

Chapter 1 Experiments and Probability

Chapter1

Contents

| Experiments and Probability | 2 |
|---|----|
| 1.1 Definition of an Experiment | 2 |
| 1.1.1 The Sample Space | 2 |
| 1.1.2 The Borel Field | 3 |
| 1.1.3 The Probability Measure | 3 |
| 1.2 Combined Experiments | 5 |
| 1.2.1 Cartesian Product of Two Experiments | 5 |
| 1.2.2 Cartesian Product of n Experiments | 7 |
| 1.2.3 Counting Experiments | 13 |
| 1.2.4 Selection Combined Experiment | 15 |
| 1.3 Conditional Probability | 16 |
| 1.3.1 Total Probability Theorem | 17 |
| 1.3.2 Bayes's Theorem | 18 |
| 1.4 Random Points | 19 |
| 1.4.1 Uniform Random Points in an Interval | 19 |
| 1.4.2 Nonuniform Random Points in an Interval | 21 |
| 1.5 Summary | 22 |



1 Experiments and Probability

1.1 Definition of an Experiment

To fully appreciate the meaning of probability and acquire a strong mathematical foundation for analytical work, it is necessary to define precisely the concept of an experiment and sample space mathematically. These definitions provide consistent methods for the assignment of elementary probabilities in paradoxical situations, and thus allow for meaningful calculation of probabilities of events other than the elementary events. Although at the beginning this approach may seem stilted, it will lead to a concrete concept of probability and an interpretation of derived probabilities.

An <u>experiment</u> E is specified by the three tuple (S, F, P(.)), where S is a finite, countable, or noncountable set called the <u>sample space</u>, F is a <u>Borel field specifying a set of events</u>, and $P(\cdot)$ is a probability measure allowing calculation of probabilities of all events.

1.1.1 The Sample Space

The <u>sample space</u> S is a set of elements called outcomes of the experiment E and the number of elements could be finite, countable, or noncountable infinite. For example, S could be the set containing the six faces of a die, $S = \{f_1, f_2, f_3, f_4, f_5, f_6\}$, or the positive integers, $S = \{i : i = 1, 2, ...\}$, or the real values between zero and one, $S = \{x : 0 < x < 1\}$, respectively.

An <u>event</u> is defined as any subset of S. On a single trial of the experiment an outcome is obtained. If that outcome is a member of an event, it is said that the event has occurred. In this way many different events occur at each trial of the experiment. For example, if f_1 is the outcome of a single trial of the experiment then the events $\{f_1\}$, $\{f_1, f_2\}$, $\{f_1, f_3\}$, ..., $\{f_1, f_6\}$, ..., $\{f_1, f_3, f_5\}$, ..., all occur. Events consisting of single elements, like f_1 , are called <u>elementary events</u>. The <u>impossible event</u> corresponds to the empty set Φ and never occurs, while the <u>certain event</u>, S, contains all outcomes and thus always occurs no matter what the outcome of the trial is. Events A and B are called <u>mutually exclusive</u> or <u>disjoint</u> if $A \cap B = \Phi$, where Φ is the **null set**.

Two events A and B are called <u>independent</u> if $P(A \cap B) = P(A) \cdot P(B)$. The events A_1, A_2, \dots, A_n are defined to be independent if the probabilities of all intersections two, three, ..., and n events can be written as products. This implies for all i, j, k, \dots , that the following conditions must be satisfied for independence

$$P(A_i \cap A_j) = P(A_i)P(A_j)$$

$$P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k)$$

$$\vdots$$

$$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1)P(A_2)...P(A_n)$$
(1.1)

1.1.2 The Borel Field

A <u>field</u> can be defined as a nonempty class of sets such that (1) if $a \in F$, then the complement of $a \in F$ and (2) if $a \in F$ and $b \in F$ then $a \bigcup b \in F$. Thus a field contains all finite unions, and by virtue of compliments and DeMorgan's theorem, all intersections of the collection. If we further require that all infinite unions and intersections are present in the collection, a <u>Borel field</u> is defined.

The set of all events of our experiment that will have probabilities assigned to them (measurable events) must be a Borel field to have mathematical consistency. If A, a collection of events, has a finite number of elements, a Borel field can be formed as the set of those events plus all possible subsets obtained by unions and intersections of those events including the null set Φ and entire set S.

If a set is noncountable, it is little harder to describe a Borel field. The most common Borel field, containing the real numbers, is the smallest Borel field containing the following intervals: $\{x : x \le x_1\}$ for all $x_1 \in$ real numbers. This will contain all finite and infinite closed and open intervals of the form [a,b],[a,b),(a,b], and (a,b), where a and b are real numbers and the intersections and unions of those intervals thereof.

1.1.3 The Probability Measure

The **probability measure**, $P(\cdot)$, must be a consistent assignment of probabilities such that the following conditions are satisfied (1) For any event $A \in F$, the probability of the event A, P(A), is such that $P(A) \ge 0$. (2) For the certain event, S, P(S)=1. (3) If A and B are any two events such that $A \cap B = \Phi$ then $P(A \cup B) = P(A) + P(B)$. (3a) If $A_i \in F$ for i = 1, 2, ..., and $A_i \cap A_i = \Phi$ for all $i \ne j$, then

$$D(A | |A | | | | |A | |) = D(A) + D(A) +$$

$$F(A_1 \cup A_2 \cup \dots \cup A_i \cup \dots) = F(A_1) + F(A_2) + \dots + F(A_i) + \dots$$

With these conditions satisfied, the probability of any event $A \in F$ can be calculated. How does one go about assigning probabilities such that we satisfy the conditions above?

For the case where S is a set with a finite number of elements the conditions above can be satisfied



by assigning probabilities to all the events with only single outcomes, $\{e_i\}$, where $e_i \in S$, such that conditions (1) and (2) above are satisfied. This mapping from the sample space S to the positive reals is called the <u>distribution function</u> for the probability measure and equivalently specifies the probability measure.

When S is a set with a noncountably infinite number of elements the assignment above is not useful as the probabilities of most elementary events will be zero. In this case the assignment of probabilities is consistent if the probabilities of the events $\{x : x \le x_1\}$ for all $x_1 \in$ real numbers are assigned such that

- (1) $0 \le P\{x : x \le x_1\} \le 1$ (bounded).
- (2) $P\{x: x \le x_2\} \ge P\{x: x \le x_1\}$ for all $x_2 > x_1$ (nondecreasing function of x) and

(3) lim as $\varepsilon \to 0$ of $P\{x : x \le x_1 + \varepsilon\}$ equals $P\{x : x \le x_1\}$ (continuous from the right side).

This mapping: $x \to P\{x : x \le x_1\}$, defined for all x, is called the <u>cumulative distribution</u> function and equivalently describes the probability measure for the noncountable case. From this distribution function we are able to calculate all probabilities of events that are members of the Borel field F. The cumulative distribution function could have also been used to specify the probability measure for the case where S has a countable or finite number of elements.

A number of examples of experiments will now be presented. They will include a couple of coin-tossing experiments and a die-rolling experiment. The experiments will be described by specifying their sample space, Borel field, and probability measures.

Example 1.1

This experiment consists of a single flipping of a coin that results in either a head or a tail showing. Give its description by specifying as $(S, F, P(\cdot))$.

Solution

The possible outcomes of the experiment are either a head or a tail. Thus the sample space can be described as the set $S = \{head, tail\}$.

The Borel field F consists of the elementary events $\{head\}$ and $\{tail\}$, the null set Φ , and S.

To complete the description of the experiment, a probability measure must be assigned. This particular assignment could be based on previous experience, careful experimentation, use of favorable to total alternatives, or any other interpretation of the concept of probability. There is no right or wrong assignment, but certain assignments (models) may be more appropriate in explaining the results of corresponding physical experiments. For the purpose of this example, we assume that this coin has been tampered with for more often a head comes up than a tail, which we specify by $P\{head\} = p$ and $P\{tail\} = 1 - p$. These two assignments comprise the distribution function and thus the probability measure $P(\cdot)$ for the

Chapter1



experiment.

Example 1.2

A more realistic assignment for the experiment of flipping a coin could be the experiment $(S, F, P(\cdot))$ defined as follows: $S = \{head, tail, edge\}$ with $P(\cdot)$ described by P(h) = 0.49, P(t) = 0.49, P(e) = 0.02. The Borel field F is defined as the power set of S.

(a) Identify all the events (elements) of the Borel field. (b) Calculate the probabilities of those events.

Solution

(a) The Borel field F specifying the measurable events is the power set of S (all possible subsets of S) given by

$$F = \{\Phi, \{h\}, \{t\}, \{e\}, \{h, t\}, \{h, e\}, \{t, e\}, \{h, t, e\}\}$$

- Where Φ is the impossible event, and $\{h, t, e\}$ is the certain event. The other events consist of all possible proper subsets of S, single elements, and combinations of two elements.
- (b) By definition $P(\Phi) = 0$ and the probabilities of the elementary events are given in the specification of the probability measure of the experiment as P(h) = 0.49, P(t) = 0.49, and P(e) = 0.02. The probabilities of the other events can be determined by repeated application of property (3) for the probability measure. For example the events $\{h\}$ and $\{e\}$ are mutually exclusive therefore $P\{h,e\}$ can be found as follows:

$$P\{h, e\} = P(\{h\} \cup \{e\}) = P\{h\} + P\{e\}$$
$$= 0.49 + 0.02 = 0.51$$

Similarly $P\{h,t\} = 0.98$, $P\{t,e\} = 0.51$, and $P\{h,t,e\} = 1$.

1.2 Combined Experiments

Combined experiments play an important role in probability theory applications. There are many ways we can combine experiments, including cartesian products in which independent trials of the same or different experiments can be described. In some cases the probabilities of events will depend on the results of previous trials of experiments or random selection of different experiments. A number examples of combined experiments are now explored beginning with the classical case of sampling with replacement.

1.2.1 Cartesian Product of Two Experiments

Consider the case of having two separate experiments specified by the following: $E_1: (S_1, F_1, P_1(.))$ and

 $E_2: (S_2, F_2.P_2(.))$. The sample spaces S_1 and S_2 are usually different sets, for example, results of a coin toss and results of a die roll, but they could be the same sets representing separate trials of the same experiment as in repeated coin-tossing experiments. We can define a new combined experiment by using the cartesian product concept as $E = E_1 \otimes E_2$, where the new sample space $S = S_1 \otimes S_2$ is the cartesian product of the two sample spaces expressible by the ordered pair of elements where the first element is from S_1 and the second is from S_2 .

Example 1.4

Let experiment $E_1: (S_1, F_1, P_1(\cdot))$ and $E_2: (S_2, F_2, P_2(\cdot))$ be defined as follows: E_1 is the experiment of flipping a coin with outcomes head (h) and tail (t) with equal probability of occurrence. E_2 is the experiment of random selection of a colored ball from a box with outcomes red (r), white (w), and blue (b) with replacement. Define the new experiment E as $E = E_1 \otimes E_2$ with experiment E_1 and E_2 being performed independently of each other; that is, the outcome of experiment E_1 in no way effects the outcome of E_2 , and vice versa. Set up a reasonable model for this new experiment.

Solution

To specify the model it suffices to give S, F, $P(\cdot)$ of the new experiment E.

The new sample space S is the Cartesian product of the two experiments and given by $S = S_1 \otimes S_2$. Elements of S are ordered pairs with the first element coming from $S_1 = \{h, t\}$ and the second from $S_2 = \{r, w, b\}$; therefore

$$S = \{(h,r), (h,w), (h,b), (t,r), (t,w), (t,b)\}$$

The new Borel field F is selected as the power set of S, that is all possible unions and intersections of S. This includes the null set, the entire set, and all possible pairs, triples, and so on, as shown below:

$$F = \begin{cases} \Phi, \{(h, r)\}, \{(h, w)\}, \{(h, b)\}, \{(t, r)\}, \{(t, w)\}, \{(t, b)\} \\ \{(h, r), (h, w)\}, \{(h, r), (h, b)\}, \{(h, r), (t, r)\}, \{(h, r), (t, w)\}, \dots \\ \{(h, r), (h, w), (h, b)\}, \{(h, r), (h, w), (t, r)\}, \{(h, r), (h, w), (t, w)\}, \dots \\ \vdots \\ \{(h, r), (h, w), (h, b), (t, r), (t, w), (t, b)\} \end{cases}$$

The probability measure $P(\cdot)$ can be described by specifying the probability of the elementary events. Once these are known, the probability of any event can be found by writing the event as a union of those events and using property (3). Since the experiments are independent, it is reasonable to assign probabilities of the elementary events as a product of the probabilities from each experiment. For example, $P\{(h,r)\} = P\{h\} \cdot P\{r\}$. If head and tail are equally probable in E_1 then it is reasonable for $P_1(\cdot)$ to be



described by P(h) = P(t) = 1/2, and if we have reason to believe that red, white, and blue are not equally probable in E_2 , then $P_2(\cdot)$ could be given by P(r) = 0.5, P(w) = 0.3, P(b) = 0.2. Thus, for this example, the probability measure $P(\cdot)$ can be described by specifying the probabilities of the elementary events (the distribution function):

$$P\{(h,r)\} = P\{(h)\} \cdot P\{(r)\} = 0.25 \qquad P\{(t,r)\} = P\{(t)\} \cdot P\{(r)\} = 0.25 P\{(h,w)\} = P\{(h)\} \cdot P\{(w)\} = 0.15 \qquad P\{(t,w)\} = P\{(t)\} \cdot P\{(w)\} = 0.15 P\{(h,b)\} = P\{(h)\} \cdot P\{(b)\} = 0.1 \qquad P\{(t,b)\} = P\{(t)\} \cdot P\{(b)\} = 0.1$$

The event of a head in the new experiment is $H = \{(h, r), (h, w), (h, b)\}$, and its probability of occurrence can be determined using property (3) as

$$P(H) = P\{(h, r), (h, w), (h, b)\} = P\{\{(h, r)\} \cup \{(h, w)\} \cup \{(h, b)\}\}$$

= 0.25 + 0.15 + 0.1 = 0.5

1.2.2 Cartesian Product of n Experiments

Consider the case of having *n* separate experiments specified by $E_k : (S_k, F_k, P_k(.))$ for k = 1, 2, ..., n. Define a new combined experiment E : (S, F, P(.)) as a cartesian product: $E = E_1 \otimes E_2 \otimes \cdots \otimes E_n$ where the new sample space $S = S_1 \otimes S_2 \otimes \cdots \otimes S_n$ is the cartesian product of the *n* spaces and expressible by the ordered n-tuples of elements whose first element is from S_1 and the second is from $S_2 \dots$, the *n*th from S_n . The E_k are, in general, different, but in many cases the experiment could be formed from independent trials of the same experiment. Also there is an important class of problems where the experiments are the same, yet they cannot be thought of as independent. A good example of this is the random selection of outcomes without replacement. Examples of each type are now presented.

Binomial Distribution. Consider the experiment $E_1: (S_1, F_1, P_1(.))$ where the outcomes of the experiment are either failure indicated by a 0 or a success indicated by a 1; therefore $S_1 = \{0,1\}$. Assume that the probability measure $P_1(.)$ for the experiment is given by $P\{success\} \stackrel{\Delta}{=} P\{1\} = p$ and $P\{failure\} \stackrel{\Delta}{=} P\{0\} = 1 - p$ and that the F_1 is the set $\{\Phi, \{0\}, \{1\}, \{0,1\}\}$. Define a new experiment by $E = E_1 \otimes E_1 \otimes \cdots \otimes E_1$ where E: (S, F, P(.)) describes the new experiment. Assume that this represents independent trials of the same experiment E_1 where the probability of success or failure is the same for each trial.

The S, F, P(.) are now described for this new experiment. The new S is the cartesian product

 $S = S_1 \otimes S_1 \otimes \cdots \otimes S_1$ and consists of all possible n-tuples where the elements are either 0 or 1 as shown below:

$$S = \begin{cases} \begin{pmatrix} 0, & 0, & \cdots & 0, & 0 \end{pmatrix} \\ \begin{pmatrix} 0, & 0, & \cdots & 0, & 1 \end{pmatrix} \\ \begin{pmatrix} 0, & 0, & \cdots & 1, & 0 \end{pmatrix} \\ \begin{pmatrix} 0, & 0, & \cdots & 1, & 1 \end{pmatrix} \\ \vdots \\ \begin{pmatrix} 1, & 1, & \cdots & 1, & 1 \end{pmatrix} \end{cases}$$
(1.2)

The new probability measure is specified once the distribution function or probabilities of the elements of S are determined. By virtue of the independent experiment assumption, the probability of each elementary event of the new experiment is the product of the probabilities for the elementary events in the single trial. For example, the $P\{(0,1,1,0,\ldots,0,1,1)\}$ is given by

$$P\{(0,1,1,0,\ldots,0,1,1)\} = P\{0\} \cdot P\{1\} \cdot P\{0\} \cdot \ldots \cdot P\{0\} \cdot P\{1\} \cdot P\{1\} = p^4(1-p)^{n-4}$$
(1.3)

As a matter of fact every sequence that has only four ones (successes) will have this same probability. The total number of these sequences is the combination of n things taken four at a time, since we have n locations and we want only four of them with ones. The probability distribution function for the new experiment can then be determined as follows, where the first entry is the elementary sequence and after the arrow is the corresponding probability:

Outcome Probablity $(0,0,...,0,0) \to P(0,0,...,0,0) = P(0) \cdot P(0) \cdots P(0) \cdot P(0) = p^{0}(1-p)^{n}$ $(0,0,...,0,1) \to P(0,0,...,0,1) = P(0) \cdot P(0) \cdots P(0) \cdot P(1) = p^{1}(1-p)^{n-1}$ $(0,0,...,1,0) \to P(0,0,...,1,0) = P(0) \cdot P(0) \cdots P(1) \cdot P(0) = p^{1}(1-p)^{n-1}$ (1.4) (1.4) (1.4) (1.4)

The Borel field will be the power set associated with S and specifies all events for which probabilities are assigned.

Probabilities of different type of events for the above example can be determined by using the distribution function described. There are a wide number of applications as a success can mean all kinds of things. For example, a success could be obtaining an ace in drawing a card from a standard deck of cards with replacement, or obtaining successful reception of a binary symbol from a random communication channel.

The event of exactly k successes out of n independent trials appears frequently in physical situations and its probability will now be derived using the results above. Define the event A_1 as the event of exactly one success out of n trials. From the results above we see that the probability of exactly one success out of n trials is **《Random Signal Processing》**

$$P(A_{1}) = P(exactly \ one \ success \ out \ of \ n \ trials)$$

$$= P(\{(0,0,...,0,1)\} \cup \{(0,0,...,1,0)\} \cup ... \cup \{(1,0,...,0,0)\})$$

$$= P\{(0,0,...,0,1)\} + P\{(0,0,...,1,0)\} + \dots + P\{(1,0,...,0,0)\}$$

$$= \binom{n}{1} p^{1} (1-p)^{n-1}$$
(1.5)

Similarly the probability of exactly k successes out of n trials can be found to be

$$P(A_k) = P(exactly \ k \ success \ out \ of \ n \ trials) = \binom{n}{k} p^k (1-p)^{n-k}$$
(1.6)

In some problems we may wish to know the probability that the number of successes out of n trials is within a range of values. The probability that the number of successes k is in the range $m \le k \le n$ can be obtained by adding the probabilities of exactly k successes for that range, since the events A_k and A_j are mutually exclusive events for all k and j such that $k \ne j$. Therefore

$$P(J \le number \ of \ successes \le K) = P\left(\left(\begin{array}{c}exactly \ J\\successes\end{array}\right) + \left(\begin{array}{c}exactly \ J+1\\successes\end{array}\right) + \dots + \left(\begin{array}{c}exactly \ K\\successes\end{array}\right)\right)$$

$$= P\left(\bigcup_{k=J}^{K} A_{k}\right) = \sum_{k=J}^{K} \binom{n}{k} p^{k} (1-p)^{n-k}$$

$$(1.7)$$

Example 1.5

Consider the experiment of tossing a fair coin with the sample space $S = \{h, t\}$ and probability distribution $P\{h\} = 0.6$ and $P\{t\} = 0.4$. Suppose the experiment is performed 10 times independently to define a new cartesian product experiment.

- (a) Determine the probability that we get exactly 5 heads in the 10 trials.
- (b) Determine the probability that we get greater than 7 heads in the 10 trials.
- (c) Determine the probability that we get less than or equal 9 heads in the 10 trials.
- (d) Determine the probability that the number of tails is greater than or equal to 4 and less than or equal 5.

Solution

The desired probabilities are determined for Eqs. (1.6) and (1.7) as follows:

(a)
$$P\{A_1\} = P(exactly \ 5 \ heads \ out \ of \ 10 \ trials)$$

$$= \binom{10}{5} (0.6)^5 (0.4)^{10-5} = 0.20066$$
(b) $P\{8 \le \frac{number}{of \ heads} \le 10\} = P\left(\binom{exactly}{8 \ heads} + \binom{exactly}{9 \ heads} + \binom{exactly}{10 \ heads}\right)$

$$= \binom{10}{8} (0.6)^8 (0.4)^2 + \binom{10}{9} (0.6)^9 (0.4)^1 + \binom{10}{10} (0.6)^{10} (0.4)^0$$
$$= 0.1269790$$

(c)
$$P(\text{less than or equal 9 heads out of 10 trials})$$

= $1 - P(\text{exactly 10 heads out of 10 trials})$
= $1 - {\binom{10}{10}} (0.6)^{10} (0.4)^0 = 0.99395$
(d) $P\left\{4 \le \frac{\text{number}}{\text{of heads}} \le 5\right\} = P\left(\binom{\text{exactly}}{4 \text{ heads}} + \binom{\text{exactly}}{5 \text{ heads}}\right)$
= ${\binom{10}{4}} (0.6)^4 (0.4)^6 + {\binom{10}{5}} (0.6)^5 (0.4)^5 = 0.1315425$

Approximations for Binomial Probabilities. The probabilities of k successes in n trials of a Bernoulli experiment was found to be as in Eq. (1.6) and evaluating probabilities of ranges of successes is given in Eq. (1.7). In Figure 1.1 the $P(k \ success \ out \ of \ n \ trials)$ is plotted for values of n equal to 5, 10, and 50 for all values of k, and a P(success) = 0.6. The graphs show a tendency toward a hill-shaped curve similar to a Gaussian function. These plots would be symmetrical around the point 0.5 n if p = 0.5. However, when p does not equal 0.5, as for the cases given, the plot is not quite symmetrical. If the number of trials is large, the calculation load due to the factorials is considerable, and certain approximations to these probabilities become useful. Many of these approximations are good provided that the number of trials n, number of successes k, and probability of success p satisfy a given set of conditions. Of these approximations we will present the DeMoivre and Poisson approximations are reliable.

<u>DeMoivre-Laplace Approximation</u>. For np(1-p) >> 1 and |k-np| of the order of $\sqrt{np(1-p)}$,

$$\binom{n}{k} p^{k} (1-p)^{n-k} \approx \frac{1}{\sqrt{2\pi n p (1-p)}} \exp\left\{-\frac{(k-np)^{2}}{2n p (1-p)}\right\}$$
(1.8)

<u>Poisson Approximation</u>. For $n \gg 1$, $p \ll 1$, and np of order 1,

$$\binom{n}{k} p^k (1-p)^{n-k} \approx e^{(-np)} \frac{(np)^k}{k!}$$
(1.9)

Approximation of Regions of Successes in Bernoulli Trials. Say that we are interested in approximating the probability that the number of successes k in n repeated trials of a Bernoulli experiment is in the

«Random Signal Processing»

range $k_1 \le k \le k_2$. If this range of successes contains values that satisfy the DeMoivre-Laplace approximation, then the summation can be approximated by using the error function or the $\phi(x)$ function as follows:

Chapter1

Experiments and Probability

$$\sum_{k=k_{2}}^{k_{2}} \binom{n}{k} p^{k} (1-p)^{n-k} \approx \phi \left(\frac{k_{2} - np}{\sqrt{np(1-p)}} \right) - \phi \left(\frac{k_{1} - np}{\sqrt{np(1-p)}} \right)$$
(1.10)
where
$$\phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^{2}/2} dy$$

The $\phi(x)$ can only be determined by numerical integration as there is no antiderivative and it is convenient to use the table given in some book appendix contents.



Figure 1.1 Plot of P(k successes out of n trails) for p=0.6 and n=5,10, 50.

<u>Multinomial Distribution</u>. A combined experiment that is an extension of the Bernoulli trial experiment just described, which resulted in the binomial distribution, is the case of multiple occurrences of several different events on multiple trials of the same experiment. Consider an experiment $E_1: (S_1, F_1, P_1(.))$. Define a set of events A_1 , i = 1 to k, consisting of elements of S_1 such that their union is the certain event, they are pairwise disjoint, and their probabilities are given by $P(A_1) = p_i$ for i = 1 to k.

Define a new experiment E by $E = E_1 \otimes E_1 \otimes \ldots \otimes E_1$ described by E: (S, F, P(.)). Assume that this represents n independent trials of the same experiment E_1 . Let n_i , i = 1 to k, be the number of times event A_i occurs in the new experiment, which is composed of n repeated trials. Also assume that $n_i \ge 0$ and $n_1 + n_2 + \cdots + n_k = n$.

It can be shown that the probability that A_i occurs exactly n_i times in the *n* trials is

$$P\begin{pmatrix}A_{1} & occurs & exactly & n_{1} & times, \\ A_{2} & occurs & exactly & n_{2} & times, \\ & \vdots & & \\ A_{k} & occurs & exactly & n_{k} & times, \end{pmatrix} = \frac{n!}{n_{1}!n_{2}!\cdots n_{k}!} p_{1}^{n_{1}} p_{2}^{n_{2}} \cdots p_{k}^{n_{k}}$$
(1.15)

Hypergeometric Distribution. In the previous compound experiments the trials of the experiment were considered to be independent. A very important class of experimentation problems is sampling done without replacement, for a finite number of elements, and without independent trials. For example, if a is the number of successful (s) elements and b is the number of failure (f) elements and n elements are drawn at random but not replaced, the corresponding outcomes of the experiment do not contain all possible *n*-tuples of *s* and *f*, So the n-tuple $\{s, s, ..., s\}$ and sequences containing more than a successes are not possible.

It can be shown for this experiment that the probability of k successes out of n trials can be determined as

$$P(k \ successes \ in \ n \ trials) = \frac{\binom{a}{k}\binom{b}{n-k}}{\binom{a+b}{n}} , \qquad k = 0,1,\dots$$
(1.16)

The next example is typical of the type of problems for which the hypergeometric formula above can be used.



1.2.3 Counting Experiments

A special class of compound experiments deal with successive application of experiments where the total experiment may be stopped at any point depending on the results from the current experiment. Two common distributions result from these compound experiments.

<u>Geometric Distribution</u>. The geometric distribution is a result of another form of compound experiment. Define a countable number of identical experiments $E_k : (S_k, F_k, P_k(.))$ for k = 1, 2, ... Each of these experiments will be of the Bernoulli type (two possible outcomes) previously discussed, with identical probabilities of success; the result of the experiment, if performed, is independent of previous or future experiments. The new experiment is described as follows: Perform experiment E_1 if a success is obtained, then stop; if a success does not happen, then perform experiment E_2 ; if a success occurs, then stop otherwise. Continue this process until a success occurs. Thus the number of trials is not the same each time the compound experiment is performed, and thus the sample space is not a cartesian product.

If p is the probability of success on each experiment, then the sample space, Borel field and probability measure can be described as follows: The elementary events are sequences of 1's (for success) and 0's for failures but with the property that they end on a 1 and only have 0's preceding the 1. Thus S can be described by

$$S = \{(1), (0,1), (0,0,1), \dots, (0,0,\dots,0,1), \dots\}$$
(1.17)

Notice that (1,0), (0,1,0), (1,1), etc., are not elements of the sample space since the new experiment would stop at a 1 and not proceed to the next one. Events that are collections of outcomes are not of the same sequence length. For example, $\{(1), (0,1)\}$ is an event where a success is obtained in less than or equal to 2 trials.

The probability measure P(.) can be specified by giving the distribution function for the elementary events. The $P\{(1)\}$ is the probability of getting a success on the very first trial. Since the probability of a success on the first experiment is p, the probability of getting $\{(1)\}$ is also p,

$$P\{(1)\} = p \tag{1.18}$$

The probability of getting $\{(0,1)\}$ equals the probability of getting a failure on performing the first experiment and getting a success on the performance of the second experiment. Since results of each experiment are independent if performed, the $P\{(0,1)\}$ can be written as a product:

$$P\{(0,1)\} = (1-p)p \tag{1.19}$$

Similarly the probability of the elementary event consisting of a sequence of k-1 failures (0's) followed by a single success (1) is

《Random Signal Processing》

$$P\{(0,0,\ldots,0,1)\} = (1-p)^{k-1}p$$
(1.20)
$$(k-1)zeros$$

These probabilities as $k \to \infty$ determine the distribution function, or equivalently the probability measure P(.). It can be shown that

$$\sum_{k=1}^{\infty} (1-p)^{k-1} p = 1$$
(1.21)

The Borel field is again defined as the power set of S, that is, all possible subsets of S. The probabilities of arbitrary events that are the subsets of S can then be found by adding up the probabilities of the elementary elements in the event. For example, define the event A to be the set of outcomes such that a success occurs on the performance of an odd number of experiments. Thus A can be written as

$$A = \{(1), (0,0,1), (0,0,0,0,1), \dots, (0^{(2k)}, 1), \dots\}$$
(1.22)

The probability of A can be determined by adding up the probabilities of the elementary events and using the properties of a geometric sequence as follows.

$$P(A) = P\{(1), (0,0,1), (0,0,0,0,1), \dots, (0^{(2k)}, 1), \dots\}$$

= $\sum_{k=0}^{\infty} (1-p)^{2k} p = p\left(\frac{1}{1-(1-p)^2}\right) = \frac{1}{2-p}$ (1.23)

Negative Binomial Distribution. An extension of the geometric experiment is a compound experiment in which the number of trials necessary to obtain k successes is desired rather than the number of trials needed for a single success. The basic underlying experiment is to define a countable number of identical experiments $E_k : (S_k, F_k, P_k(.))$ for k = 1, 2, ... Each of these experiments will be of the Bernoulli type with identical probabilities of success, and the result of the experiment, if performed, is independent of previous or future experiments. The new experiment is described as follows: Perform experiment E_i if a success is obtained and it is the *k*th, then stop; if not, continue to the next experiment E_{i+1} . Continue this process until the *k*th success occurs.

Let x be the number of trials in which the *k*th success occurs; then the probability of that event can be shown to be

P(event that the kth success occurs on the xth trial)

$$= \binom{x-1}{k-1} p^{k} (1-p)^{x-k} \qquad x = k, k+1, k+2, \dots$$
(1.24)

The following example illustrates the type of problems that will use the negative Binomial distribution given above.

Example:

《Random Signal Processing》

Suppose the probability of getting a f_2 on the single toss of a die is 0.2. Find (a) the probability that the fourth f_2 occurs on the 6th trail; (b) that the fifth f_2 occurs before the 7th trail.

Solution:

(a)

$$P(event that the 4th success occurs on the 6th trial) = \binom{6-1}{4-1} 0.2^4 (1-0.2)^{6-4} = 0.02048$$

(b)

$$P(event that the 5th success occurs before the 7th trial) = P\begin{pmatrix} (event that the 5th success occurs on the 5th trial) \\ \cup (event that the 5th success occurs on the 6th trial) \\ = \binom{5-1}{5-1} 0.2^5 (1-0.2)^{5-5} + \binom{6-1}{5-1} 0.2^5 (1-0.2)^{6-5} = 0.02048$$

1.2.4 Selection Combined Experiment

A different type of combined experiment will now be considered that is a combination of three or more experiments. For purpose of illustration only three component experiments are presented. Let the three experiments be defined as

$$E_0: (S_0, F_0, P_0(.)), \qquad E_1: (S_1, F_1, P_1(.)), \qquad E_2: (S_2, F_2, P_2(.))$$

The sample space for E_0 consists of only two outcomes $(e_1 \text{ and } e_2)$. These outcomes serve to help us select which one of the other experiments will be performed. If the outcome is e_1 , then E_1 is performed with outcome α_i ; if the outcome is e_2 then E_2 is performed with outcome β_i . Thus the combination experiment E:(S, F, P(.)) can be described as follows: The sample space S will consist of ordered pairs of elements the first either e_1 or e_2 and the second either $\alpha_i \in S_1$ or $\beta_i \in S_2$. Thus a trial of the combined experiment results in an outcome $e_c = (e, \delta)$, where $e \in \{e_1, e_2\}$ and $\delta \in S_1 \cup S_2$.

The Borel field F will be defined to be the set of all possible subsets of S, and the associated probability measure can be described as

$$P(\{(e_i, \delta_j)\}) = P_0(e_i) \cdot P(\delta_j)$$
where
$$P(\delta_j) = P_1(\alpha_j) \quad if \quad e_i = e_1 \quad or$$

$$P(\delta_j) = P_2(\beta_j) \quad if \quad e_i = e_2$$

$$(1.27)$$

The product of the probabilities is because of the assumption that the experiment E_0 is independent of



the experiments E_1 and E_2 .

1.3 Conditional Probability

The probability measure allows us to calculate the probabilities of all events that are members of the Borel field for the defined experiment. It will become useful to define the concept of conditional probability of events. Given a conditioning event C such that P(C) > 0, the <u>conditional probability of any event A</u> assuming C is defined as

$$P(A \mid C) \stackrel{\scriptscriptstyle \Delta}{=} \frac{p(A \cap C)}{P(C)} \tag{1.28}$$

The following examples consider the calculation of conditional probabilities for discrete and continuous sample spaces. It is seen that the fundamental set operations and the probability measure for the underlying experiment are used to obtain conditional probabilities.

Example 1.11

Consider an experiment defined as a single toss of a "crooked" die with sample space $S = \{f_1, f_2, f_3, f_4, f_5, f_6\}$. The probability measure, maybe based on past history, is known to be

$$P\{f_1\} = \frac{1}{2} \qquad P\{f_2\} = \frac{1}{4} \qquad P\{f_3\} = \frac{1}{8}$$
$$P\{f_4\} = \frac{1}{16} \qquad P\{f_5\} = \frac{1}{32} \qquad P\{f_6\} = \frac{1}{32}$$

Let the conditioning event *C* be given as $C = \{f_1, f_3, f_5\}$, the face of the die is odd, and calculate the P(A | C) where $A = \{f_1, f_2\}$, the face is less than or equal 2.

Solution

By the definition of conditional probability, Eq. (1.28), P(A | C) is

$$P(A \mid C) = \frac{P(A \cap C)}{P(C)} = \frac{P(\{f_1, f_2\} \cap \{f_1, f_3, f_5\})}{P(\{f_1, f_3, f_5\})}$$
$$= \frac{P(\{f_1\})}{P(\{f_1, f_3, f_5\})} = \frac{1/2}{1/2 + 1/8 + 1/32} = \frac{16}{21}$$

Example 1.13

Define an experiment that has as outcomes, t, the set of real numbers greater than or equal to zero and that the outcome represents the time until a certain device fails. The probability measure for the experiment is described as



Experiments and Probability

for $x \ge 0$

$$P(failure time t \le x) = 1 - e^{-x}$$

The Borel field is given by the smallest field containing the intervals $\{t: 0 \le t \le x\}$ for all $x \ge 0$. Find the *P*(*failure time* $t \le 2 \mid$ *failure time* t > 1).

Solution

From the definition of conditional probability the probability that the time to failure is less than or equal 2 for the device given that the failure time is greater than 1 is

$$P(\{failure \ t \le 2\} \mid \{failure \ t > 1\}) = \frac{P(\{failure \ t \le 2\} \cap \{failure \ t > 1\})}{P(\{failure \ t > 1\})}$$

Using set operation gives

{failure
$$t \le 2$$
} \cap {failure $t > 1$ } = {failure $1 < t \le 2$ }

And the denominator is seen to be

$$P(\{failure \ t > 1\}) = 1 - P(\{failure \ t \le 1\})$$
$$= 1 - (1 - e^{-1}) = e^{-1}$$

Substituting these two results into the first equation gives us the following result

$$P(\{failure \ t \le 2\} | \{failure \ t > 1\}) = \frac{P(\{failure \ 1 < t \le 2\})}{P(\{failure \ t > 1\})} = \frac{P(\{failure \ 1 < t \le 2\})}{P(\{failure \ t > 1\})} = \frac{(1 - e^{-2}) - (1 - e^{-1})}{e^{-1}} = 1 - e^{-1}$$

1.3.1 Total Probability Theorem

Given *n* events A_1, A_2, \ldots, A_n such that

$$A_i \cap A_j = \Phi$$
 for all $i \neq j$ (mutually exclusive)
 $\bigcup_{i=1}^n A_i = S$ (exhaustive) (1.31)

Then it can be shown that the **total probability** of an arbitrary event B can be written in terms of the following conditional probabilities as

$$P(B) = P(B | A_1)P(A_1) + P(B | A_2)P(A_2) + \dots + P(B | A_n)P(A_n)$$

= $\sum_{i=1}^{n} P(B | A_i)P(A_i)$ (1.32)



1.3.2 Bayes's Theorem

A very important theorem that has many applications is **Bayes's theorem** which involves the determination of conditional probabilities $P(A_k | B)$ under the same framework as above. It is easily shown that

$$P(A_{k} | B) = \frac{p(A_{k} \cap B)}{P(B)}$$

$$= \frac{P(B | A_{k})P(A_{k})}{P(B | A_{1})P(A_{1}) + P(B | A_{2})P(A_{2}) + \dots + P(B | A_{n})P(A_{n})}$$

$$= \frac{P(B | A_{k})P(A_{k})}{\sum_{i=1}^{n} P(B | A_{i})P(A_{i})}$$
(1.33)

The communication receiver and radar signal detector are the typical examples that are representative of the type of problem that can be solved using Bayes's theorem and the total probability theorem.

Example 1.14

Consider an experiment involving a random selection of one of three boxes. The random selection is of a single ball from the box chosen. The boxes contain red, white, and blue balls with specified probabilities of selection. Assume that

$$P(box1) = 0.5$$
 $P(box2) = 0.3$ $P(box3) = 0.2$ $P(red | box1) = 0.4$ $P(red | box2) = 0.5$ $P(red | box3) = 0.1$ $P(white | box1) = 0.3$ $P(white | box2) = 0.3$ $P(white | box3) = 0.2$ $P(blue | box1) = 0.3$ $P(blue | box2) = 0.2$ $P(blue | box3) = 0.7$

- (a) Find the probability of getting a red ball.
- (b) Find the conditional probability, P = (A | B), where B is the event box 2 selected and A is the event a red ball is selected.

Solution

(a) The total probability theorem can be used to obtain the probability of getting a red ball as follows:

$$P(red) = P(red | box1)P(box1) + P(red | box2)P(box2) + P(red | box3)P(box3)$$

= 0.4.0.5 + 0.3.0.5 + 0.1.0.2 = 0.37

(b) The probability of box 2 given that the ball is red can be determined by using Bayes's theorem and the results of part (a) as

$$P(box2 \mid red) = \frac{P(red \mid box2)P(box2)}{P(red)}$$
$$= \frac{0.5 \cdot 0.3}{0.37} = 0.4054$$



1.4 Random Points

The random placement of points in an interval is an important problem. Conceptually it is analogous to random arrival times used in basic inventory problems, and it is used as a basis for shot noise in communication theory. The random placement of the points can be uniformly or nonuniformly distributed on an interval as seen in the following sections.

1.4.1 Uniform Random Points in an Interval

Define an experiment E:(S, F, P(.)) as the random placement of a point t somewhere in the closed interval [0,T] as shown in Figure 1.3(a). The sample space is $S = \{t: 0 \le t \le T\}$, F is the smallest field containing the sets $\{t: t \le t_1\}$ for all $t_1 \in S$, and P(.) is defined by $P\{t: t \le t_1\} = t_1/T$, for all $t_1 \in S$. Thus it can be seen that the probability that the point selected will be in any given interval is the ratio of that interval's length to the total length T. Probabilities of other events that are unions of nonoverlapping intervals can be obtained by adding up the probabilities for each of the intervals.



(c) k points in interval t_a out of n points

Figure 1.3 Random times in interval [0, T]

The purpose of this section is to talk about the random placement of n points, not just one point, in an interval [0,T] see Figure 1.3(b). A convenient way to design such an experiment is to form a new compound experiment $E_n: (S_n, F_n, P_n(.))$ composed of ordered n-tuples of times, $(t_1, t_2, ..., t_n)$, obtained from repeating the experiment E defined above independently. Assume an independence of the trials so that the probability measure can be described as the product

$$P\{t_1 \le x_1, t_2 \le x_2, \dots, t_n \le x_n\} = P\{t_1 \le x_1\}P\{t_2 \le x_2\}\dots P\{t_n \le x_n\}$$
(1.34)

《Random Signal Processing》

(Randor

The Borel field is defined to be the smallest field containing the events $\{t_1 \le x_1, t_2 \le x_2, \dots, t_n \le x_n\}$ for all $x_1, x_2, \dots, x_n \in [0, T]$. We are interested in answering questions relating to calculating the probabilities that a certain number of points fall in a given interval or intervals.

Say that the probability that exactly k of the points fall in a given interval $(t_1, t_2]$ of length t_a as shown in Figure 1.3(c) is desired. Let A be the event that on a single trial the point selected at random falls in the given interval. Using the uniform probability measure P(.), we give the probability of A by $P(A) = (t_2 - t_1) = t_a$ Thus, on n independent trials, the probability of getting k points in the interval is binomially distributed as

$$P(k \text{ point in } (t_1, t_2) \text{ for } n \text{ trials}) = \binom{n}{k} p^k (1-p)^{n-k}$$
where $p = \frac{t_a}{T}$
(1.35)

Further assume that n, the number of trials, is very large, $n \gg 1$, and that the relative width of the interval is very small, $t_a/T \ll 1$, and k is of the order nt_a/T , The resulting Poisson approximation is written as

$$P\begin{pmatrix} exactly & point & in & (t_1, t_2) \\ out & of & n & trials \end{pmatrix} \approx \exp\left(\frac{-nt_a}{T}\right) \frac{(nt_a/T)^k}{k!}$$
(1.36)
where $t_a = t_2 - t_1$

In the limiting case this result will give an interpretation in terms of an average number of points per unit interval. If $n \to \infty$, $T \to \infty$, and $n/T \to \lambda$, then it can be shown that

$$\lim_{t_a \to 0} \frac{P(exactly \ 1 \ point \ in \ t_a)}{t_a} = \lambda$$
(1.37)

Another probability of interest is that of getting exactly k_a points in interval t_a and exactly k_b in interval t_b as indicated in Figure 1.4.



Figure 1.4 Exactly ka points in interval ta and exactly kb in interval tb

Let A, B, C equal the events that on a single trial of the experiment exactly one point falls in t_a , t_b , and not in t_a or t_b , respectively. Then the probability of getting exactly k_a points in interval t_a , exactly k_b points in interval t_b , and exactly $n - k_a - k_b$ points not in t_a or t_b out of n trials can

«Random Signal Processing»



be obtained from the multinomial result as

$$P\begin{pmatrix}k_{a} & in \ t_{a} \\ k_{b} & in \ t_{b} \\ n-k_{a}-k_{b} \notin t_{a} & or \ t_{b}\end{pmatrix} = \frac{n!}{k_{a}!k_{b}!(n-k_{a}-k_{b})!} \left(\frac{t_{a}}{T}\right)^{k_{a}} \left(\frac{t_{b}}{T}\right)^{k_{b}} \left(1-\frac{t_{a}}{T}-\frac{t_{b}}{T}\right)^{n-k_{a}-k_{b}}$$
(1.38)

The individual events exactly k_a points in t_a and exactly k_b points in t_b out of n trials have probabilities as follows:

$$P(k_{a} \quad in \quad t_{a}) = \binom{n}{k_{a}} \left(\frac{t_{a}}{T}\right)^{k_{a}} \left(1 - \frac{t_{a}}{T}\right)^{n-k_{a}}$$

$$P(k_{a} \quad in \quad t_{a}) = \binom{n}{k_{a}} \left(\frac{t_{b}}{T}\right)^{k_{b}} \left(1 - \frac{t_{b}}{T}\right)^{n-k_{b}}$$
(1.39)

Since the product of the above two probabilities does not equal that of Eq. (1.38), the events exactly k_a points in t_a and exactly k_b points in t_b out of n trials are not independent events.

1.4.2 Nonuniform Random Points in an Interval

In certain problems the random points are not placed uniformly in the interval. A common way to describe a nonuniform rate is to assign the probability measure by using a weighting function $\alpha(t)$ that satisfies the following properties:

$$\alpha(t) \ge 0 \quad for \quad all \quad t \in [0,T]$$

$$\int_0^T \alpha(t) dt = 1$$
(1.40)

On a single trial of the experiment, the random placement of a single point in the interval [0,T], the probability that the point selected is in $\{t_1, t_2\}$ is given by

$$P\{t \in (t_1, t_2]\} = \int_{t_1}^{t_2} \alpha(t) dt$$
(1.41)

Thus, if $\alpha(t)$ has a peak at $t = t_1$, it means that the point selected at random has a higher probability of being close to t_1 than other values of t. If n independent trials of this experiment are performed, the probability of exactly k points out of n trials being in the interval $\{t_1, t_2\}$ can be determined from Eq. (1.6) as

$$P\begin{pmatrix} k & point & in & (t_1, t_2) \\ out & of & n & trials \end{pmatrix} = \binom{n}{k} p^k (1-p)^{n-k}$$

$$where \quad p = P\{t \in (t_1, t_2]\} = \int_{t_1}^{t_2} \alpha(t) dt$$
(1.42)

For the case of a nonuniform rate with the assumption that n, the number of trials, is very large, n >> 1,

and that the relative width of the interval $t_a = (t_1, t_2)$ is very small, $t_a/T \ll 1$, and k of the order nt_a/T , the Poisson approximation results in the following probability where p is given as above:

$$P\begin{pmatrix} exactly \ k \ point \ in \ (t_1, t_2) \\ out \ of \ n \ trials \end{pmatrix} \approx \exp(-np) \frac{(np)^k}{k!}$$
(1.43)

1.5 Summary

In this chapter the mathematical definitions of an experiment in terms of a sample space, a Borel field, and a probability measure were given. The concept of an experiment is basic to the understanding of random variables and random processes to be discussed in later chapters. The assignment of the probability measure for several experiments obtained by combining other experiments led to the binomial, multinonnal, and hypergeometric distributions.

The important concept of conditional probability, the total probability theorem and the Bayes theorem are defined. Using this definition of the total probability theorem and the Bayes theorem for obtaining the a posteriori probability followed. These concepts have a fundamental role in the detection and estimation of random variables and random processes as will be seen in Chapters 2, 8, and 9.

The chapter concluded with a short discussion on the random placement of points in a given interval. These experiments are important in analyzing problems involving random times of arrival and other related problems.

It was not the intent of this chapter to give an exhaustive presentation on experiments and their use but to provide background material that would be used in the remainder of the class. For those wishing a more thorough presentation there are many excellent texts, such as the book, <**Probability, Random Variables**, and Stochastic Processes>(Papoulis, Athanasios, McGraw Hill, 1965).

-----This is the end of Chapter01-----