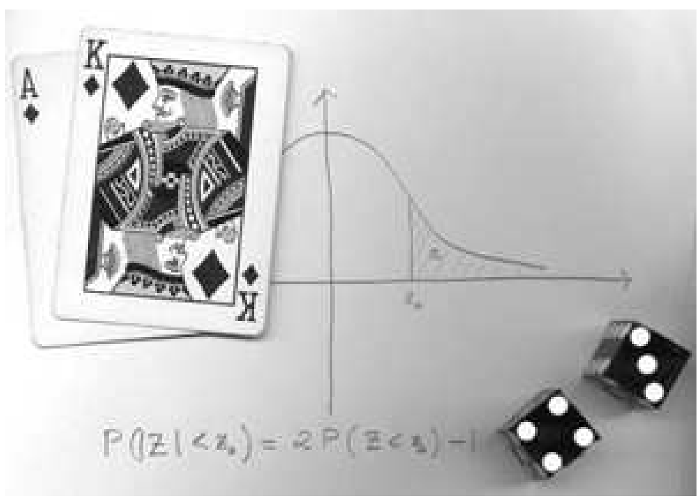
Probability and Statistical Theory

with

Applications to Games and Gambling



C.Friedman
Math.Dept. RLM 8.100
University of Texas
Austin, Texas 78712

References

[1] A. Church, *The Concept of a Random Sequence*,
Bull. Amer. Math. Soc., 46, 130-135, 1940.

[2] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 18th printing, 1991.

[3] J.L. Doob, *Note on Probability*, Annals of Mathematics, Vol. 37, No. 2, 363-367, April, 1936.

[4] Mason Malmuth, Lynn Loomis, *Fundamentals of Poker*, Two Plus Two Publishing, Creel Printing Co, Las Vegas, Nevada, 1992.

[5] Ming Li, Paul Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*, Springer-Verlag, New York, 1993.

[6] P. Martin-Löf, *On the Concept of a Random Sequence*, Theory Probability Appl., 11, 177-179, 1966.

[7] L.J. Savage, *The Foundations of Statistics*, Dover, New York, 1972.

[8] Richard von Mises, *Probability, Statistics, and Truth*, Dover, New York, 1957.

0. Preface

The present work is intended to be a comprehensive introduction to probability and statistics. The mathematical prerequisites are primarily the contents of a standard (2 term) university undergaduate calculus course (including differentiation and integration of functions of one and several variables, infinite series, *etc*) and a bit of linear algebra (matrices), although the parts of the text dealing with finite sample spaces involves little more than high school algebra, and some results involving calculus may make sense to readers lacking the background in calculus if they are willing to accept these without following the derivations. Actually, to really develop the subject properly, much more is necessary - in particular a large portion of the apparatus of *measure theory* is required. Such notions usually appear in graduate mathematics courses, and cannot be presented here. However, we do make comme comments about existence of countably additive measures which the reader can simply accept without any difficulty. These latter notions will not play any crucial role in our presentation.

Section 1 introduces some basic notions in the context of coin tossing. These ideas are quite important even though they occur in a very simple context. The reader should make an effort to understand all the ideas in this section (which might even be considered a "mini" course in probability b itself.) Later sections develop various topics in much more detail.

Although the foundations of probability involve some abstract mathematical notions, the theory originated in an attempt to provide models for real situations and processes; many of the early writers were concerned with calculations related to gambling situations (rolling dice, *etc*) or physics (statistical behavior of gasses and "random" phenomena). It is our intent to preserve this "practical" aspect of the theory. After studying our presentation, the reader should be able to understand the odds involved in gambling situations and make informed decisions concerning these. For example, in poker, (when) should one draw cards to an "inside" straight and in cases that this should be done, exactly what are the quantitative advantages? What is the "best" strategy in various situations? If one really has an advantage in some game, how long must one play to be fairly well assured of being an overall winner, and what are the risks? (These latter questions involve some statistical notions.) When can one draw significant conclusions from certain data? *Etc.*

It is our opinion that many of the elementary notions that occur in probability are actually quite difficult to explain precisely. *E.g.*, what does it mean when we say we toss a "fair" coin? What does it mean when we are told we have a 1 in 4 chance of getting cancer? Usually, these notions are not really precisely defined. It is true that one can consider such notions as axiomatic and then use them to calculate other more complicated probabilities. However, it would be nice to have a more operational meaning for some of these ideas, and we devote some space to this task (mostly in an appendix.)

Our numbering conventions for equations, theorems, definitions, *etc* are fairly simple. All numbering is in the form $< sec\# > . < item\# >$, *e.g.*, Theorem 3.2.1 is the 1st numbered theorem in section 3.2. Theorems, propositions, lemmas, corollaries and definitions use the same numbering counter; this helps in locating a particular item in the text (I find it confusing to look for a proposition 3.2.5 and find theorem 3.2.5, corollary 3.2.10, definition 3.2.4 *etc*, all in the same vicinity as the proposition 3.2.5.) Equations and figures have their own separate numbering counters, but follow the convention cited.

1. Introduction

Probability theory provides a model of certain real situations which often provides very useful insight and understanding. As is the case generally with models, the precise connection with reality is problematical, but there is little doubt that the theory is highly appropriate in many cases. We will have more to say about this later. For now, we will discuss some examples which should serve to introduce the basic framework, terminology, *etc.*

Consider the experiment $\mathcal{E}$ of tossing a "fair" coin once. Instead of using the term "fair", one also says that the "probability of obtaining a head (H) is 1/2." Undoubtedly, it is extremely difficult to give this notion a precise meaning in any operational way. One sometimes hears the explanation that what is meant is that we have no knowledge concerning the outcome. I don't care for this explanation. For example, suppose we had a coin which tended to produce H a substantially larger fraction of the time than T (*e.g.* it might be 2-Headed), but we didn't know that; in such a case, we have no knowledge, but it does not seem correct to say that the probability of H is 1/2 ! So, lack of knowledge seems a poor reason to assume a particular value for the probability of H. We would prefer that the assignment of such a value is closely related to some intrinsic properties of the coin and/or tossing procedure. One such approach is the "frequency interpretation" - we explain the assignment of probability 1/2 to the outcome H, by stating that we believe that in an unending sequence of tosses, the asymptotic relative frequency of H is 1/2, *i.e.*

$$\lim_{n\to\infty} \frac{\#H \text{ in } 1st\ n \text{ tosses}}{n} = 1/2. \tag{1.1}$$

(The term "relative frequency" refers to the ratio of H outcomes to number of tosses; usually, the terminology "frequency of H in $1st$ n tosses" means just the number of H in the $1st$ n tosses. "Asymptotic" refers to the taking of the limit.) Of course, one may question whether this is actually true, but it is possible to just consider it an assumption which is part of our model for coin tossing. There are problems with this idea at least in the overly simple-minded way it has been expressed here. One difficulty is that really much more is probably implicitly assumed concerning the sequence of outcomes of independent tosses of a fair coin. (The notion of "independence" will be discussed later, but it suffices now to understand that the results of some particular toss have no bearing on the results of some other toss.) For example, the sequence

$$H, T, H, T, H, T, \ldots \ (etc)$$

satisfies (1.1), but this sequence is much too regular to serve as an example of what the results of a sequence of coin tosses looks like. We return to this discussion much later when a more satisfactory discussion of the frequency interpretation will be given. Another general problem with the frequency interpretation (even when properly formulated) is that often the notion of repeating some experiment may be unconvincing. Consider, for example, the question "What is the probability that a cure for cancer will be discovered in the next year?" Whatever value might be assigned to the probability of such an event, it seems hard to understand this by means of a frequency interpretation. Another approach and one that is usually taken in modern presentations of the theory is the axiomatic one. We define a *sample space* S whose members are the *elementary outcomes* H and T, *i.e.*

$$S = \{H, T\} \tag{1.2}$$

and a *probability* P for which

$$P(H) = 1/2, \quad P(T) = 1/2. \tag{1.3}$$

This is our explanation of what is meant by the statement that the coin is fair; clearly, there is nothing operational here - we are in essence refusing to assign any "explanation" to the notion of fairness other than the assertion that it corresponds to this mathematical construct. This may be somewhat philosophically unsatisfying, but nicely avoids the difficulty of saying what is "really" meant. (In fairness, it must be admitted that it is likely not possible to say in a completely satisfactory way what is "really" meant!) We might think of the statement $P(H) = 1/2$ as an undefined notion somewhat like undefined notions in geometry (point, line, plane *etc*, in the axiomatic development.) Actually, the probability P is considered as a function assigning real values to *subsets* of S; in the present example, there are not too many of these, namely: the empty subset $\emptyset$, S itself, $\{H\}$, and $\{T\}$. We can define

$$P(\emptyset) = 0, \quad P(S) = 1, \quad P(\{H\}) = 1/2, \quad P(\{T\}) = 1/2. \tag{1.4}$$

We usually identify a singleton set with its single member, in which case $P(\{H\}) = P(\{T\}) = 1/2$ is equivalent to (1.3). The subsets of S are called *events* (for larger sample spaces, the events are generally more interesting than in the present small example; it is also the case that when S contains infinitely many elements, sometimes not every subset of S is considered to be an event.)

A property which holds in the previous situation and which holds generally for sample spaces and the associated probability P is:

$$P(A \cup B) = P(A) + P(B) \text{ if } A, B \text{ are events and } A \cap B = \emptyset. \tag{1.5}$$

A consequence of this is the property:

$$P(A) = \sum_{s \in A} P(s) \text{ if A is finite.} \tag{1.6}$$

These properties have more general forms which will be considered later.

An alternate way to express probabilities is by using the notion of *odds*. As an example, one says that when tossing a fair coin, the odds of obtaining a H are 1 to 1 (either "against" or "for" in this case.) This expresses the idea that H and T are equally likely. The odds terminology is useful when considering what the payoffs on various bets should be. For example, in the case being considered of tossing a fair coin, a fair bet might consist of betting \$1.00 on the result H; If you are wrong, you lose your \$1.00; if you are correct, you win \$1.00 (in addition to getting your \$1.00 bet returned.) Why is this "fair"? The explanation one often finds in elementary discussions of probability is as follows: in 2 such bets, "on average" you will win one time and lose one time for a net profit of \$0. Although this seems reasonable, the precise meaning of the term "on average" is a bit involved; to properly understand what is meant requires familiarity with the notions of *expectation* and the *law of large numbers*. At the present time, we describe very briefly the notion of expectation in order to give a possible meaning to the notion of "fair" in the present example. Consider the function

$$X : S = \{H, T\} \to \mathbb{R}^1 \tag{1.7}$$

defined by

$$X(H) = 1, \quad X(T) = -1. \tag{1.8}$$

(This notation means that X is a function from S to $\mathbb{R}^1$ [the real numbers].)
Generally, a function from a sample space S to $\mathbb{R}^1$ is called a *random variable*. In the case that S is finite, the *expectation* (or *expected value*) of X, denoted $E(X)$ is defined by:

$$E(X) = \sum_{s \in S} X(s)P(s). \tag{1.9}$$

Notice that $E(X)$ is a "weighted average" of the values of X (the "weights" are the values $P(s)$.) We interpret $E(X)$ as an "average" value of X (the precise operational meaning of this will have to wait till more machinery has been developed.) In the case of X given by (1.7), (1.8) associated with the bet on H when a fair coin is tossed, the values $X(H), X(T)$ are your winnings and the expectation of X is

$$E(X) = 1 \cdot (\frac{1}{2}) + (-1) \cdot (\frac{1}{2}) = 0. \tag{1.10}$$

The fact that your expected winnings ($= E(X)$) are 0 means that you have no advantage or disadvantage; this is what is meant by the statement that the bet is "fair".

Now consider the experiment of tossing a coin where it is postulated that $P(H) = 1/3$, $P(T) = 2/3$. Our sample space is still $S = \{H, T\}$, but P is defined differently. We now say that that the odds are 2 to 1 against obtaining a H when we toss the coin (note that T is twice as likely to occur as H.) If you bet \$1.00 on H, the bet is fair if you lose your \$1.00 if T occurs, but win \$2.00 (plus get your \$1.00 back) if H occurs - if X is the random variable whose values are your winnings, then

$$X(H) = 2, \quad X(T) = -1 \tag{1.11}$$

and

$$E(X) = 2 \cdot (\frac{1}{3}) + (-1) \cdot (\frac{2}{3}) = 0. \tag{1.12}$$

(Again you have no advantage or disadvantage.) Note that the winnings of \$2.00 when \$1.00 is bet on H are in the same proportion as the odds against H occurring (2-1.)
Now suppose that in the present situation, you could bet \$1.00 on H and someone would pay you \$3.00 if H occurs (and return your original \$1.00), but you would lose your \$1.00 if T occurs. If this were the case, then you should be happy to make this bet repeatedly. In this case, your winnings are

$$X(H) = 3, \quad X(T) = -1 \tag{1.13}$$

and the expectation is

$$E(X) = 3 \cdot (\frac{1}{3}) + (-1) \cdot (\frac{2}{3}) = 1/3. \tag{1.14}$$

Note that the elementary reasoning mentioned earlier leads to the same result: on average, in 3 such bets, one wins one for a \$3.00 gain and loses two for a \$2.00 loss *i.e.*, a \$1.00 win in 3 bets, hence an (average) gain of \$1/3 *per* bet. You have an advantage (or an "overlay" or "the best of it") in this situation. Of course, you are not likely to find someone to make such a bet with you unless they don't like money. However, in more complicated situations where the odds are not quite as obvious, it is sometimes possible to find betting situations where you have such an advantage. (We will see later that in almost all casino games played against the house, the house has the best of it; nevertheless, people continue to play these games. One reason for this is that in most of these games there is a significant short term luck factor; *i.e.*, even though the player can't win in the long run, it is possible to win significantly for a while with good luck. We discuss this more later. But occasionally, casinos either make mistakes or to lure customers actually allow the player to have an advantage in some betting situation.)

Remark: We have made use of the notions of *probability* and *odds*. One should practice converting between these two notions. For example, if an event A has probability $2/7$ of occurring, then the odds against its occurrence are 5 to 2. (If an event has probability less than $\frac{1}{2}$ of occurring, then one usually states the odds *against* its occurrence.)

We present another application of the notion of expectation. Consider the experiment of tossing a fair coin again ($P(H) = P(T) = 1/2$.) Let X be the random variable equal to the "number of H", i.e.,

$$X(H) = 1; \quad X(T) = 0. \tag{1.15}$$

The expectation of X is

$$E(X) = 1 \cdot (\frac{1}{2}) + 0 \cdot (\frac{1}{2}) = \frac{1}{2}. \tag{1.16}$$

This makes good heuristic sense - the "average" number of H in one toss is $\frac{1}{2}$ since H and T are equally likely. Notice that the expectation of X is not a value of X which we "expect" to actually occur (even though the expectation is also called the "expected value"); obviously, the outcome can't be $\frac{1}{2}H$! How should one interpret the meaning of the expectation of X in this case? Of course, it has a well-defined mathematical meaning, but if one wishes to have a more operational meaning, then $E(X)$ can be thought of as an average over many independent repetitions of the experiment of tossing the coin, because we will see that (1.1) holds with probability 1 (this is one form of the *law of large numbers*.)

We now return to the experiment of tossing a fair coin, but consider the case of 2 tosses. A natural sample space associated with this experiment consists of the 4 possible outcomes:

$$S = \{HH, TT, HT, TH\} \tag{1.17}$$

and it seems reasonable in this situation to set

$$P(HH) = P(TT) = P(HT) = P(TH) = 1/4. \tag{1.18}$$

There are now some non-trivial events, *e.g.*, the event $A = \{HH, HT, TH\}$. When we say an "event occurs" we mean that <u>the outcome of the experiment is a member of the event</u>. In this case, A is clearly the "event that at least one H occurs", and (using (1.6)) we see that $P(A) = 3/4$. We can define a random variable X equal to the "number of H" as before:

$$X(TT) = 0; \quad X(HT) = X(TH) = 1; \quad X(HH) = 2. \tag{1.19}$$

The event A just considered can then be expressed

$$A = \{s \in S \mid X(s) \geq 1\}. \tag{1.20}$$

It is standard probabilistic notation to express A by

$$A = \{X \geq 1\} \tag{1.21}$$

and to write

$$P(X \geq 1) = 3/4. \tag{1.22}$$

In a similar way

$$\{X = 0\} = \{TT\}; \quad \{X = 1\} = \{HT, TH\}; \quad \{X = 2\} = \{HH\} \tag{1.23}$$

and

$$P(X = 0) = 1/4; \quad P(X = 1) = 1/2; \quad P(X = 2) = 1/4. \tag{1.24}$$

We can calculate the expectation of X by using (1.9):

$$E(X) = 0 \cdot P(TT) + 1 \cdot P(HT) + 1 \cdot P(TH) + 2 \cdot P(TT) = 1. \tag{1.25}$$

Notice that if we group the 2 terms with coefficient 1, we have:

$$E(X) = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) + 2 \cdot P(X = 2) = 1. \tag{1.26}$$

This generalizes; if X is a random variable on a finite sample space S with probability P, then

$$E(X) = \sum_{s \in S} X(s)P(s) = \sum_{x} x \cdot P(X = x) \tag{1.27}$$

where the sum is over all the values x of the random variable X. The second form of the sum in (1.27) is just obtained from the first by grouping terms as in the particular example above; it is often simpler to compute since it may involve a much smaller number of terms. In passing, we mention that when we know the values $P(X = x)$, then we say that we know the *distribution* of X; a more complete discussion of this notion will be presented later.

There is another way to compute the expectation of X which was done in (1.25), (1.26) which will be quite important and useful in other examples. We first note that it follows from (1.9) that for two random variables X and Y on a finite sample space

$$E(X + Y) = E(X) + E(Y). \tag{1.28}$$

This is elementary:

$$E(X + Y) = \sum_{s \in S} (X(s) + Y(s)) = \sum_{s \in S} X(s) + \sum_{s \in S} Y(s) = E(X) + E(Y). \tag{1.29}$$

(Note that this additivity property is less obvious from the second form of the expectation in (1.27) because generally the expression of $P(X + Y = \xi)$ in terms of $P(X = x)$ and $P(Y = y)$ is a little complicated.)

Now we use this property of the expectation to compute $E(X)$ for the X of (1.19) which was done in (1.25). We note that

$$X = X_1 + X_2 \tag{1.30}$$

where X_i (for $i = 1, 2$) is the random variable whose value is 0 or 1 depending on whether the *ith* toss yields T or H respectively. (*e.g.*, $X_1(HH) = X_1(HT) = 1$, $X_1(TH) = X_1(TT) = 0$.) It should be clear that X_1 and X_2 both have the same expectation as the random variable equal to the number of H when we toss a fair coin once (since X_i is the number of H on the *ith* toss), and this expectation is 1/2. Hence from (1.29) and (1.30) we get

$$E(X) = E(X_1) + E(X_2) = 1/2 + 1/2 = 1. \tag{1.31}$$

Although this may not seem especially useful in the simple case being considered here, a little reflection will indicate that this same line of reasoning is applicable to the case of tossing a fair coin n times, leading to the result that the expected number of H will be $n/2$. This result is much more complicated to compute directly from the definition of expectation as we will see below.

We emphasize once again that the expectation is a kind of "average" of X and should **not** be thought of as the value of X that we expect to occur; in the case of n tosses, $n/2$ is not a possible number of H if n is odd, (and even if n is even, the event that $n/2$ H occur is not very likely when n is large as we will also see in what follows..)

Before leaving our discussion of tossing a fair coin twice, we make a final remark concerning the sample space associated with this experiment. If we were only interested in the number of H obtained, it might be appropriate to consider as our sample space

$$S = \{0H, 1H, 2H\}. \tag{1.32}$$

There is nothing wrong with this choice, but in order to be consistent with the way we assigned probabilities previously, we see that that we must have

$$P(0H) = 1/4;\ P(1H) = 1/2;\ P(2H) = 1/4. \tag{1.33}$$

That is, the members of this S are not equaly likely outcomes. Generally, if it is possible to choose a sample space for which the members all have equal probabilities, then the computation of the probabilities of events amounts to just counting how many members these events contain (*i.e.* combinatorial arguments.) However, if the members of S have different probabilities, then the calculation of the probabilities of events is more complicated. For this reason, it is usually preferable to choose a sample space consisting of equally likely outcomes *if this is possible.*

In order to have some less trivial example and to develop some further useful results, we now consider in some detail the experiment of tossing a coin n times. We take the probability of H to be a number $0 \le p \le 1$ on each toss, and assume that the result of any toss has no effect on the results of other tosses. (This means that the tosses are "independent"; we discuss this notion in more detail later.) As sample space for this experiment, we take the set of all sequences of length n consisting of H and T:

$$S = \{HH \ldots H,\ HH \ldots HT,\ \ldots\}. \tag{1.34}$$

(In (1.34), the sequence $HH \ldots H$ is a single member of S.)
There are 2^n sequences in S, because to form a sequence n-long of H and T, we have to decide whether the first entry in the sequence is H or T, *i.e.*, we have 2 choices; then for each of these choices, we have 2 choices for the next entry, *etc.* So altogether we can make the choices for the n entries in $2 \cdot 2 \cdot \ldots \cdot 2$ ways, which means there are 2^n possibilities for the result of all these choices. We indicate this by writing

$$\text{card}(S) = 2^n \qquad \text{or} \qquad |S| = 2^n. \tag{1.35}$$

(card(S) or $|S|$ denote the *cardinality* of S which for finite sets means the number of members of the set.)

The reasoning just used can be stated as a useful combinatorial principle in the following result (we also include a related principle):

Prop. 1.1 Suppose we have to select 1 from among m distinct possible choices, followed by 1 from among n distinct possible choices. The number of ways of doing this is mn. If, instead, we have to choose *either* 1 from among the m choices *or* 1 from the n choices, and all the choices are distinct, then the number of ways of doing this is $m + n$.

Pf. In the first situation, we can make the choice from among the m possibilities in m ways, and then for each of these ways, we can make the choice from among the n in n ways. If we list all pairs of choices where the 1st choice can be any of m possibilities and the 2nd choice can be any of n possibilities, then we will list mn pairs. In the second situation, we essentially have to select 1 from a collection consisting of the m and the n possibiities, which amounts to selecting 1 from $m + n$ total possibilities; this can be done in $m + n$ ways. ∎
(*Note*: This extends in an obvious way to making a sequence of more than 2 choices or choosing 1 from among more than 2 sets of choices.)

We have to decide how to assign probabilities to the members of S in (1.34). If $p = 1/2$, then we would consider all the members of S as equally likely, so we could assign probability 2^{-n} to each. In the general situation where p is not necessarily $1/2$, we assign the probabilities as follows:

$$P(s) = p^k q^{n-k}, \text{ if } s \text{ is any sequence with } k\ H \text{ and } n-k\ T \quad (q = 1-p). \tag{1.36}$$

(Note that there may be many such sequences with k H and $n-k$ T unless $k=0$ or n.) The justification of (1.36) involves the notion of independence which hasn't been covered yet, so we present an *ad hoc* argument for its validity. Think about the frequency interpretation of these probabilities. We envision an unending (or very large number N) of repetitions of the experiment of tossing the coin n times.
(That is, we toss the coin n times, note the result, toss another n times, note the result, and keep repeating this.)
Consider a member of S of form $HT\ldots H$ (a sequence of length n.) We are assigning this sequence probability $p \cdot q \cdot \ldots \cdot p$. The idea is that in our repeated trials of the n tosses, we expect the proportion of outcomes where the first toss yields H to be p; **among these**, the proportion with the second toss yielding T should be q, *etc*; hence, the proportion of outcomes which are of form $HT\ldots H$ should be $p \cdot q \cdot \ldots \cdot p$.

Once we know the probabilities of the members of S, we can calculate probabilities of *events* (subsets of S), by addition. For example, let A_k be the event that exactly k H occur.

$$A_k = \{s \in S \mid s \text{ contains exactly } k\ H\} \tag{1.37}$$

If we define a random variable $X : S \to \mathbb{R}$ by

$$X(s) = k \text{ if } s \text{ contains exactly } k\ H \tag{1.38}$$

then

$$A_k = \{X = k\}. \tag{1.39}$$

We want to compute $P(A_k)$ (or equivalently $P(X = k)$.) To do this, we only need to compute the number of members of A_k, since by (1.36) and (1.37), we have

$$P(A_k) = P(X = k) = |A_k| \cdot p^k q^{n-k}. \tag{1.40}$$

(Recall that $|A_k| = \text{card}(A_k) =$ the number of members of A_k.)
Since A_k consists of sequences n-long of H, T with exactly k H appearing, it follows that $|A_k|$ is the same as the number of subsets of of size k of a set with n members (to form

a sequence of A_k, we have to pick a subset of size k of the n positions in a sequence of length n and set these k positions equal to H.) Now sets are *unordered* objects (a set is determined by knowing its members; the notion of order has no meaning for sets, even though the elements of a set are often listed in some particular order. Similarly, the notion of *repeated elements* makes no sense for sets.) However, in counting members of a set, it is often useful to consider order initially and then discount this later. In addition, ordered collections of objects are important in their own right. We digress to present some combinatorial notions. These will be useful in computing $|A_k|$.

Suppose we have a set W with n (distinct) members. Suppose we want to form an *ordered* collection of k members of W. In how many ways can this be done? Well, to do this we have to choose a 1st member which can be done in n ways, then a 2nd member which can be done in $n-1$ ways, ..., and finally a kth member which can be done in $n-k+1$ ways. By Prop. 1.1, this can be done in $n \cdot (n-1) \cdot \ldots \cdot (n-k+1)$ ways. This last number is equal to $n!/(n-k)!$ (where $r! = r \cdot (r-1) \cdot \ldots \cdot 1$.) An ordered collection of objects is called a *permutation*, and an ordered collection of k members of W is thus a permutation of k members of W or a permutation of size k of the members of W. We have shown:

Prop. 1.2 The number of permutations of size k of the members of a set of cardinality (size) n is

$$P(n,k) = n \cdot (n-1) \cdot \ldots \cdot (n-k+1) = n!/(n-k)!. \tag{1.41}$$

$P(n,k)$ is also called "the number of permutations of n distinct objects taken k at a time." Another notation for $P(n,k)$ is ${}_nP_k$.

An ordering of a set with n members is called a permutation of the n members, so we have

Cor. 1.3 The number of permutations of n distinct objects is $P(n,n) = n!$.

As an example, consider the set $W = \{1,2,3,4\}$. We know that $P(4,2) = 4!/(4-2)! = 12$. Notice that there are 6 *subsets* of W, namely:

$$\{1,2\},\ \{1,3\},\ \{1,4\},\ \{2,3\},\ \{2,4\},\ \{3,4\}. \tag{1.42}$$

Each of these subsets can be ordered in 2 ways, which produces exactly 12 permutations of size 2.

The 6 subsets in (1.42) are called *combinations* of size 2 of the members of W.

Combinations are *unordered* collections of objects. If we have n distinct objects, the notation $C(n,k)$ (or sometimes ${}_nC_k$) is used for the number of combinations (unordered collections) of size k from the n. $C(n,k)$ is also called "the number of combinations of n distinct objects taken k at a time." The notation $\binom{n}{k}$ is also used for the quantity $C(n,k)$.

The above example indicates the relationship between $P(n,k)$ and $C(n,k)$; we have:

Prop. 1.4 The number of combinations of n distinct objects taken k at a time is

$$\binom{n}{k} = C(n,k) = P(n,k)/k! = \frac{n!}{(n-k)!k!}. \tag{1.43}$$

This is also the number of subsets of size k formed from the members of a set of cardinality (size) n.

Pf. Since each combination can be ordered in $k!$ ways, $P(n,k) = k!C(n,k)$. ∎

Because $C(n,k)$ is the number of ways of choosing k objects from among n distinct objects (with no ordering), $C(n,k)$ is sometimes referred to by the locution "n choose k".
Remark: An elementary property of the quantity $\binom{n}{k}$ is the following

$$\binom{n}{k} = \binom{n}{n-k}. \tag{1.44}$$

Equation (1.44) follows immediately from the observation that every choice of k from n distinct objects is paired with a unique subset of $n-k$ of the objects (the ones not chosen.)

Finally, consider a set W with n members. How many subsets (of any size) does W have? (This includes the empty set $\emptyset$, and W itself.) The easiest way to calculate this is to reason that to form a subset we have to decide, for each member of W, whether it is in the subset. So we have to decide "yes" or "no" n times to form the subset, and it follows from Prop 1.1 that there are 2^n subsets of W.

As an example, the sample space S associated with n tosses of a coin (see (1.34)) has 2^n members. Hence the number of events (subsets of S) associated with this experiment is $2^{(2^n)}$. (Note that this is **not** the same as $(2^2)^n$ which is just 2^{2n}.) When $n = 8$, $2^n = 512$, and the number of events is thus 2^{512}. By comparison, the number of molecules in the universe is estimated at about 10^{75} which is much smaller than 2^{512}.
This indicates why some general combinatorial principles are necessary in dealing with such situations.

There is another way to compute the number of subsets of a set W (with card(W)=n.) We could add the numbers of subsets of size $k = 0, 1, \ldots, n$, *i.e.*,

$$C(n,0) + C(n,1) + \ldots + C(n,n) = \binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{n}. \tag{1.45}$$

To compute the sum in (1.45), we need the following result:

Thm. 1.5 (Binomial Theorem)

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k}\, b^k$$

(Note that $0! = 1$)

Pf. This may be proved by induction, but there is an elementary combinatorial proof which uses the notions we have developed. When we expand

$$(a+b)^n = (a+b) \cdot \ldots \cdot (a+b) \qquad [n \text{ factors}] \tag{1.46}$$

we get a sum of terms each formed by choosing either a or b from each factor $(a+b)$ and multiplying these choices. The total number of such terms which contain b chosen k times (and hence a chosen $n-k$ times) is the number of ways of choosing b from exactly k of the factors $(a+b)$. This is just $\binom{n}{k}$. ∎

By Thm. 1.5 the sum in (1.45) is $(1+1)^n = 2^n$.

Now we can return to finish the calculation of the probability $P(X=k)$ of obtaining exactly k H when a coin with $P(H) = p$ is tossed n times. Combining the combinatorial results we have discussed and equation (1.40), we have

$$P(X=k) = \binom{n}{k} \cdot p^k q^{n-k} = \frac{n!}{(n-k)!k!} \cdot p^k q^{n-k}. \tag{1.47}$$

Definition 1.6 A random variable X with values $0, 1, \ldots, n$ satisfying (1.47) is said to have a *binomial distribution* based on n and p.

We now calculate the expectation, $E(X)$, of X. It is possible to obtain this by using (1.47) directly:

$$E(X) = \sum_{k=0}^{n} k \cdot \frac{n!}{(n-k)!k!} \cdot p^k q^{n-k}. \tag{1.48}$$

In fact, we have

$$\sum_{k=0}^{n} k \cdot \frac{n!}{(n-k)!k!} \cdot p^k q^{n-k} = \sum_{k=1}^{n} \frac{n!}{(n-k)!(k-1)!} \cdot p^k q^{n-k} \tag{1.49}$$

and setting $\kappa = k - 1$, the last sum becomes

$$np \sum_{\kappa=0}^{n-1} \frac{(n-1)!}{(n-1-\kappa)!(\kappa)!} \cdot p^\kappa q^{n-1-\kappa} = np \tag{1.50}$$

because the sum $\sum_{\kappa=0}^{n-1} \frac{(n-1)!}{(n-1-\kappa)!(\kappa)!} \cdot p^\kappa q^{n-1-\kappa}$ is the sum of the binomial probabilities for the distribution based on $n-1$ and p, and the sum of these is 1 (this is just the probability that the associated random variable has *any* value.)
A much simpler calculation of $E(X)$ is performed by mimicking the method used in (1.31). We define random variables X_i for $i = 1, \ldots n$, where

$$X_i = \begin{Bmatrix} 1 & \text{if the } ith \text{ toss is } H \\ 0 & \text{otherwise} \end{Bmatrix} \tag{1.51}$$

From the way in which the probabilities were defined in (1.36), it should be fairly clear that $P(X_i = 1) = p, P(X_i = 0) = q$ (this is easily checked), so that

$$E(X_i) = 1 \cdot p + 0 \cdot q = p. \tag{1.52}$$

Since, $X = X_1 + \ldots + X_n$, we have

$$E(X) = E(X_1) + \ldots + E(X_n) = np. \tag{1.53}$$

What is the operational meaning of (1.53)? It is, of course, not that we "expect" the number of heads obtained in n tosses to be np or even close to np. Suppose, for example, that $p = 1/2$. In this case, $n/2$ is not even a possible number of heads if n is odd. Suppose we consider the case of an even number of tosses and the probability of obtaining n heads in $2n$ tosses (with $P(H) = 1/2$). Using (1.47), we have

$$P(n\ H \text{ in } 2n \text{ tosses}) = \frac{(2n)!}{n!n!} 2^{-2n}. \tag{1.54}$$

To estimate this quantity, it is useful to have an estimate of $n!$ for large n, and this is provided by *Stirling's formula*:

$$n! \sim n^n e^{-n} \sqrt{2\pi n}. \tag{1.55}$$

The meaning of (1.55) is that the *ratio* of the terms on the right and left of the symbol $\sim$ tends to 1 as $n \to \infty$ (however it is not true that these 2 terms are close in value for large n). Using (1.55) for the factorials in (1.54), produces (after some simplification):

$$P(n\ H \text{ in } 2n \text{ tosses}) \sim \frac{1}{\sqrt{\pi n}} \tag{1.56}$$

and so

$$P(n\ H \text{ in } 2n \text{ tosses}) \to 0 \text{ as } n \to \infty. \tag{1.57}$$

(*Remark*: It is also true that the probability - that in $2n$ tosses of a fair coin the number of H differs from n by not more than some fixed bound - still tends to 0 as $n \to \infty$.) This probably leads one to wonder what the "law of averages" actually says concerning fair coin tossing! Roughly, the idea is that after a large number n of independent tosses of a fair coin (with sample space as in (1.34) and $P(H) = p$), one expects the *relative* frequency of H to be close to $1/2$, even though the frequency of H generally won't be too close to $n/2$. More generally, suppose we are considering independent tosses of a coin for which $P(H) = p$. Then the *Weak Law of Large Numbers* holds:

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \ldots + X_n}{n} - p\right| \geq \epsilon\right) = 0, \text{ for } \epsilon > 0. \tag{1.58}$$

(The X_i are as in (1.51).) This will be proved later (in a more general form.)

Equation (1.58)can be written in the form

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - p)\right| \geq \epsilon\right) = 0, \text{ for } \epsilon > 0. \tag{1.59}$$

This indicates that perhaps the factor $\frac{1}{n}$ multiplying the sum in (1.59)is "too large" in a certain sense, and that a more interesting result might be obtained by dividing by a factor which decreases more slowly. This is in fact the case. The *Central Limit Theorem* (for the special case of the X_i as in (1.51)and with $P(H) = p$) implies:

$$\lim_{n\to\infty} P\left(a < \frac{\sum_{i=1}^{n}(X_i - p)}{\sqrt{npq}} < b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}\,dx. \tag{1.60}$$

This result is extremely useful in understanding the behavior of the binomial distribution (equation (1.47)) for large n. If we want to know the probability that X of (1.47) is in a certain range, it is cumbersome and difficult to have to deal with a sum of terms of the type appearing in (1.47); (1.60) is much easier to analyze, because tables of values of the integral appearing in (1.60) are readily available. For example, it is known that about 95%

of the area under the graph of $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ is contained in the interval $-2 \leq x \leq 2$. We can make use of this fact by setting $a = -2, b = 2$ in (1.60) and rearranging a bit to get:

$$P(np - 2\sqrt{npq} < X < np + 2\sqrt{npq}) \approx .95, \; n \text{ large.} \tag{1.61}$$

(In (1.61), $X = \sum_{i=1}^{n} X_i$. How large n has to be depends on the value of p; for $p \approx .5$, $n \geq 35$ is sufficient; more generally, the result in (1.61) is accurate if $0 < p \pm 2\sqrt{pq/n} < 1$.) What (1.61) indicates is that in n tosses of a coin (with $P(H) = p$), fluctuations in the random variable X (equal to the number of H) of the order of $2\sqrt{npq}$ from the expectation np are not all that unlikely. For the case $p = .5$ (fair coin), the expectation of the number of heads in n tosses is $n/2$, and fluctuations of size $\sqrt{n}$ are not surprising. Note that $\sqrt{n}$ tends to ∞ with n. What this means, for example, is that if 2 players continually bet \$1.00 on the result of tossing a fair coin (with \$1.00 won or lost each time), then after many bets, one of the players may be quite a bit ahead in \$ won, even though the *fraction* of the \$ bet held by each player will likely be about 1/2 of the total wagered so far! (In a similar way, if a group of poker players of essentially equal skill play for a long period of time [*e.g.*, months or years], then it is not unlikely that a few of the players will be significantly ahead in winnings.) This is a result that is often not well understood even by professsional gambler/writers. The case in which $p \neq .5$ is somewhat different; (1.61) provides useful information concerning this case also. Suppose, for example, that $p = .55$, so the coin is biased (H is more likely than T), and suppose you know this, but your opponent doesn't. Then if you always bet on H, you should eventually break your opponent (assuming he keeps playing the game - probably an unlikely supposition in practice!) In fact, if W_i is your winnings on the *ith* toss,

$$E(W_i) = 1 \cdot (.55) - 1 \cdot (.45) = .1 \tag{1.62}$$

so you have expected winnings of \$.10 *per* toss. Note that your winnings after n tosses, W, and X of (1.61)are related by the relation

$$W = X - (n - X) = 2X - n. \tag{1.63}$$

Clearly, the *expectation* of W is $.1n$ since $W = \sum_{i=1}^{n} W_i$. However, in order to sure of winning *anything*, we need $W = 2X - n > 0$ or $X > n/2$. In order to be 95% sure that $W > 0$, we therefore need, according to (1.61) that

$$np - 2\sqrt{npq} > n/2. \tag{1.64}$$

(According to (1.61) , X could be as small as $np - 2\sqrt{npq}$ even when we restrict to an interval with probability .95 for X.)

If we solve for n in (1.64), we find

$$n > 4pq/(p - .5)^2 \tag{1.65}$$

which, when $p = .55$ gives

$$n > 396. \tag{1.66}$$

If $p = .51$, the analogous calculation gives

$$n > 9996. \tag{1.67}$$

What this means is that even when you have an advantage, you may have to play for a long time in order to be reasonably sure of being ahead. (A similar calculation shows that even if you have a disadvantage, you can sometimes be ahead for a long time with reasonable luck.) These kinds of calculations show that it can be quite difficult to know whether you actually have an advantage by looking at your results even for a fairly long time. In casino poker, for example, the advantage an expert player has is small compared to the possible fluctuations that can occur, and even though he figures to be a winner in the long run, the "long run" can sometimes be a matter of many months (or longer) of daily play. This explains why experts can run bad and non-experts can do well for long periods.

1.1 Some Examples - Poker Probabilities

In the present section we calculate various probabilities related to the game of poker. There are many variants of the game of poker, but in all of them, the object is to make the best 5-card poker hand. The ranking of hands is (from lowest to highest):
(i) High card hand (no pair, straight or flush).
(ii) One pair
(iii) Two pair
(iv) Three of a kind ("trips")
(v) Straight (5 cards in numerical sequence. Ace counts as "1" or "14")
(vi) Flush (5 cards of the same suit)
(vii) Full house (3 of one kind, 2 of another)
(viii) Four of a kind
(ix) Straight flush (5 cards in numerical sequence, all of same suit)
Within each group, ranking is determined by comparing highest cards, or 2nd highest cards if highest cards are equal, *etc.*

In *5 card draw*, usually each player antes some fixed amount (so there is something to play for) and then is dealt 5 cards. The players then bet. (Sometimes there is a minimum requirement to open the betting, *e.g.*, a pair of Jacks.) Then each player may discard up to 5 cards if desired and draw (receive) replacements for those discarded (in order to try to make a better hand usually.) Then there is another round of betting after which the hands are shown down. Highest hand takes the pot, unless there is a tie (in which case the pot is split.) (This is just a bare outline of the game, of course.)

We calculate the probability of being dealt various hands in 5 card draw. There are $\binom{52}{5} = 2,598,960$ possible 5 card hands (more if a Joker is being used.)

One pair: There are $\binom{4}{2}$ ways of picking a pair from some particular rank, and 13 ranks. To get a pair you must be dealt 2 of one rank and then 3 other cards of different ranks (with no further pair.) The number of ways of getting the latter is $\binom{12}{3} \cdot 4^3$ (the factor 4^3 occurs because there are 4 suits each of the 3 can be.) Thus the probability of being dealt a pair is

$$\frac{\binom{4}{2} \cdot 13 \cdot \binom{12}{3} \cdot 4^3}{\binom{52}{5}} \approx .422569. \tag{1.1.1}$$

Two pair: To get two pair, you must be dealt pairs from 2 different ranks and then 1 other card of a further rank. The probability of this is

$$\frac{\frac{1}{2} \cdot \binom{4}{2} \cdot 13 \cdot \binom{4}{2} \cdot 12 \cdot 11 \cdot 4}{\binom{52}{5}} \approx .047539. \tag{1.1.2}$$

The initial factor $\frac{1}{2}$ is present so that the ways of picking one pair and then another are not counted twice in different orders.

Trips: Probability of being dealt trips is

$$\frac{\binom{4}{3} \cdot 13 \cdot \binom{12}{2} \cdot 4^2}{\binom{52}{5}} \approx .021128. \tag{1.1.3}$$

Straight: Probability of being dealt a straight is

$$\frac{10 \cdot 4^5}{\binom{52}{5}} \approx .003940 \tag{1.1.4}$$

Flush: Probability of being dealt a flush is

$$\frac{\binom{13}{5}\cdot 4}{\binom{52}{5}} \approx .001981. \tag{1.1.5}$$

Full house: Probability of being dealt a full house is

$$\frac{\binom{4}{3}\cdot 13\cdot\binom{4}{2}\cdot 12}{\binom{52}{5}} \approx .001441. \tag{1.1.6}$$

4 of a kind: Probability of being dealt 4 of a kind is

$$\frac{\binom{4}{4}\cdot 13\cdot 48}{\binom{52}{5}} \approx .000240. \tag{1.1.7}$$

Straight flush: Probability of being dealt a straight flush is

$$\frac{10\cdot 4}{\binom{52}{5}} \approx .000015. \tag{1.1.8}$$

(*Note*: The probabi;lities computed for straights and flushes include the probability of being dealt a straight flush. To get the probabilities for exactly straights or flushes, one should subtract the probability of a straight flush.)

One can also compute the probabilities of improving various hands by discarding and drawing cards. Suppose you have the hand

$$A\heartsuit,\ A\diamondsuit,\ K\spadesuit,\ 7\clubsuit,\ 8\diamondsuit\ . \tag{1.1.9}$$

You could keep the pair of Aces and discard 3, or keep the Aces and King kicker and discard 2.

<u>**First we consider what happens if you discard 3.**</u>

Probability of improving to exactly two pair: This can occur if you draw a pair of one of the 3 ranks you discarded (K,7,8) and then another card from 42 others (52 − original 5 − 3 of rank that produced the new pair − 2 remaining Aces), or a pair of one of the 9 ranks different from the ranks originally held and then another card from 41 others

(52 − original 5 − 4 of rank that produced the new pair − 2 Aces.) The probability is thus

$$\frac{\binom{3}{2}\cdot 3\cdot 42+\binom{4}{2}\cdot 9\cdot 41}{\binom{47}{3}}\approx .159852. \tag{1.1.10}$$

Probability of improving to exactly trips: This occurs if you draw one of the Aces and one each from two of the 3 ranks you discarded, or one Ace and one of the 3 ranks you discarded and one from one of the remaining 9 ranks, or 1 Ace and 2 others from different ranks from the 9 ranks remaining. The probability of this is

$$\frac{2\cdot\binom{3}{1}^2\cdot\binom{3}{2}+2\cdot\binom{3}{1}\cdot 3\cdot\binom{4}{1}\cdot 9+2\cdot\binom{4}{1}^2\cdot\binom{9}{2}}{\binom{47}{3}}\approx .114339. \tag{1.1.11}$$

Probability of improving to a full house: This occurs if you draw another Ace and either a pair from one of the 3 discarded ranks or one of the 9 remaining ranks, or 3 from one of the 3 discarded ranks or 9 remaining ranks. The probability is thus

$$\frac{2\cdot\left[\binom{3}{2}\cdot 3+\binom{4}{2}\cdot 9\right]+\binom{3}{3}\cdot 3+\binom{4}{3}\cdot 9}{\binom{47}{3}}\approx .010176. \tag{1.1.12}$$

Probability of improving to 4 of a kind: There are 45 combinations of 3 cards that include the remaining 2 Aces from the 47 you draw from, so the probability is

$$\frac{45}{\binom{47}{3}}\approx .002775. \tag{1.1.13}$$

Now suppose you discard 2, keeping A, A, K

Probability of improving to exactly two pair: This occurs if you draw one of the remaining Kings and 1 from among 42 cards (52 − the original 5 − 2 Aces − 3 Kings), or 2 7s or 2 8s, or 2 of one of 9 remaining ranks, so the probability is

$$\frac{3\cdot 42+\binom{3}{2}\cdot 2+\binom{4}{2}\cdot 9}{\binom{47}{2}}\approx .172063. \tag{1.1.14}$$

Probability of improving to exactly trips: For this to occur, you must draw 1 Ace and 1 of 42 (52 − 5 original − 3 Kings − 2 Aces) cards, so the probability is

$$\frac{2 \cdot 42}{\binom{47}{2}} \approx .077706. \tag{1.1.15}$$

Probability of improving to a full house: You must draw 1 Ace and 1 King, or 2 Kings, so the probability is

$$\frac{\binom{2}{1} \cdot \binom{3}{1} + \binom{3}{2}}{\binom{47}{2}} \approx .008326. \tag{1.1.16}$$

Probability of improving to 4 of a kind: You must draw the 2 remaining Aces. The probability is

$$\frac{1}{\binom{47}{2}} \approx .000925. \tag{1.1.17}$$

Notice that if you keep the King kicker and draw 3, you have a better chance of improving to 2 pair, but the chances of making the other (stronger) hands are decreased. So what should you do? As usual, "it depends". If you are trying to beat 2 pair smaller than Aces up, you should keep the kicker, but if you think an opponent has trips, you should draw 3, *etc*.

How do you decide what actions to take in playing a hand (*i.e.*, bet, raise, check, call, *etc*)? Generally, it is important to determine whether a given action has positive expectation with regard to the amount that can be won. In 5 card draw, when deciding whether to bet or call after receiving the initial cards, this is often fairly straightforward and involves calculating whether the odds you will be getting from the pot are greater than the odds against winning. For example, if you start with 5 cards that include 4 of one suit, then the probability of completing the flush by drawing 1 card is 9/47 (9 cards out of 47 unseen help you), so the odds against making your flush are 38 to 9, or approximately 4 to 1. Suppose you have to call a $20 bet to stay in the hand and draw, but you estimate that if you make the flush you will be ahead $100. Since you are getting 5 to 1 odds, you may decide to play. Generally, you want the pot odds to be somewhat greater than the odds against making your hand, because you are playing to make money, not break even; and in addition, you must take into account that there may be raises which decrease your pot odds, and you might make your hand and lose to a better hand if someone makes a miracle draw (or you understimated the strengths of your opponents hands.) Your calculations will always be based on what you *think* your opponents hold (you can't see any of their cards), so the probabilities or odds you calculate are at best approximate. We give some further examples of these calculations later.

Remark: When we calculate the probability of completing a 4-flush by drawing one as 9/47, we are ignoring the possibility that some of our opponents may hold the cards we need, and in fact, there aren't 47 cards left that we are drawing from. In actuality, sometimes the probability of making the flush is greater than 9/47 and sometimes less; but it is not difficult to see that on average the probability is 9/47, and we get the correct result by ignoring the fact that some cards are unavailable because they are in other players' hands. All that matters is how many *unseen* cards remain. Of course, in games where some of the opponents' cards are seen, these should be taken into account.

We make a few comments concerning straights. If you have

$$7\heartsuit,\ 8\diamondsuit,\ 9\clubsuit,\ 10\heartsuit,\ A\spadesuit \tag{1.1.18}$$

you can discard the Ace and draw to the *open-ended straight*; there are 8 cards that help you (4 Sixes, 4 Jacks), so the probability of making the straight is 8/47.

If you have instead

$$7\heartsuit,\ 8\diamondsuit,\ 9\clubsuit,\ J\heartsuit,\ Q\spadesuit \tag{1.1.19}$$

you can discard the Seven and draw to the *inside straight*. Here there are only 4 Tens that help you, so the probability of making your straight is 4/47. One sometimes hears the advice "never draw to an inside straight", but this can be correct if the pot odds are right.

(Of course, you can also discard the Seven and Eight and draw 2 to the King high straight, but you must draw both a Ten and a King to make this hand. The probability of this is $\binom{4}{1} \cdot \binom{4}{1} \Big/ \binom{47}{2} \approx .014801$. We hope our opponents routinely make draws like this! When they make the hand, congratulate them on their play and be sure to send a limo to pick them up for the next game!)

5 card draw is not played so much anymore; another game which is much more popular is *(Texas) Hold'em*. This is played as follows. First, there are forced "blind" bets usually by 2 of the players (this rotates) which act as antes (again so there is something to play for; generally one of the blind bets is twice the other.) Then each player receives 2 cards. In order to play, each player must call the larger of the blind bets, and the player making the small blind bet must put in enough to equal the large blind bet if she wishes to continue in the hand. Raises are also allowed. (See [4] for a more complete description of the mechanics of this and other poker games.) Next, the dealer puts 3 cards from the deck on the table; this is called the "flop" (usually the cards are dealt out overlapping and face

down and then "flopped" over.) There is then another round of betting. Then a 4th card is dealt on the table (this is called the "turn"), and there is more betting. Finally, a 5th card is dealt on the table (this is called the "river"), and there is a final round of betting. Usually the bet limit on the last 2 cards is twice that after the deal and the flop (when there are limits on the bet size.) Then the hands are shown down; each player makes the best 5-card poker hand using 0, 1, or 2 of his cards, and 5, 4, or 3 of the cards on the table (called the "board" cards) which are community cards used by all players. (It is fairly rare that a player uses none of his original 2 cards to make his hand; this is "playing the board", and could happen, for example, if there were a flush or 4 of a kind on the board. If no one could beat the board, the pot would be split among all players.) Hold'em is probably the most common casino poker game and is the game played in the final \$10,000 buy-in event at the World Series of Poker (which takes place every year in May at the Horseshoe in Las Vegas; this event is played no-limit.) There are many interesting and fairly easy probabilistic calculations that can be done for this game, and we present some of these.

Good starting hands are pairs ("pocket pairs"), the larger the better, and big cards, better if "suited" (*i.e.*, both of the same suit.) Almost everyone agrees that a pair of Aces is the best starting hand, followed by a pair of Kings. There is disagreement about what is next, some preferring QQ, then perhaps JJ, others rating AK suited above QQ. But certainly AK. AQ, are strong hands, although they are drawing hands - with AK you don't have anything really strong before the flop. Small pairs usually need another card of the same rank on the flop in order to continue. For example, if you start with 22 and the flop contains no 2, the hand is pretty much worthless if there is any action - any card can pair a card in someone's hand to beat you. If you have a pair and another card of that rank appears on the board, then you have a "set". (If there are 2 cards of a given rank on the board matching one in your hand, you have "trips" - this is not nearly as strong as having a set.) Suppose you have a pair, say 22. What is the probability of making a set on the flop (*e.g.* the probability that there is at least one more 2 on the flop)? For this to happen, the flop must contain one 2 and 2 from among 48 other cards or two 2s and 1 from among 48 other cards, so the probability is

$$\frac{\binom{2}{1}\cdot\binom{48}{2}+\binom{2}{2}\cdot 48}{\binom{50}{3}} \approx .117551. \tag{1.1.20}$$

This could also be calculated by computing 1 $-$ the probability that there is no 2 on the flop. (The latter is $\binom{48}{3}\Big/\binom{50}{3}$.)

Remark: It is easy to reason incorrectly in some of these calculations. As an example, what is wrong with the reasoning that to have another 2 on the flop, we need to choose one 2 in $\binom{2}{1}$ ways and then 2 of 49 other cards (one of which could be another 2.) This gives

$$\frac{\binom{2}{1}\cdot\binom{49}{2}}{\binom{50}{3}} = .12. \quad (\textbf{Wrong!}) \tag{1.1.21}$$

(*Hint*: Somethings are being counted twice in (1.1.21).)

Even though it is 8 to 1 against making a set on the flop, if you make it you may win a big pot and get significantly better than 8 to 1 on your investment.

Here is an interesting calculation. Suppose you are "heads-up" against an opponent who has AK, and you have 22. ("Heads-up" means only the two of you are contesting the pot.) It is non-trivial to compare these hands in such a way that every possibility is taken into account, but one can get a fairly good estimate of what happens if these are played to the end against each other. In order to win, the 22 generally needs that no other A or K appears in the 5 board cards. The probability of this is

$$\binom{42}{5}\Big/\binom{48}{5} \approx .503203. \tag{1.1.22}$$

This indicates that the 22 is a slight favorite over AK when played to the end; in fact, a simulation (dealing out random hands using Mike Caro's Poker Probe software) shows that actually the 22 is about an 11 to 10 favorite when played to the end. This, however, is quite misleading. Hands are generally not played to the end, and the AK is a **much** beter hand to have. The problem with 22 is that, as indicated earlier, any card can potentially beat you, and you never know where you are in the hand. The AK can win if an A or K comes on the flop and can fold if there is no improvement and a lot of action. (If there is not much action on the flop, the AK can continue and hope for improvement on the turn, *etc.*) Simulations are useful in many contexts, but often don't reflect the dynamics of actual play.

Suppose you "pick up a draw" on the flop (this means that you have 4 to a straight or flush and hope to make it in the next 2 cards. You might also have 4 to a straight and 4 to a flush, or 4 to a flush and a big pair; these give you extra possibilities.) The number of cards which will complete your hand is your "number of outs". If you have a 4-flush, then you have 9 outs; with 4 to a straight, you have 4 or 8 outs. If you have $A\heartsuit$, $J\heartsuit$, and the board is $10\heartsuit$, $5\heartsuit$, $Q\spadesuit$, then any of 9 $\heartsuit$s will win for you, and K ($\spadesuit, \clubsuit,$ or $\diamondsuit$) makes you the best straight (and will likely win the whole pot), so you have 12 outs. With 2 cards to come, the probability that one of the 12 cards comes that makes your hand is

$$1-\frac{\binom{35}{2}}{\binom{47}{2}}\approx .449584 \tag{1.1.23}$$

which is quite good. Note that with 14 outs and 2 cards to come, the probability of making your hand is

$$1-\frac{\binom{33}{2}}{\binom{47}{2}}\approx .511563 \tag{1.1.24}$$

so you are a favorite to make the hand.

The probability of picking up a flush draw on the flop (if you start with 2 suited cards) is

$$\frac{\binom{11}{2}\cdot\binom{39}{1}}{\binom{50}{3}}\approx .109439 \tag{1.1.25}$$

or about 9 to 1 against. The probability of actually completing the flush on the flop is

$$\frac{\binom{11}{3}}{\binom{50}{3}}\approx .008418 \tag{1.1.26}$$

or more than 100 to 1 against. If only one card of the suit you hold appears on the flop, then you have a "backdoor flush draw". The probability of making the flush (the last 2 cards must be your suit) is then

$$\frac{\binom{10}{2}}{\binom{47}{2}}\approx .041628. \tag{1.1.27}$$

Finally, if you start with 2 suited cards, the probability that by the 5th card you have made a flush is

$$\frac{\binom{11}{3}\binom{39}{2}+\binom{11}{4}\cdot\binom{39}{1}+\binom{11}{5}}{\binom{50}{5}}\approx .063998 \tag{1.1.28}$$

which is about 15 to 1 against. (It is usually not worth playing 2 suited cards only to try for a flush. Of course, if the cards you have don't have much value except that they are suited, then they are probably not too large, and the flush you could make won't be the nut flush, so playing for a flush in this situation is even worse than (1.1.28) indicates.)

The situation that often arises that is usually worth further play is the following. Suppose you have a 4-flush on the flop. With 9 outs, the probability of making your hand is

$$1-\frac{\binom{38}{2}}{\binom{47}{2}}\approx .349676 \tag{1.1.29}$$

so the odds against making the flush with 2 cards to come are approximately 2 to 1. If you figure on getting somewhat better than 2 to 1 on your investment if you play to the end, then you will probably play. (However, you have to also consider the possiblilty that you make the flush and lose to a better hand. This could happen if the flush cards you hold are not very big, or the board pairs and someone makes a full house, *etc.*) Another problem occurs if you don't make the flush on the turn. Then with one card to come, the probability of making the flush is

$$1-\frac{\binom{37}{1}}{\binom{46}{1}}=\frac{9}{46}\approx .195652 \tag{1.1.30}$$

so you are approximately a 4 to 1 underdog.

It is difficult to do many of the calculations presented above without a calculator, but one can remember some of the results and use them during actual play. The calculations involving "outs" after the flop are fairly easy, however, and can be done quickly if one estimates a bit. The following is a simple rule that can be useful in these situations. Since $1/47$ is slightly more than .02, when we want to calculate the probability of completing a Hold'em hand with 1 or 2 cards to come, we can multiply the number of outs by .02 if there is 1 card to come and double this figure if there are 2 cards to come. (The results with 1 card to come are underestimated a little, and the ones with 2 cards to come are a bit larger than the true values but the results are still fairly good estimates, especially when the number of outs is not very large.) For example, completing a 4-flush with 1 card to come: 9 outs $\times$ .02 = .18 or approximately 4 to 1 against since this method underestimates a bit. For completing a 4-flush with 2 cards to come: 9 outs $\times$ .02 $\times 2$ = .36 or approximately 2 to 1 against. Another example - with 14 outs and 2 cards to come, $14 \times .02 \times 2 = .56$ (a favorite to make the hand; note that the actual value computed in (1.1.24) is .511563, so our estimate is a little too large.)

2. Foundations and Basic Results

The basic mathematical ingredients of a probabilistic model are a *sample space*, S consisting of the elementary outcomes of some experiment $\mathcal{E}$ and a *probability* (or *probability measure*) defined on the events (a class of subsets) of S. We refer to the pair $\{S, P\}$ as a *probability space*.

We begin with some general remarks on sets and set theory. A set is a collection of objects referred to as the *members* or *elements* of S. (Actually this is too broad a notion and leads to inconsistencies which can only be removed by a careful and systematic development which is a whole subject in itself and can hardly be even hinted at here. Even so, one can't prove that inconsistencies don't remain, but the obvious ones can be eliminated. The type of treatment we pursue is usually referred to as *naive set theory*, although we barely scratch the surface of that approach.) The symbol $\in$ is used to denote membership, *i.e.,*

$$s \in S \tag{2.1}$$

is read "s is a member of the set S". If s is **not** a member of S, we write

$$s \notin S. \tag{2.2}$$

The *universal set*, $\mathcal{U}$, is the largest set one is interested in; for example, if one were considering sets of real numbers, then the universal set might be taken to be the set, $\mathbb{R}^1$, of all real numbers, and in probabilistic models, the universal set is usually the sample space S. The *empty set*, $\emptyset$, is the set with no members ($\emptyset = \{\}$.) For sets A and B, we say A is a *subset* of B, and write $A \subset B$ if every member of A is a member of B (this includes the possibility that $A = B$.) The basic operations on sets are *union* (denoted $\cup$), *intersection* (denoted $\cap$), and *difference* (denoted $\setminus$.) For sets A and B,

$$A \cup B = \{x | x \in A \text{ or } x \in B\}, \qquad A \cap B = \{x | x \in A \text{ and } x \in B\},$$

$$A \setminus B = \{x | x \in A \text{ and } x \notin B\}. \tag{2.3}$$

(For the construct $A \setminus B$, it is not assumed or necessary that $A \subset B$.) A special case of set difference (one of the most common cases) occurs when there is an implicit universal set $\mathcal{U}$. Then one denotes the difference $\mathcal{U} \setminus A$ by $\overline{A}$ or A^c; this is called the *complement* of A. Some important relations involving these operations are

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \tag{2.4}$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \tag{2.5}$$

$$\overline{A \cup B} = \overline{A} \cap \overline{B}, \qquad \overline{A \cap B} = \overline{A} \cup \overline{B} \tag{2.6}$$

((2.4) and (2.5) are called distributive laws for union over intersection and intersection over union respectively, and (2.6) are the DeMorgan formulas.) The operations of union and intersection and various relations extend to arbitrary families of sets; *e.g.*, if $\{A_i\}_{i \in I}$ is a family of sets indexed by I, then one defines

$$\bigcup_{i \in I} A_i = \{x | x \in A_i \text{ for some } i \in I\}, \qquad \bigcap_{i \in I} A_i = \{x | x \in A_i \text{ for all } i \in I\} \tag{2.7}$$

and the DeMorgan formulas are $\overline{\bigcup_{i \in I} A_i} = \bigcap_{i \in I} \overline{A}_i$, *etc.*

(*Exercise*: What are $\bigcup_{i \in I} A_i$ and $\bigcap_{i \in I} A_i$ if $I = \emptyset$?)

Index

The Author

Chas. Friedman was born at an early age. After an uneventful childhood, he arrived at the point of choosing a career; possibilities considered were mandolin builder, traveling tuba player and mathematician. Following some meditation on the matter, he decided on the latter, attending the Graduate School at Princeton which resulted in a Ph.D. in 1971. He spent 2 years as an instructor at M.I.T., and then came to University of Texas-Austin in 1973 where he has remained to the present time. His interests have included mathematical physics, differential equations and probability, and he has published articles on these subjects and others (*e.g.* number theory.) He resides in the Texas Hill Country near San Marcos and Wimberley with his wife, 3 dogs, 4 cats, 2 miniature goats and a pot-bellied pig. In his spare time, he plays various instruments, does woodworking and makes jewelry, plays poker with the local gamblers and thinks about mathematics and its applications.