Error Correcting Codes: Combinatorics, Algorithms and Applications(Fall 2007)Lecture 4: Probability and Discrete Random Variables

Wednesday, January 21, 2009

Lecturer: Atri Rudra

Scribe: Anonymous

1 Counting and Probability

This lecture reviews elementary combinatorics and probability theory. We begin by first reviewing elementary results in counting theory, including standard formulas for counting permutations and combinations. Then, the axioms of probability and basic facts concerning probability distributions are presented.

2 Counting

Counting theory tries to answer the question "How many?" or "How many orderings of n distinct elements are there?" In this section, we review the elements of counting theory. A set of items that we wish to count can sometimes be expressed as a union of disjoint sets or as a Cartesian product of sets. The *rule of sum* says that the number of ways to choose an element from one of two *disjoint* sets is the sum of the cardinalities of the sets. That is, if A and B are two finite sets with no members in common, then $|A \cup B| = |A| + |B|$. The *rule of product* says that the number of ways to choose the first element times the number of ways to choose the second element. That is, if A and B are two finite sets, then $|A \times B| = |A| \cdot |B|$.

A string over a finite set S is a sequence of elements of S. We sometimes call a string of length k a k-string. A substring s' of a string s is an ordered sequence of consecutive elements of s. A k-substring of a string is a substring of length k. For example, 010 is a 3-substring of 01101001 (the 3-substring that begins in position 4), but 111 is not a substring of 01101001. A k-string over a set S can be viewed as an element of the Cartesian product S^k of k-tuples; thus, there are $|S|^k$ strings of length k. For example, the number of binary k-strings is 2^k . Intuitively, to construct a k-string over an n-set, we have n ways to pick the first element; for each of these choices, we have n ways to pick the second element; and so forth k times. This construction leads to the k-fold product $n \cdot n \cdots n = n^k$ as the number of k-strings.

A *permutation* of a finite set S is an ordered sequence of all the elements of S, with each element appearing exactly once. For example, if $S = \{a, b, c\}$, there are 6 permutations of S:

abc, acb, bac, bca, cab, cba.

There are n! permutations of a set of n elements, since the first element of the sequence can be chosen in n ways, the second in n - 1 ways, the third in n - 2 ways, and so on.

A *k-permutation* of S is an ordered sequence of k elements of S, with no element appearing more than once in the sequence. Thus, an ordinary permutation is just an *n*-permutation of an n-set. The twelve 2-permutations of the set $\{a, b, c, d\}$ are

$$ab, ac, ad, ba, bc, bd, ca, cb, cd, da, db, dc,$$

where we have used the shorthand of denoting the 2-set $\{a, b\}$ by ab, and on on. The number of k-permutations of an n-set is

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

since there are n ways of choosing the first element, n - 1 ways of choosing the second element, and so on until k elements are selected, the las being a selection from n - k + 1 elements.

A *k*-combination of an *n*-set *S* is simply a *k*-subset of *S*. There are six 2-combinations of the 4-set $\{a, b, c, d\}$:

We can construct a k-combination of an n-set by choosing k distinct elements from the n-set. The number of k-combinations of an n-set can be expressed in terms of the number of k-permutations of an n-set. For every k-combination, there are exactly k! permutations of its elements, each of which is a distinct k-permutation of the n-set. Thus the number of k-combinations of an n-set is the number of k-permutations divided by k!; from Equation (1), this quantity is

$$\frac{n!}{k!(n-k)!}.$$
(2)

For k = 0, this formula tells us that the number of ways to choose 0 elements from an *n*-set is 1 (not 0), since 0! = 1.

We use the notation $\binom{n}{k}$ (read "*n* choose *k*") to denote the number of *k*-combinations of an *n*-set. From Equation (2), we have

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$
(3)

This formula is symmetric in k and n - k:

$$\binom{n}{k} = \binom{n}{n-k}.$$
(4)

These two numbers are known as *binomial coefficients*, due to their appearance in the *binomial expansion*:

$$(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}.$$
 (5)

A special case of the binomial expansion occurs when s = y = 1:

$$2^n = \sum_{k=0}^n \binom{n}{k}.$$
(6)

This formula corresponds to counting the 2^n binary *n*-strings by the number of 1's they contain: there are $\binom{n}{k}$ binary *n*-strings containing exactly *k* 1's, since there are $\binom{n}{k}$ ways to choose *k* out of the *n* positions in which to place the 1's.

We sometimes need to bound the size of a binomial coefficient. For $1 \le k \le n$, we have the lower bound

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots1}$$
(7)

$$= \left(\frac{n}{k}\right) \left(\frac{n-1}{k-1}\right) \cdots \left(\frac{n-k+1}{1}\right) \tag{8}$$

$$\geq \left(\frac{n}{k}\right)^k.\tag{9}$$

Taking advantage of the inequality $k! \ge (k/e)^k$, we obtain the upper bounds

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots1}$$
(10)

$$\leq \frac{n^k}{k!} \tag{11}$$

$$\leq \left(\frac{en}{k}\right)^k.$$
(12)

For all $0 \le k \le n$, we can use induction to prove the bound

$$\binom{n}{k} \le \frac{n^n}{k^k (n-k)^{n-k}},\tag{13}$$

where for convenience we assume that $0^0 = 1$. For $k = \lambda n$, where $0 \le \lambda \le 1$, this bound can be rewritten as

$$\binom{n}{\lambda n} \leq \frac{n^n}{(\lambda n)^{\lambda n} ((1-\lambda)n)^{(1-\lambda)n}}$$
(14)

$$=\left(\left(\frac{1}{\lambda}\right)^{\lambda}\left(\frac{1}{1-\lambda}\right)^{1-\lambda}\right)^{\lambda} \tag{15}$$

$$=2^{nH(\lambda)},\tag{16}$$

where

$$H(\lambda) = -\lambda \lg \lambda - (1 - \lambda) \lg (1 - \lambda)$$
(17)

is the (binary) entropy function and where, for convenience, we assume that $0 \cdot lg(0) = 0$, so that H(0) = H(1) = 0.

3 Probability

This section reviews basic probability theory. We define probability in terms of a *sample space* S, which is a set whose elements are called *elementary events*. Each elementary event can be viewed as a possible outcome of an experiment. For the experiment of flipping two distinguishable coins, we can view the sample space as consisting of the set of all possible 2-strings over $\{H, T\}$:

$$S = \{HH, HT, TH, TT\}$$

An *event* is a subset¹ of the sample space S. For example, in the experiment of flipping two coins, the event of obtaining one head and one tail is $\{HT, TH\}$. The event S is called the *certain event*, and the event \emptyset is called the *null event*. We say that two events A and B are *mutually exclusive* if $A \cap B = \emptyset$. We sometimes treat an elementary event $s \in S$ as the event $\{s\}$. By definition, all elementary events are mutually exclusive.

A *probability distribution* $Pr[\cdot]$ on a sample space S is a mapping from events of S to real numbers such that the following *probability axioms* are satisfied:

- 1. $Pr[A] \ge 0$ for any event A.
- 2. Pr[S] = 1.
- 3. $Pr[A \cup B] = Pr[A] + Pr[B]$ for any two mutually exclusive events A and B. More generally, for any (finite or countably infinite) sequence of events A_1, A_2, \ldots that are pairwise mutually exclusive,

$$Pr\left[\bigcup_{i} A_{i}\right] = \sum_{i} Pr\left[A_{i}\right]$$

We call Pr[A] probability of the event A. We note here that axiom 2 is a normalization requirement: there is really nothing fundamental about choosing 1 as the probability of the certain event, except that it is natural and convenient. Several results follow immediately from these axioms and basic set theory. The null event \emptyset has probability $Pr[\emptyset] = 0$. If $A \subseteq B$, then $Pr[A] \leq Pr[B]$. Using \overline{A} to denote the event S - A (the *complement* of A), we have $Pr[\overline{A}] = 1 - Pr[A]$. For any two events A and B,

$$Pr[A \cup B] = Pr[A] + Pr[B] - Pr[A \cap B]$$
(18)

$$\leq \Pr\left[A\right] + \Pr\left[B\right]. \tag{19}$$

The generalization of Equation (18) is also known as the *union bound* and it is written as

$$Pr\left[\bigcup_{i=1}^{n} A_i\right] \leq \sum_{i=1}^{n} Pr\left[A_i\right]$$

¹For a general probability distribution, there may be some subsets of the sample space S that are not considered to be events. This situation usually arises when the sample space is uncountably infinite. The main requirement is that the set of events of a sample space be closed under the operations of taking the complement of an event, forming the union of a finite or countable number of events, and taking the intersection of a finite or countable number of events. Most of the probability distributions are over finite or countable sample spaces, and in general, all subsets of a sample space are considered to be events. With the notable exception of the continuous probability distribution.

4. Markov's inequality:

Theorem 3.1. For any nonnegative random variable X and any t > 0,

$$Pr[X \ge t] \le \frac{\mathbb{E}[X]}{t}.$$

Proof.

$$\mathbb{E}\left[X\right] = \int_0^\infty x f_X(x) dx \ge \int_t^\infty x f_X(x) dx \ge t \int_t^\infty f_X(x) dx.$$

5. Suppose a random variable X takes values X = i, where i = 0, 1, 2, ... with probabilities

$$Pr\left[X=i\right]=P_X(i).$$

Define

$$u(n-k) = \begin{cases} 1 & n \ge k, \\ 0 & \text{otherwise} \end{cases}$$

It follows, that

$$Pr [X \ge k] = \sum_{n=k}^{\infty} P_X(n)$$
$$= \sum_{n=0}^{\infty} P_X(n)u(n-k)$$
$$\le \sum_{n=0}^{\infty} P_X(n)e^{t(n-k)} \text{ for } t \ge 0.$$

The last line follows from the fact that

$$e^{t(n-k)} \ge u(n-k)$$
 for $t \ge 0$.

Next, let $\theta_X(t) = P_X(n)e^{tn}$ and note that

$$\sum_{n=0}^{\infty} P_X(n)e^{t(n-k)} = e^{-tk}P_X(n)e^{tn}$$
$$= e^{-tk}\theta_X(t).$$

Hence, we have established the important result

$$Pr\left[X \ge k\right] \le e^{-tk}\theta_X(t).$$

The *Chernoff bound* is determined by minimizing $e^{-tk}\theta_X(t)$:

$$Pr[X \ge k] \le \min_{t \ge 0} e^{-tk} \theta_X(t).$$

Going back to the coin-flipping example, suppose that each of the four elementary events has probability 1/4. Then the probability of getting at least one head is

$$Pr[HH, HT, TH] = Pr[HH] + Pr[HT] + Pr[TH] = \frac{3}{4}$$

Alternatively, since the probability of getting strictly less than one head is Pr[TT] = 1/4, the probability of getting at least one head is 1 - 1/4 = 3/4. A probability distribution is discrete if it is defined over a finite or countably infinite sample space. Let S be the sample space. Then for any event A,

$$Pr[A] = \sum_{s \in A} Pr[s],$$

since elementary events, specifically those in A, are mutually exclusive. If S is finite and every elementary event $s \in S$ has probability

$$Pr\left[s\right] = \frac{1}{|S|},$$

then we have the *uniform probability distribution* on S. In such case the experiment is often described as "picking an element of S at random." As an example, consider the process of flipping a *fair coin*, one for which the probability of obtaining head is the same as the probability of obtaining a tail, that is, 1/2. If we flip the coin n times, we have the uniform probability distribution defined on the sample space $S = \{H, T\}^n$, a set of size 2^n . Each elementary event in S can be represented as a string of length n over $\{H, T\}$, and each occurs with probability $1/2^n$. The event

$$A = \{ \text{exactly } k \text{ heads and exactly } n - k \text{ tails occur} \}$$

is a subset of S of size $|A| = \binom{n}{k}$, since there are $\binom{n}{k}$ strings of length n over $\{H, T\}$ that contain exactly k H's. The probability of event A is thus $Pr[A] = \binom{n}{k}/2^n$.

Sometimes we have some prior partial knowledge about the outcome of an experiment. For example, suppose that a friend has flipped two fair coins and has told you that at least one of the coins showed a head. what is the probability that both coins are heads? The information given eliminates the possibility of two tails. the three remaining elementary events are equally likely, so we infer that each occurs with probability 1/3. Since only one of these elementary events shows two heads, the answer to our question is 1/3.

Conditional probability formalizes the notion of having prior partial knowledge of the outcome of an experiment. The *conditional probability* of an event A given that another event B occurs is defined to be

$$Pr[A|B] = \frac{Pr[A \cap B]}{Pr[B]}$$
(20)

whenever $Pr[B] \neq 0$. We read "Pr[A|B]" as "the probability of A given B." Intuitively, since we are given that event B occurs, the event that A also occurs is $A \cap B$. That is, $A \cap B$ is the set of outcomes in which both A and B occur. Since the outcome is one of the elementary events in B, we normalize the probabilities of all the elementary events in B by dividing them by Pr[B], so that

they sum to 1. The conditional probability of A given B is, therefore, the ratio of the probability of event $A \cap B$ to the probability of event B. In the example above, A is the event that both coins are heads, and B is the event that at least one coin is a head. Thus

$$Pr[A|B] = (1/4)/(3/4) = 1/3.$$

Two events are "independent" if

$$Pr[A \cap B] = Pr[A] Pr[B],$$

which is equivalent, if $Pr[B] \neq 0$, to the condition

$$Pr\left[A|B\right] = Pr\left[A\right].$$

For example, suppose that two fair coins are flipped and that the outcomes are independent. Then the probability of two heads is (1/2)(1/2) = 1/4. Now suppose that one event is that the first coin comes up heads and the other event is that the coins come up differently. Each of these events occurs with probability 1/2, and the probability that both events occur is 1/4; thus, according to the definition of independence, the events are independent–even though one might think that both events depend on the first coin. Finally, suppose that the coins are welded together so that they both fall heads or both fall tails and that the two possibilities are equally likely. The the probability that each coin comes up heads is 1/2, but the probability that they both come up heads is $1/2 \neq (1/2)(1/2)$. Consequently, the event that one comes up heads and the event that the other comes up heads are not independent.

A collection A_1, A_2, \ldots, A_n of events is said to be *pairwise independent* if

$$Pr\left[A_i \cap A_j\right] = Pr\left[A_i\right] Pr\left[A_j\right]$$

for all $1 \le i < j \le n$. We say that they are *(mutually) independent* if every k-subset $A_{i_1}, A_{i_2}, \ldots, A_{i_n}$ of the collection, where $2 \le k \le n$ and $1 \le i_1 < i_2 < \cdots < i_k \le n$, satisfies

$$Pr\left[A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}\right] = Pr\left[A_{i_k}\right] Pr\left[A_{i_2}\right] \cdots Pr\left[A_{i_k}\right].$$

For example, suppose we flip two fair coins. Let A_1 be the event that the first coin is heads, let A_2 be the event that the second coin is heads, and let A_3 be the event that the two coins are different. We have

$$Pr [A_1] = 1/2,$$

$$Pr [A_2] = 1/2,$$

$$Pr [A_3] = 1/2,$$

$$Pr [A_1 \cap A_2] = 1/4,$$

$$Pr [A_1 \cap A_3] = 1/4,$$

$$Pr [A_2 \cap A_3] = 1/4,$$

$$Pr [A_1 \cap A_2 \cap A_3] = 0.$$

Since for $1 \le i < j \le 3$, we have $Pr[A_i \cap A_j] = Pr[A_i]Pr[A_j] = 1/4$, the events A_1, A_2 , and A_3 are pairwise independent. The events are not mutually independent, however, because $Pr[A_1 \cap A_2 \cap A_3] = 0$ and $Pr[A_1]Pr[A_2]Pr[A_3] = 1/8 \ne 0$. From the definition of conditional probability (20), it follows that for two events A and B, each with nonzero probability,

$$Pr[A \cap B] = Pr[B]Pr[A|B]$$
(21)

$$= Pr[A] Pr[B|A].$$
⁽²²⁾

Solving for Pr[A|B], we obtain

$$Pr[A|B] = \frac{Pr[A]Pr[B|A]}{Pr[B]},$$
(23)

which is known as *Bayes's theorem*. The denominator Pr[B] is a normalizing constant that we can express as follows. Since $B = (B \cap A) \cup (B \cap \overline{A})$ and $B \cap A$ and $B \cap \overline{A}$ are mutually exclusive events,

$$Pr[B] = Pr[B \cap A] + Pr[B \cap \overline{A}]$$
$$= Pr[A] Pr[B|A] + Pr[\overline{A}] Pr[B|\overline{A}].$$

Substituting into Equation (23), we obtain an equivalent form of Bayes's theorem:

$$Pr[A|B] = \frac{Pr[A]Pr[B|A]}{Pr[A]Pr[B|A] + Pr[\overline{A}]Pr[B|\overline{A}]}.$$

Bayes's theorem can simplify the computing of conditional probabilities. For example, suppose that we have a fair coin and a biased coin that always comes up heads. We run an experiment consisting of three independent events: one of the two coins is chosen at random, the coin is flipped once, and then it is flipped again. Suppose that the chosen coin comes up heads both times. What is the probability that is is biased? We solve this problem using Bayes's theorem. Let A be the event that the biased coin is chosen, and let B be the event that the coin comes up heads both times. We wish to determine Pr[A|B]. We have Pr[A] = 1/2, Pr[B|A] = 1, $Pr[\overline{A}] = 1/2$, and $Pr[B|\overline{A}] = 1/4$; hence,

$$Pr[A|B] = \frac{(1/2) \cdot 1}{(1/2) \cdot 1 + (1/2) \cdot (1/4)}$$

= 4/5.

4 Discrete random variables

A (discrete) random variable X is a function from a finite or countably infinite sample space S to the real numbers. It associates a real number with each possible outcome of an experiment, which allows us to work with the probability distribution induced on the resulting set of numbers.

Random variables can also be defined for uncountably infinite sample spaces. For our purposes, we shall assume that random variables are discrete.

For a random variable X and a real number x, we define the event X = x to be $\{s \in S : X(s) = x\}$; thus,

$$Pr[X = x] = \sum_{\{s \in S: X(s) = x\}} Pr[s].$$

The function

$$f(x) = \Pr\left[X = x\right]$$

is the *probability density function* of the random variable X. From the probability axioms, $Pr[X = x] \ge 0$ and $\sum_{x} Pr[X = x] = 1$. As an example, consider the experiment of rolling a pair of ordinary 6-sided dice. There are 36 possible elementary events in the sample space. We assume that the probability distribution is uniform, so that each elementary event $s \in S$ is equally likely: Pr[s] = 1/36. Define the random variable X to be the *maximum* of the two values showing on the dice. We have Pr[X = 3] = 5/36, since X assigns a value of 3 to 5 of the 36 possible elementary events, namely (1,3), (2,3), (3,3), (3,2), and (3,1). It is common for several random variables to be defined on the same sample space. If X and Y are random variables, the function

$$f(x, y) = Pr[X = x \text{ and } Y = y]$$

is the *joint probability density function* of X and Y. For a fixed value y,

$$Pr[Y = y] = \sum_{x} Pr[X = x \text{ and } Y = y]$$

and similarly, for a fixed value x,

$$Pr[X = x] = \sum_{y} Pr[X = x \text{ and } Y = y].$$

Using the definition (20) of conditional probability, we have

$$Pr[X = x|Y = y] = \frac{Pr[X = x \text{ and } Y = y]}{Pr[Y = y]}$$

We define two random variables X and Y to be *independent* if for all x and y, the events X = x and Y = y are independent or, equivalently, if for all x and y, we have Pr[X = x and Y = y] = Pr[X = x] Pr[Y = y].

Given a set of random variables defined over the same sample space, one can define new random variables as sums, products, or other functions of the original variables. The simplest and most useful summary of the distribution of a random variable is the "average" of the values it takes on. The *expected value* (or, synonymously, *expectation* or *mean*) of a discrete random variable X is

$$\mathbb{E}[x] = \sum_{x} x Pr\left[X = x\right],\tag{24}$$

which is well defined if the sum is finite or converges absolutely. Sometimes the expectation of X is denoted by μ_X or, when the random variable is apparent from context, simply by μ .

Consider a game in which you flip two fair coins. You earn \$3 for each head but lose \$2 for each tail. The expected value of the random variable X representing your earnings is

$$\mathbb{E}[X] = 6 \cdot Pr[2H] + 1 \cdot Pr[1H, 1T] - 4 \cdot Pr[2T]$$

= 6(1/4) + 1(1/2) - 4(1/4)
= 1.

The expectation of the sum of two random variables is the sum of their expectations, that is,

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y], \tag{25}$$

whenever $\mathbb{E}[X]$ and $\mathbb{E}[Y]$ are defined. This property extends to finite and absolutely convergent summations of expectations, and it is called *linearity of expectation*:

$$\mathbb{E}\left[\sum_{i=1}^{n} X_{i}\right] = \mathbb{E}\left[X_{1}\right] + \mathbb{E}\left[X_{2}\right] + \dots + \mathbb{E}\left[X_{n}\right].$$

If X is any random variable, any function g(X) defines a new random variable g(X). If the expectation of g(X) is defined, then

$$\mathbb{E}[g(x)] = \sum_{x} g(x) Pr \left[X = x \right]$$

Letting g(x) = ax, we have for any constant a,

$$\mathbb{E}[aX] = a\mathbb{E}[X]. \tag{26}$$

Consequently, expectations are linear: for any two random variables X and Y and any constant a,

$$\mathbb{E}[aX+Y] = a\mathbb{E}[X] + \mathbb{E}[Y].$$
(27)

When two random variables X and Y are independent and each has a defined expectation,

$$\mathbb{E}[XY] = \sum_{x} \sum_{y} xyPr [X = x \text{ and } Y = y]$$

=
$$\sum_{x} \sum_{y} xyPr [X = x] Pr [Y = y]$$

=
$$\left(\sum_{x} xPr [X = x]\right) \left(\sum_{y} yPr [Y = y]\right)$$

=
$$\mathbb{E}[X]\mathbb{E}[Y].$$

In general, when n random variables X_1, X_2, \ldots, X_n are mutually independent,

$$\mathbb{E}[X_1 X_2 \cdots X_n] = \mathbb{E}[X_1] \mathbb{E}[X_2] \cdots \mathbb{E}[X_n]$$
(28)

When a random variable X takes on values from the natural numbers $\mathbb{N} = \{0, 1, 2, ...\}$, there is a nice formula for its expectation:

$$\mathbb{E}[X] = \sum_{i=0}^{\infty} i \Pr\left[X=i\right]$$
(29)

$$= \sum_{i=0}^{\infty} i \left(\Pr\left[X \ge i \right] - \Pr\left[X \ge i + 1 \right] \right)$$
(30)

$$=\sum_{i=0}^{\infty} \Pr\left[X \ge i\right],\tag{31}$$

since each term $Pr[X \ge i]$ is added in *i* times and subtracted out i - 1 times (except $Pr[X \ge 0]$, which is added in 0 times and not subtracted out at all). The *variance* of a random variable X with mean $\mathbb{E}[X]$ is

$$Var[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2 \right]$$
(32)

$$= \mathbb{E}\left[X^2 - 2X\mathbb{E}[X] + \mathbb{E}^2[X]\right]$$
(33)

$$= \mathbb{E}[X^2] - 2\mathbb{E}[X\mathbb{E}[X]] + \mathbb{E}^2[X]$$
(34)

$$= \mathbb{E}[X^2] - \mathbb{E}^2[X]. \tag{35}$$

The justification for the equalities $\mathbb{E}[\mathbb{E}^2[X]] = \mathbb{E}^2[X]$ and $\mathbb{E}[X\mathbb{E}[X]] = \mathbb{E}^2[X]$ is that $\mathbb{E}[X]$ is not a random variable but simply a real number, which means that Equation (26) applies (with $a = \mathbb{E}[X]$). Equation (32) can be rewritten to obtain an expression for the expectation of the square of a random variable:

$$\mathbb{E}\left[X^2\right] = Var[X] + \mathbb{E}^2[X].$$
(36)

The variance of a random variable X and the variance of aX are related:

$$Var[aX] = a^2 Var[X].$$

When X and Y are independent random variables,

$$Var[X+Y] = Var[X] + Var[Y].$$

In general, if n random variables X_1, X_2, \ldots, X_n are pairwise independent, then

$$Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} Var[X_i].$$
(37)

The *standard deviation* of a random variable X is the positive square root of the variance of X. The standard deviation of a random variable X is sometimes denoted σ_X or simply σ when the random variable X is understood from context. With this notation, the variance of X is denoted σ^2 .

5 The geometric and binomial distributions

A coin flip is an instance of a **Bernoulli trial**, which is defined as an experiment with only two possible outcomes: **success**, which occurs with probability p, and **failure**, which occurs with probability q = 1 - p. When we speak of **Bernoulli trials** collectively, we mean that the trials are mutually independent and, unless we specifically say otherwise, that each has the same probability p for success. Two important distributions arise from Bernoulli trials: the geometric distribution and the binomial distribution.

Suppose we have a sequence of Bernoulli trials, each with a probability p of success and a probability q = 1 - p of failure. How many trials occur before we obtain a success? Let the random variable X be the number of trials needed to obtain a success. Then X has values in the range $\{1, 2, \ldots\}$, and for $k \ge 1$,

$$Pr[X = k] = q^{k-1}p,$$
(38)

since we have k - 1 failures before the one success. A probability distribution satisfying Equation (38) is said to be a *geometric distribution*.

Assuming p < 1, the expectation of a geometric distribution can be calculated using

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} kq^{k-1}p \tag{39}$$

$$=\frac{p}{q}\sum_{k=0}^{\infty}kq^k\tag{40}$$

$$=\frac{p}{q}\cdot\frac{q}{(1-q)^2}\tag{41}$$

$$=\frac{1}{p}.$$
(42)

Thus, on average, it takes 1/p trials before we obtain a success, an intuitive result. The variance, which can be calculated similarly, is

$$Var[X] = q/p^2. \tag{43}$$

As an example, suppose we repeatedly roll two dice until we obtain either a seven or an eleven. Of the 36 possible outcomes, 6 yield a seven and 2 yield an eleven. Thus, the probability of success is p = 8/36 = 2/9, and we must roll 1/p = 9/2 = 4.5 times on average to obtain a seven or eleven. How many successes occur during n Bernoulli trials, where a success occurs with probability p and a failure with probability q = 1 - p? Define the random variable X to be the number of successes in n trials. Then X has values in the range $\{0, 1, \ldots, n\}$, and for $k = 0, \ldots, n$,

$$Pr\left[X=k\right] = \binom{n}{k} p^k q^{n-k},\tag{44}$$

since there are $\binom{n}{k}$ ways to pick which k of the n trials are successes, and the probability that each occurs is $p^k q^{n-k}$. A probability distribution satisfying Equation (44) is said to be a *binomial*

distribution. For convenience, we define the family of binomial distributions using the notation

$$b(k;n,p) = \binom{n}{k} p^k (1-p)^{n-k}.$$
(45)

The name "binomial" comes from the fact that Equation (45) is the *k*th term of the expansion of $(p+q)^n$. Consequently, since p+q=1,

$$\sum_{k=0}^{n} b(k; n, p) = 1,$$
(46)

as is required by axiom 2 of the probability axioms. We can compute the expectation of a random variable having a binomial distribution from Equation (46). Let X be a random variable that follows the binomial distribution b(k; n, p), and let q = 1 - p. By the definition of expectation, we have

$$\mathbb{E}[X] = \sum_{k=0}^{n} kb(k; n, p) \tag{47}$$

$$=\sum_{k=0}^{n} k \binom{n}{k} p^{k} q^{n-k} \tag{48}$$

$$= np \sum_{k=0}^{n} \binom{n-1}{k-1} p^{k-1} q^{n-k}$$
(49)

$$= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k q^{(n-1)-k}$$
(50)

$$= np \sum_{k=0}^{n-1} b(k; n-1, p)$$
(51)

$$= np.$$
(52)

By using the linearity of expectation, we obtain the same result with substantially less algebra. Let X_i be the random variable describing the number of successes in the *i*th trial. Then $\mathbb{E}[X_i] = p \cdot 1 + q \cdot 0 = p$, and by linearity of expectation Equation (27), the expected number of successes for *n* trials is

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_{i}\right]$$
$$= \sum_{i=1}^{n} \mathbb{E}[X_{i}]$$
$$= \sum_{i=1}^{n} p$$
$$= np.$$

The same approach can be used to calculate the variance of the distribution. Using Equation (32), we have $Var[X_i] = \mathbb{E}[X_i^2] - \mathbb{E}^2[X_i]$. Since X_i only takes on the values 0 and 1, we have $\mathbb{E}[X_i^2] = \mathbb{E}[X_i] = p$, and hence

$$Var[X_i] = p - q^2 = pq.$$
⁽⁵³⁾

To compute the variance of X, we take advantage of the independence of the n trials; thus, by Equation (53),

$$Var[X] = Var\left[\sum_{i=1}^{n} X_i\right]$$
(54)

$$=\sum_{i=1}^{n} Var[X_i]$$
(55)

$$=\sum_{i=1}^{n} pq \tag{56}$$

$$= npq. \tag{57}$$

The binomial distribution b(k; n, p) increases as k runs from 0 to n until it reaches the mean np, and then it decreases. We can prove that the distribution always behaves in this manner by looking at the ratio of successive terms:

$$\frac{b(k;n,p)}{b(k-1;n,p)} = \frac{\binom{n}{k}p^kq^{n-k}}{\binom{n}{k-1}p^{k-1}q^{n-k+1}}$$
(58)

$$=\frac{n!(k-1)!(n-k+1)!p}{k!(n-k)!n!q}$$
(59)

$$=\frac{(n-k+1)p}{kq}\tag{60}$$

$$=1 + \frac{(n+1)p - k}{kq}.$$
 (61)

This ratio is greater than 1 precisely when (n+1)p-k is positive. Consequently, b(k; n, p) > b(k-1; n, p) for k < (n+1)p (the distribution increases), and b(k; n, p) < b(k-1; n, p) for k > (n+1)p (the distribution decreases). If k = (n+1)p is an integer, then b(k; n, p) = b(k-1; n, p), so the distribution has two maxima: at k = (n+1)p and at k-1 = (n+1)p-1 = np-q. Otherwise, it attains a maximum at the unique integer k that lies in the range np - q < k < (n+1)p. The following lemma provides an upper bound on the binomial distribution.

Lemma 5.1. Let $n \ge 0$, let 0 , let <math>q = 1 - p, and let $0 \le k \le n$. Then

$$b(k;n,p) \le \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}.$$

Proof. Using Equation (13), we have

$$b(k;n,p) = \binom{n}{k} p^k q^{n-k}$$

$$\leq \left(\frac{n}{k}\right) \left(\frac{n}{n-k}\right)^{n-k} p^k q^{n-k}$$

$$= \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}.$$

6 Introduction to the Probabilistic Method

The goal of the *probabilistic method* is to prove that there exists a code C with property P. Therefore define D over all possible codes Ω and prove that

$$\Pr_{\mathcal{C}\in\mathcal{P}}\left[\mathcal{C}\text{ has property }\mathcal{P}\right]>0.$$

The typical approach is to define P_1, \ldots, P_m such that $\mathcal{P} = P_1 \wedge P_2 \wedge P_3 \ldots \wedge P_m$ and show that:

 $Pr\left[\mathcal{C} \text{ doesn't have property } \mathcal{P}\right] < \frac{1}{m}.$

7 Summary of Important Results

1. Linearity of Expectation:

$$\mathbb{E}\left[\sum_{i=1}^{n} X_{i}\right] = \mathbb{E}\left[X_{1}\right] + \mathbb{E}\left[X_{2}\right] + \dots + \mathbb{E}\left[X_{n}\right].$$
(62)

2. Union Bound:

$$Pr\left[\bigcup_{i=1}^{n} A_{i}\right] \leq \sum_{i=1}^{n} Pr\left[A_{i}\right].$$
(63)

3. Markov's Inequality:

$$Pr\left[X \ge t\right] \le \frac{\mathbb{E}\left[X\right]}{t}.$$
(64)

4. Chernoff Bound:

$$Pr\left[X \ge k\right] \le \min_{t \ge 0} e^{-tk} \theta_X(t).$$
(65)