# Notes for A First Course in Probability

**CHAPTER 1: COMBINATORIAL ANALYSIS**

****

1. The basic principle of counting

**The basic principle of counting**

Suppose that two experiments are to be performed. Then if experiment 1 can result in any one of *m* possible outcomes and if, for each outcome of experiment 1, there are *n* possible outcomes of experiment 2, then together there are *mn* possible outcomes of the two experiments.

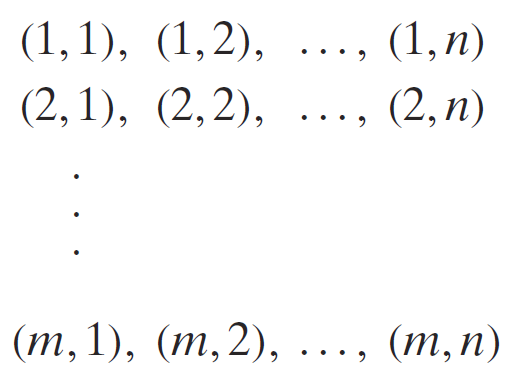


Figure 1. Proof of the Basic Principle

1. Permutation (排列)(fit)

* How many different ordered arrangements of the letters *a*, *b*, and *c* are possible? By direct enumeration we see that there are 6, namely, *abc*, *acb*, *bac*, *bca*, *cab*, and *cba*. Each arrangement is known as a permutation. (has the order of selection)

Suppose now that we have n objects. Reasoning similar to that we have just used for the 3 letters then shows that there are

different permutations of the n objects.

Whereas *n*! (read as “n factorial”) is defined to equal 1∙2…*n* when *n* is a positive integer, it is convenient to define 0! To equal 1.

In general, the same reasoning as that used in Example 3d (Page 17) shows that there are

different permutations of *n* objects, of which are alike, are alike, , are alike.

1. Combination (组合)(out)

* Page 18 (no the order of selection)

We define , for , by

and say that (read as “n choose r”) represents the number of possible combinations of *n* objects taken *r* at a time.

* Example 4b and 4c (Page 19)
* A useful combinatorial identity (Page 20)
* The binomial theorem (Page 20)

like

1. Multinomial Coefficients

* Groups have order but their members have not order.

**Notation**

If , we define by

Thus, represents the number of possible divisions of *n* distinct

objects into r distinct groups of respective sizes .

**CHAPTER 2: COMBINATORIAL ANALYSIS**

****

1. Sample space and events

* The *sample space* of the experiment and is denoted by ***S.***

Like

* is called the *union* of the event E and the event F, which consists of all outcomes that are either in *E* or in *F* or in both *E* and *F*.

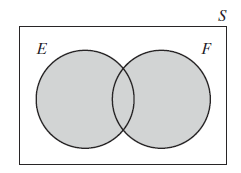


Figure 2. Venn diagram:

* is called the *intersection* of *E* and *F*, which consists of all outcomes that are both in E and in *F*.

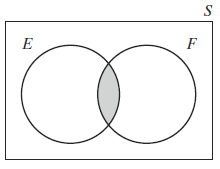


Figure 3. Venn diagram:

* is the null event.
* is the union of events-
* is the event consisting of those outcomes that are in all of the events
* , referred to as the complement of . ()

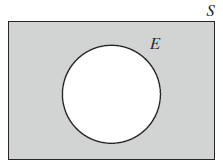


Figure 4. Venn diagram:

* : *E* is a subset of *F*.

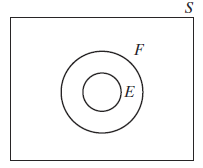


Figure 5. Venn diagram:

* : *E* and *F* are equal.
* Rules:

Commutative laws

Associative laws

Distributive laws

* *DeMorgan’s* laws:

1. Axioms of probability

* We define to be the number of times in the first repetitions of the experiment that event occurs. Then , the probability of the event , is defined as

It is the limiting *relative frequency* of .

* The three axioms of probability

**The three axioms of probability**

***Axiom 1***

***Axiom 2***

***Axiom 3***

For any sequence of mutually exclusive events (that is, events for which when

We refer to as the probability of the event .

1. Some simple propositions
2. Probability as a continuous set function

* A sequence of events is said to be an increasing sequence if

Whereas it is said to be a decreasing sequence if

If is an increasing sequence of events, then we define a new event, denoted by

, by

Similarly, if is a decreasing sequence of events, we define, by

* If is either an increasing or a decreasing sequence of events, then

**CHAPTER 3: CONDITIONAL PROBABILITY AND INDEPENDENCE**

****

1. Conditional probabilities

* Suppose that we toss 2 dice, and suppose that each of the 36 possible outcomes is equally likely to occur. If we let *E* and *F* denote, respectively, the event that the sum of the dice is 8 and the event that the first die is a 3, then the probability just obtained is called the *conditional probability* that *E* occurs given that *F* has occurred and is denoted by

Now, since we know that F has occurred, it follows that F becomes our new, or reduced, sample

space; hence, the probability that the event EF occurs will equal the probability of EF relative to the

probability of F.

**Definition**

If , then

1. Bayes’s formula

* Let *E* and *F* be events. We may express *E* as

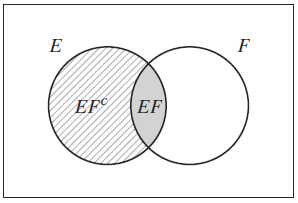


Figure 6.

As and are clearly mutually exclusive, we have, by Axiom 3,

which can be used to compute by “conditioning” on whether F occurs.

It may be generalized as follows: Suppose that are mutually exclusive events such that

In other words, exactly one of the events must occur. By writing

and using the fact that the events are mutually exclusive, we obtain

**The law of total probability (Page 82)**

* The odds of an event *A* (Page 81)

**Definition**

The odds of an event *A* are defined by

That is, the odds of an event *A* tell how much more likely it is that the event *A*

occurs than it is that it does not occur. For instance, if , then , so the odds are 2. If the odds are equal to α, then it is common to say

that the odds are “α to 1” in favor of the hypothesis.

* *Bayes’s formula* (Page 82)

**Proposition**

1. Independent events

* In the special cases where does in fact equal , we say that *E* is independent of *F*. That is, *E* is independent of *F* if knowledge that *F* has occurred does not change the probability that *E* occurs.

Since , it follows that *E* is independent of *F* if

**Definition**

Two events *E* and *F* are said to be *independent* if the above holds.

Two events *E* and *F* that are not independent are said to be *dependent*.

* If *E* and *F* are independent, then so are *E* and .

**Definition**

Three events *E*, *F*, and *G* are said to be independent if

1. is a probability

* Conditional probabilities satisfy all of the properties of ordinary probabilities, which shows that satisfies the three axioms of a probability.

1. .
2. If are mutually exclusive events, then

**CHAPTER 4: RANDOM VARIABLES**

****

1. Random variables

* When an experiment is performed, we are frequently interested mainly in some function of the outcome as opposed to the actual outcome itself. These real-valued functions defined on the sample space, are known as *random variables*.
* Because the value of a random variable is determined by the outcome of the experiment, we may assign probabilities to the possible values of the random variable.

1. Discrete random variables

* A random variable that can take on at most a countable number of possible values is said to be discrete. For a discrete random variable ***X***, we define the *probability* *mass function* of ***X*** by

The probability mass function is positive for at most a countable number of values of . That is, if ***X*** must assume one of the values then

for

for all other values of

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

The cumulative distribution function ***F*** can be expressed in terms of by

1. Expected values

* The expectation of a random variable. If ***X*** is a discrete random variable having a probability mass function , then the *expectation*, or the *expected* *value*, of ***X***, denoted by , is defined by

1. Expectation of a function of a random variable

* To compute the expected value of some function of ***X***, say, . Since is itself a discrete random variable, it has a probability mass function, which can be determined from the probability mass function of X.
* If ***X*** is a discrete random variable that takes on one of the values , with respective probabilities , then, for any real-valued function ,
* If *a* and *b* are constants, then

is also referred to as the *mean* or the *first moment of* ***X***. The quantity , is called the *nth moment* of ***X***.

1. Variance

* One possible way to measure this variation would be to consider the quantity , where. However, it turns out to be mathematically inconvenient to deal with this quantity, so a more tractable quantity is usually considered—namely, the expectation of the square of the difference between ***X*** and its mean.

**Definition**

If is a random variable with mean , then the variance of , denoted by , is defined by

The variance of is equal to the expected value of minus the square of its expected value. In practice, this formula frequently offers the easiest way to compute

* A useful identity is that for any constants and ,
* The square root of the is called the *standard deviation* of , and we denote it by . That is,

1. The Bernoulli and binomial random variables

* A random variable is said to be a *Bernoulli random variable* (after the Swiss mathematician James Bernoulli) if its probability mass function is given by

for some .

* A *binomial random variable* with parameters represents trials, each of them is a Bernoulli random variable with parameters . The probability mass function of a binomial random variable having parameters is given by
* Properties of binomial random variables

If is a binomial random variable with parameters and , then

* If is a binomial random variable with parameters , where , then as goes from 0 to, first increases monotonically and then decreases monotonically, reaching its largest value when is the largest integer less than or equal to.

Hence, if and only if

or equivalently, if and only if

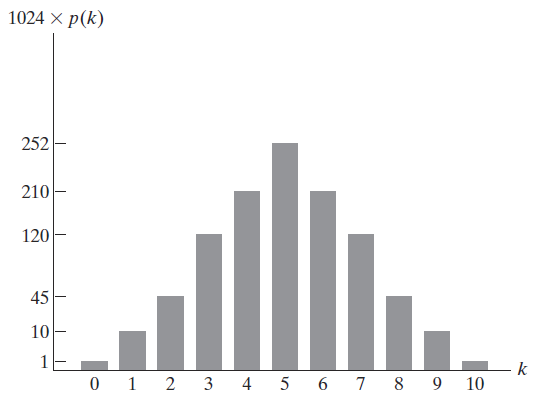


Figure 7. Graph of

* Suppose that is binomial with parameters . The key to computing its distribution function

is to utilize the following relationship between and , which was established in last proof:

1. The Poisson random variable

* A random variable that takes on one of the values is said to be a *Poisson* random variable with parameter if, for some ,