilon)}^{n(p+\varepsilon)} \binom{n}{k} p^k q^{n-k} \\ &\approx G \left(\frac{n\varepsilon}{\sqrt{npq}} \right) - G \left(\frac{-n\varepsilon}{\sqrt{npq}} \right) = 2G \left(\frac{n\varepsilon}{\sqrt{npq}} \right) - 1. \end{aligned}$$

But $\frac{n\varepsilon}{\sqrt{npq}} \rightarrow \infty$ and $G \left(\frac{n\varepsilon}{\sqrt{npq}} \right) \rightarrow 1$ as $n \rightarrow \infty$. Therefore,

$$\mathbf{P} \left[\left| \frac{k}{n} - p \right| \leq \varepsilon \right] \approx 2G \left(\varepsilon \sqrt{\frac{n}{pq}} \right) - 1 \rightarrow 1$$

as $n \rightarrow \infty$, and the Law of Large Numbers is proved.

Example 1-25: Let $p = .6$, and find large n such that the probability that k/n is between .59 and .61 is at least 98/100. That is, choose n so that $\mathbf{P} [.59 < k/n < .61] \geq .98$. This requires

$$\begin{aligned} \mathbf{P} (.59n \leq k \leq .61n) &\approx G \left(\frac{.61n - .60n}{\sqrt{npq}} \right) - G \left(\frac{.59n - .60n}{\sqrt{npq}} \right) = G \left(\frac{.01n}{\sqrt{npq}} \right) - G \left(\frac{-.01n}{\sqrt{npq}} \right) \\ &= 2G \left(\frac{.01n}{\sqrt{npq}} \right) - 1 \geq .98 \end{aligned}$$

or

$$G \left(.01 \sqrt{n/pq} \right) \geq 1.98 / 2 = .9900$$

From a table of the Gaussian distribution function, we see that $G(2.33) = .9901$, a value that is close enough for our problem. Hence, we equate

$$.01 \sqrt{\frac{n}{pq}} = 2.33$$

and solve this for

$$n = (2.33)^2 \frac{(.6)(.4)}{10^{-4}} \approx 13,029$$

so we must choose $n > 13,029$.

Generalization of Bernoulli Trials

Suppose $[A_1, A_2, \dots, A_r]$ is a *partition* of the sample space. That is,

$$\bigcup_{i=1}^r A_i = \mathcal{S}$$

$$A_i \cap A_j = \{\emptyset\}, \quad i \neq j.$$

Furthermore, for $1 \leq i \leq r$, let $\mathbf{P}(A_i) = p_i$ and $p_1 + p_2 + \dots + p_r = 1$. Now, perform n independent trials of the experiment and denote by $p_n(k_1, k_2, \dots, k_r)$ the probability of the event

$\{A_1 \text{ occurs } k_1 \text{ times, } A_2 \text{ occurs } k_2 \text{ times, } \dots, A_r \text{ occurs } k_r \text{ times}\}$,

where $k_1 + k_2 + \dots + k_r = n$. Order is not important here. This is a generalization of previous work which had the two event A and \bar{A} . Here, the claim is

$$p_n(k_1, k_2, \dots, k_r) = \frac{n!}{k_1! k_2! \dots k_r!} p_1^{k_1} p_2^{k_2} \dots p_r^{k_r}. \quad (1-42)$$

Proof: First, consider a "counting problem". From n distinct objects, how many ways can you form a first subset of size k_1 , a second subset of size k_2 , \dots , an r^{th} subset of size k_r ?

Number of ways of forming a first subset of size k_1 and a second of size $n-k_1$ is

$$\frac{n!}{k_1!(n-k_1)!}.$$

Number of ways of forming a subset of size k_1 , a second of size k_2 and a third of size $n-k_1-k_2$ is

$$\frac{n!}{k_1!(n-k_1)!} \times \frac{(n-k_1)!}{k_2!(n-k_1-k_2)!} \cdot$$

Number of ways of forming a subset of size k_1 , a second of size k_2 , a third of size k_3 and a fourth of size $n-k_1-k_2-k_3$ is

$$\frac{n!}{k_1!(n-k_1)!} \times \frac{(n-k_1)!}{k_2!(n-k_1-k_2)!} \times \frac{(n-k_1-k_2)!}{k_3!(n-k_1-k_2-k_3)!} \cdot$$

Number of ways of forming a first subset of size k_1 , a second subset of size k_2 , ... , an r^{th} subset of size k_r (where $k_1 + k_2 + \dots + k_r = n$) is

$$\begin{aligned} & \frac{n!}{k_1!(n-k_1)!} \times \frac{(n-k_1)!}{k_2!(n-k_1-k_2)!} \times \frac{(n-k_1-k_2)!}{k_3!(n-k_1-k_2-k_3)!} \times \dots \times \frac{(n-k_1-k_2-\dots-k_{r-1})!}{k_r!(n-k_1-k_2-\dots-k_r)!} \\ &= \frac{n!}{k_1!k_2!\dots k_r!} \end{aligned}$$

Hence, the probability of A_1 occurring k_1 times, A_2 occurring k_2 times ... A_r occurring k_r times (these occur in any order) is

$$p_n(k_1, k_2, \dots, k_r) = \frac{n!}{k_1!k_2!\dots k_r!} p_1^{k_1} p_2^{k_2} \dots p_r^{k_r}$$

as claimed.

Example 1-26: A fair die is rolled 10 times. Determine the probability that f_1 shows 3 times and "even" shows 6 times.

$$A_1 = \{f_1 \text{ shows}\}$$

$$A_2 = \{f_2 \text{ or } f_4 \text{ or } f_6 \text{ shows}\}$$

$$A_3 = \{f_3 \text{ or } f_5 \text{ shows}\}$$

$$A_1 \cup A_2 \cup A_3 = \mathcal{S} = \{f_1, f_2, f_3, f_4, f_5, f_6\}$$

$$A_i \cap A_j = \{\emptyset\} \text{ for } i \neq j$$

$$\mathbf{P}(A_1) = 1/6, \mathbf{P}(A_2) = 1/2, \mathbf{P}(A_3) = 1/3$$

$$n = 10, k_1 = \# \text{ times } A_1 \text{ occurs} = 3, k_2 = \# \text{ times } A_2 \text{ occurs} = 6 \text{ and } k_3 = \# \text{ times } A_3 \text{ occurs} = 1$$

$$\mathbf{P}(f_1 \text{ shows 3 times, "even" shows 6 times, not } (f_1 \text{ or even) shows 1 time}) = \mathbf{P}_{10}(3, 6, 1)$$

$$= \frac{10!}{3!6!1!} \left(\frac{1}{6}\right)^3 \left(\frac{1}{2}\right)^6 \frac{1}{3} = .0203$$

Poisson Theorem and Random Points

The probability that A occurs k times in n independent trials is

$$\mathbf{P}[A \text{ occurs } k \text{ time in } n \text{ independent trials}] = \binom{n}{k} p^k q^{n-k}. \quad (1-43)$$

If n is large and $npq \gg 1$, we can use the DeMoivre Laplace theorem to approximate the probability (1-43). However, the DeMoivre Laplace theorem is no good if n is large and p is small so that np is on the order of 1. However, for this case, we can use the *Poisson Approximation*.

Theorem 1-6 (Poisson Theorem): As $n \rightarrow \infty$ and $p \rightarrow 0$, such that $np \rightarrow \lambda$ (a constant), we have

$$\binom{n}{k} p^k q^{n-k} \xrightarrow[\substack{n \rightarrow \infty \\ p \rightarrow 0 \\ np \rightarrow \lambda}]{\lambda} e^{-\lambda} \frac{\lambda^k}{k!}. \quad (1-44)$$

Proof

$$\begin{aligned}
 \binom{n}{k} p^k q^{n-k} &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n!}{k!(n-k)!} \left(\frac{\lambda^k}{n^k}\right) \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
 &= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k} \frac{n(n-1)(n-2) \cdots (n-k+1)}{n^k} \\
 &= \left[\left(1 - \frac{\lambda}{n}\right)^n \frac{\lambda^k}{k!} \right] \left(1 - \frac{\lambda}{n}\right)^{-k} \frac{\overbrace{n(n-1)(n-2) \cdots (n-k+1)}^{k \text{ terms in numerator product}}}{n^k}
 \end{aligned}$$

Now, as $n \rightarrow \infty$ we have

$$\left(1 - \frac{\lambda}{n}\right)^n \xrightarrow[\substack{n \rightarrow \infty \\ p \rightarrow 0 \\ np \rightarrow \lambda}]{e^{-\lambda}}$$

$$\left(1 - \frac{\lambda}{n}\right)^{-k} \xrightarrow[\substack{n \rightarrow \infty \\ p \rightarrow 0 \\ np \rightarrow \lambda}]{1}$$

$$\frac{\overbrace{n(n-1)(n-2) \cdots (n-k+1)}^{k \text{ terms in numerator product}}}{n^k} \xrightarrow[\substack{n \rightarrow \infty \\ p \rightarrow 0 \\ np \rightarrow \lambda}]{1}$$

Putting it all together, we have

$$\binom{n}{k} p^k q^{n-k} \xrightarrow[\substack{n \rightarrow \infty \\ p \rightarrow 0 \\ np \rightarrow \lambda}]{e^{-\lambda} \frac{\lambda^k}{k!}}$$

as claimed.

Example 1-27: Suppose that 3000 parts are received. The probability p that a part is defective is 10^{-3} . Consider part defects to be independent events. Find the probability that there will be

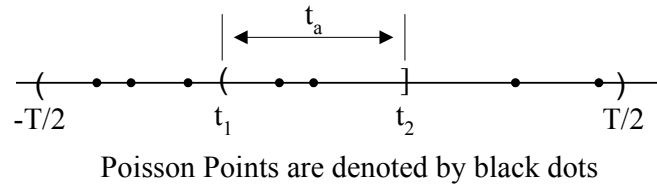


Figure 1-10: n random Poisson Points in $(-T/2, T/2)$.

more than five defective parts. Let k denote the number of defective parts, and note that $np = 3$.

Then

$$\mathbf{P}(k > 5) = 1 - \mathbf{P}(k \leq 5) = 1 - \sum_{k=0}^5 \binom{3000}{k} (10^{-3})^k (1 - 10^{-3})^{3000-k} .$$

$$\text{But } \mathbf{P}(k \leq 5) \approx e^{-3} \sum_{k=0}^5 \frac{3^k}{k!} = .916 \text{ so that } \mathbf{P}(k > 5) \approx .084 .$$

The function $\mathbf{P}(k) = e^{-\lambda} \frac{\lambda^k}{k!}$ is known as the *Poisson Function with parameter λ* .

Random Poisson Points

In a *random* manner, place n points in the interval $(-T/2, T/2)$. Denote by $\mathbf{P}(k \text{ in } t_a)$ the probability that k of these points will lie in an interval $(t_1, t_2] \subset (-T/2, T/2)$, where $t_a = t_2 - t_1$ (see Fig. 1-10). Find $\mathbf{P}(k \text{ in } t_a)$. First, note that the probability of placing a single point in $(t_1, t_2]$ is

$$p = \frac{t_2 - t_1}{T} = \frac{t_a}{T} . \tag{1-45}$$

Now, place n points in $(-T/2, T/2)$, and do it independently. The probability of finding k points in a sub-interval of length t_a is

$$\mathbf{P}(k \text{ in } t_a) = \binom{n}{k} p^k q^{n-k} ,$$

where $p = t_a/T$.

Now, assume that $n \rightarrow \infty$ and $T \rightarrow \infty$ such that $n/T \rightarrow \lambda_d$, a constant. Then, $np = n(t_a/T) \rightarrow t_a \lambda_d$ and

$$\mathbf{P}(k \text{ in } t_a) = \binom{n}{k} p^k q^{n-k} \xrightarrow[\substack{n \rightarrow \infty \\ T \rightarrow \infty \\ np = n(t_a/T) \rightarrow t_a \lambda_d}]{\quad} e^{-\lambda_d t_a} \frac{(\lambda_d t_a)^k}{k!} . \quad (1-46)$$

The constant λ_d is the *average point density* (the average number of points in a unit length interval).

In the limiting case, as $n \rightarrow \infty$ and $T \rightarrow \infty$ such that $n/T \rightarrow \lambda_d$, a constant, the points are known as **Random Poisson Points**. They are used to describe many arrival time problems, including those that deal with electron emission in vacuum tubes and semiconductors (*i.e.*, *shot noise*), the frequency of telephone calls, and the arrival of vehicle traffic.

Alternative Development of $\mathbf{P}(k \text{ in } t_a)$ as Expressed by (1-46)

We arrive at (1-46) in another way that gives further insight into Poisson points. As above, we consider the infinite time line $-\infty < t < \infty$, and we place an infinite number of points on this line where λ_d is the *average point density* (λ_d points per unit length, on the average).

To first-order in Δt , the probability of finding exactly one point in $(t, t + \Delta t]$ is $\lambda_d \Delta t$. That is, this probability can be formulated as

$$\mathbf{P}[\text{exactly one point in } (t, t + \Delta t]] = \lambda_d \Delta t + \text{Higher-Order Terms} , \quad (1-47)$$

where “Higher-Order Terms” are terms involving $(\Delta t)^2$, $(\Delta t)^3$, \dots . Also, we can express the probability of finding no points in $(t, t + \Delta t]$ as

$$\mathbf{P}[\text{no points in } (t, t + \Delta t]] = (1 - \lambda_d \Delta t) + \text{Higher-Order Terms} . \quad (1-48)$$

Consider the arbitrary interval $(0, t]$, $t > 0$ (nothing is gained here by assuming the more general case $(t_0, t]$, $t > t_0$). Denote as $p_k(t)$ the probability of finding exactly k points in $(0, t]$; we write

$$p_k(t) \equiv \mathbf{P}[\text{exactly } k \text{ points in } (0, t]]. \quad (1-49)$$

Now, k points in $(0, t + \Delta t]$ can happen in two mutually exclusive ways. You could have k points in $(0, t]$ and no point in $(t, t + \Delta t]$ or you could have $k-1$ points in $(0, t]$ and exactly one point in $(t, t + \Delta t]$. Formulating this notion in terms of p_k , (1-47) and (1-48), we write

$$p_k(t + \Delta t) = p_k(t)(1 - \lambda_d \Delta t) + p_{k-1}(t)(\lambda_d \Delta t) + \text{Higher-Order Terms} , \quad (1-50)$$

where Higher-Order Terms are those involving second-and-higher-order powers of Δt .

Equation (1-50) can be used to write the first-order-in- Δt relationship

$$\frac{p_k(t + \Delta t) - p_k(t)}{\Delta t} = \lambda_d [p_{k-1}(t) - p_k(t)], \quad (1-51)$$

where terms of order Δt and higher are omitted from the right-hand side. In (1-51), take the limit as Δt approaches zero to obtain

$$\frac{d}{dt} p_k(t) = \lambda_d [p_{k-1}(t) - p_k(t)], \quad (1-52)$$

an equation that can be solved for the desired $p_k(t)$.

Starting with p_0 , we can solve (1-52) recursively. With $k = 0$, Equation (1-52) becomes

$$\frac{d}{dt} p_0(t) = -\lambda_d p_0(t) \quad (1-53)$$

$$\lim_{t \rightarrow 0^+} p_0(t) = 1$$

(the probability is one of finding zero points in a zero length interval), so that $p_0(t) = \exp[-\lambda_d t]$.

Setting $k = 1$ in (1-52) leads to

$$\frac{d}{dt} p_1(t) + \lambda_d p_1(t) = \lambda_d e^{-\lambda_d t} \quad (1-54)$$

$$\lim_{t \rightarrow 0^+} p_1(t) = 0$$

(the probability is zero of finding one point in a zero length interval), so that $p_1(t) = (\lambda_d t) \exp[-\lambda_d t]$. This process can be continued to obtain

$$p_k(t) = e^{-\lambda_d t} \frac{(\lambda_d t)^k}{k!}, \quad k = 0, 1, 2, \dots, \quad (1-55)$$

a formula that satisfies (1-52) as can be seen from direct substitution. Note the equivalence of (1-55) and (1-46) (with an interval length of t instead of t_a). Poisson points arise naturally in applications where a large number of points are distributed at random and independently of one another (think of the large number of applications where Poisson point models can be applied!).

Poisson Points In Non-Overlapping Intervals

Consider again a $(-T/2, T/2)$ long interval that contains n points. Consider two **non-overlapping** subintervals of length t_a and t_b . See Figure 1-11 where points are denoted as black dots. We want to find the probability $\mathbf{P}(k_a \text{ in } t_a, k_b \text{ in } t_b)$ that k_a points are in interval t_a **and** k_b points are in interval t_b . Using the generalized Bernoulli trials formula developed previously, we claim that

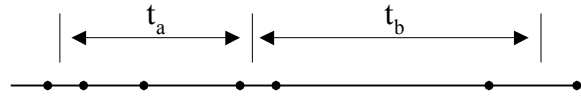


Figure 1-11: Poisson Points in non-overlapping intervals.

$$\mathbf{P}(k_a \text{ in } t_a, k_b \text{ in } t_b) = \frac{n!}{k_a! k_b! (n - k_a - k_b)!} \left(\frac{t_a}{T}\right)^{k_a} \left(\frac{t_b}{T}\right)^{k_b} \left(1 - \frac{t_a}{T} - \frac{t_b}{T}\right)^{n - k_a - k_b}. \quad (1-56)$$

Proof: This can be established by using the idea of a generalized Bernoulli Trial. The events

$A_1 = \{\text{point in } t_a\}$ with $\mathbf{P}(A_1) = t_a/T$,

$A_2 = \{\text{point in } t_b\}$ with $\mathbf{P}(A_2) = t_b/T$

$A_3 = \{\text{point outside } t_a \text{ and } t_b\}$ with $\mathbf{P}(A_3) = 1 - t_a/T - t_b/T$

form a disjoint partition of $(-T/2, T/2)$. The event $\{k_a \text{ in } t_a \text{ and } k_b \text{ in } t_b\}$ is equivalent to the event $\{A_1 \text{ occurs } k_a \text{ times, } A_2 \text{ occurs } k_b \text{ times, } A_3 \text{ occurs } n - k_a - k_b \text{ times}\}$. Hence, from the Generalized Bernoulli theory

$$\mathbf{P}(k_a \text{ in } t_a, k_b \text{ in } t_b) = \frac{n!}{k_a! k_b! (n - k_a - k_b)!} \left(\frac{t_a}{T}\right)^{k_a} \left(\frac{t_b}{T}\right)^{k_b} \left(1 - \frac{t_a}{T} - \frac{t_b}{T}\right)^{n - k_a - k_b} \quad (1-57)$$

as claimed.

Note that the events $\{k_a \text{ in } t_a\}$ and $\{k_b \text{ in } t_b\}$ **are not independent**. This intuitive result follows from the fact that

$$\begin{aligned}
\mathbf{P}(k_a \text{ in } t_a, k_b \text{ in } t_b) &= \frac{n!}{k_a! k_b! (n - k_a - k_b)!} \left(\frac{t_a}{T}\right)^{k_a} \left(\frac{t_b}{T}\right)^{k_b} \left(1 - \frac{t_a}{T} - \frac{t_b}{T}\right)^{n - k_a - k_b} \\
&\neq \left(\frac{n!}{k_a! (n - k_a)!} \left(\frac{t_a}{T}\right)^{k_a} \left(1 - \frac{t_a}{T}\right)^{n - k_a} \right) \left(\frac{n!}{k_b! (n - k_b)!} \left(\frac{t_b}{T}\right)^{k_b} \left(1 - \frac{t_b}{T}\right)^{n - k_b} \right) \quad (1-58) \\
&= \mathbf{P}(k_a \text{ in } t_a) \mathbf{P}(k_b \text{ in } t_b).
\end{aligned}$$

That is, the joint probability $\mathbf{P}(k_a \text{ in } t_a, k_b \text{ in } t_b)$ **does not** factor into $\mathbf{P}(k_a \text{ in } t_a) \mathbf{P}(k_b \text{ in } t_b)$.

The fact is intuitive that the events $\{k_a \text{ in } t_a\}$ and $\{k_b \text{ in } t_b\}$ are *dependent* for the finite case outlined above. Since the number n of points is finite, the more you put into the t_a interval the fewer you have to put into the t_b interval.

Limiting Case

Now, suppose that $n/T = \lambda_d$ and $n \rightarrow \infty$, $T \rightarrow \infty$. Note that $nt_a/T = \lambda_d t_a$, $nt_b/T = \lambda_d t_b$ so that

$$\begin{aligned}
&\frac{n!}{k_a! k_b! (n - k_a - k_b)!} \left(\frac{t_a}{T}\right)^{k_a} \left(\frac{t_b}{T}\right)^{k_b} \left(1 - \frac{t_a}{T} - \frac{t_b}{T}\right)^{n - k_a - k_b} \\
&= \frac{n(n-1) \cdots (n - k_a - k_b + 1)}{n^{k_a + k_b}} \left(1 - \lambda_d \frac{t_a + t_b}{n}\right)^{-k_a - k_b} \frac{(\lambda_d t_a)^{k_a}}{k_a!} \frac{(\lambda_d t_b)^{k_b}}{k_b!} \left(1 - \lambda_d \frac{t_a + t_b}{n}\right)^n
\end{aligned}$$

as $n \rightarrow \infty$ and $T \rightarrow \infty$

$$\frac{n(n-1) \cdots (n - k_a - k_b + 1)}{n^{k_a + k_b}} \xrightarrow[n/T \rightarrow \lambda_d]{\substack{n \rightarrow \infty \\ T \rightarrow \infty}} 1$$

$$\left(1 - \lambda_d \frac{t_a + t_b}{n}\right)^{-k_a - k_b} \xrightarrow[n \rightarrow \infty, T \rightarrow \infty, n/T \rightarrow \lambda_d]{} 1$$

$$\left(1 - \lambda_d \frac{t_a + t_b}{n}\right)^n \xrightarrow[n \rightarrow \infty, T \rightarrow \infty, n/T \rightarrow \lambda_d]{} e^{-\lambda_d(t_a + t_b)}$$

so that

$$\begin{aligned} \mathbf{P}(k_a \text{ in } t_a, k_b \text{ in } t_b) &\xrightarrow[n \rightarrow \infty, T \rightarrow \infty, n/T \rightarrow \lambda_d]{} e^{-\lambda_d t_a} \frac{(\lambda_d t_a)^{k_a}}{k_a!} e^{-\lambda_d t_b} \frac{(\lambda_d t_b)^{k_b}}{k_b!} \\ &= \mathbf{P}(k_a \text{ in } t_a) \mathbf{P}(k_b \text{ in } t_b) \end{aligned} \quad (1-59)$$

Thus, the events $\{k_a \text{ in } t_a\}$ and $\{k_b \text{ in } t_b\}$ **approach independence** in the limit, as described above.

In the limiting case, the points are distributed on the real axis, λ_d is the average number of points per unit length (the average point density), and the points are known as **Random Poisson Points**. Random Poisson Points are completely specified by

1. $\mathbf{P}(k_a \text{ in } t_a) = e^{-\lambda_d t_a} \frac{(\lambda_d t_a)^{k_a}}{k_a!}$
2. If intervals $(t_1, t_2]$ and $(t_3, t_4]$ are non-overlapping, the events $\{k_a \text{ in } (t_1, t_2]\}$ and $\{k_b \text{ in } (t_3, t_4]\}$ are independent.

Random Poisson points have been applied to many problems in science and engineering. They are used to describe arrival time problems such as the arrival of telephone calls and vehicle traffic. Also, they play a significant role in the development of the theory of *shot noise* in electron devices (vacuum tubes and semiconductors). This type of common noise results from

the random arrival of electrons at a vacuum tube anode or semiconductor junction (see Chapters 7 and 9 of these class notes for a discussion of shot noise).