Preface

Education is not preparation for life, education is life itself.
John Dewey

Probability and Statistics are studied by most students of Tomsk Polytechnic University, usually as a second- or third year course. The text will be useful for students with diverse backgrounds, and a broad range of interests: from freshmen to beginning graduate students, from the engineering to management departments. In essence, the secondary school mathematics is enough for understanding Probability I (excluding time series, but the parts involving this technique can be omitted for first reading). The textbook contains the Review of Elementary Mathematical Prerequisites.

The text does not emphasize an overly rigorous or formal view of probability. Some of the more mathematically rigorous analysis has been just sketched or intuitively explained in the text. The main objective for this course is to learn probabilistic and statistical thinking, to emphasize more on concepts and finally to foster active self-learning. Many current texts in the area are just cookbooks and, as a result, students do not know why they perform the methods they are taught, or why the methods work. This text is an attempt to readdresses these shortcomings. Numerous easy to more challenging exercises are provided, often from real life, to show how the fundamentals of probabilistic and statistical theories arise intuitively.

Plugging numbers into the formulas and crunching them have no value by themselves. You should continue to put effort into the concepts and concentrate on interpreting the results.

The concept of probability occupies an important place in the decision-making process under uncertainty, whether the problem is one faced in business, in government, in sciences, or just in one's own everyday personal life. Most decisions are made in the face of uncertainty. In a broad sense probability theory can be understood as a mathematical model for the intuitive notion of uncertainty. The statistician’s prime concern lies in drawing conclusions or making inferences from experiments which involve uncertainties. The concepts of probability make it possible for the statistician to generalize from the known (sample) to the unknown (population) and to place a high degree of confidence in these generalizations. In essence, probability is the “machinery” that allows us to draw conclusions from the sample to the population.

This course introduces the basic notions of probability theory and develops them to the stage where one can begin to use probabilistic ideas in statistical inference and modeling. Probability I. Discrete Distributions includes probability axioms and properties, conditional probability and independence, main discrete distributions. Probability II will be devoted to continuous distributions, which need more sophisticated mathematical technique.

The text is not original one. It partly consists of materials thoroughly chosen from Internet resources, probability and statistics textbooks that will be of interest and useful for undergraduates. A lot of exercises and some theoretical topics were taken from “Lecture Notes. Probability” by Jan Vrbik from Brock University (Canada) and “Syllabus” by Joran Elias from Montana University (USA); a lot of materials (especially concerning puzzles, generalities, history of probability) − from the books by C.M. Grinstead (Swarthmore College, USA), and J.L. Snell (Dartmouth College, USA) “Introduction to Probability” (1997) and “Elementary Probability” by David Stirzaker (2003, Cambridge University Press). At the end of the text there are references to the resources and useful links.

I’d be glad to have corrigenda and other suggestions (kit1157@yandex.ru).

Anna Kitaeva
December 2010

Content

Introduction

The most important questions of life are,
for the most part, really only problems of probability.
Laplace, Pierre Simon

Probability is the very guide of life.
Cicero.

Uncertainty and randomness are unavoidable aspects of our everyday life. Your income and spending are subject to erratic strokes of good or bad fortune. Your genetic makeup is a random selection from those of your parents. The weather is notoriously fickle in many areas of the globe. You may decide to play cards, invest in shares, bet on horses, buy lottery tickets, or engage in other forms of gambling; the results are necessarily uncertain (otherwise, gambling could not occur).

At a different level, society has to organize itself in the context of similar sorts of uncertainty. Engineers have to build structures to withstand stressful events of unknown magnitude and frequency. The telephone network, call centers, and airline companies with their randomly fluctuating loads could not have been economically designed without calculations of chances what is the subject of the probability theory. Any system should be designed to have a small chance of failing and a high chance of performing as it was intended. Financial markets of any kind should function so as to share out risks in an efficient and transparent way, for example, when you insure your car or house, buy an annuity, or mortgage your house. The stock market, "the largest casino in the world," cannot do without probability theory.

Everyone must have some internal concept of chance to live in the real world, although such ideas may be implicit or even unacknowledged. Concepts of chance have long been incorporated into many cultures in mythological or religious form. The Romans, for example, had gods of chance named Fortuna and Fors, and even today Englishmen have Lady Luck. The casting of lots to make choices at random is widespread; we are all familiar with "the short straw" and the "lucky number".

The subject of probability can be traced back to the 17th century when it arose out of the study of gambles. And nowadays the range of applications extends beyond games into business decisions, insurance, law, medical tests, and the social sciences. Probabilistic modeling is used to control the flow of traffic through a highway system, a telephone interchange, or a computer processor; find the genetic makeup of individuals or populations; quality control; investment; and other sectors of business and industry.

Our task is to find a way of describing and analyzing the concepts of chance and uncertainty that we intuitively see are common to the otherwise remarkably diverse examples mentioned above.

Uncertain things are susceptible to judgment and insight. We know that casinos invariably make profits; we believe that it does not really matter whether you call heads or tails when a coin is flipped; we learn not to be on top of the mountain during a thunderstorm; and so on. Most people would agree that in roulette, black is more likely than green (the zeros); a bookmaker is more likely to show a profit than a loss on a book; the chance of a thunderstorm is greater later in the day; and so on. The point is that because probabilities are often comparable in this way it is natural to represent them on a numerical scale. Agreeing that probability is a number is the first step on our path to constructing our model of chance.

The advantage of modeling is concise description of otherwise incomprehensibly complicated systems. It provides not only a description of corresponding systems, but also predictions about how they will behave in the future. Adequate models may predict how the systems would behave in different circumstances, or shed light on their (unobserved) past behavior.

As mentioned above a primitive model for chance, used by many cultures, represents it as a supernatural entity, or God. Unfortunately, it is useless for practical purposes, such as prediction and judgment. There is no evidence to suggest that any of various techniques, such as augury or casting lots (sortilege) to discover Fortune's inclination rate better than useless.

Fortunately, our experience has shown that we can do much better by using a mathematical model. First, a useful model must be simpler than reality. Mathematical models have this stripped-down quality in abundance. Second, mathematical models are abstract and are therefore quite unconstrained in their applications. When we define the probabilities of events, and the rules that govern them, our conclusions will apply to all events of whatever kind (e.g., insurance claims, computer algorithms, crop failures, scientific experiments, games of chance and so on). Third, the great majority of practical problems about chance deal with questions that either are intrinsically numerical or can readily be rephrased in numerical terms. The use of a mathematical model becomes almost inescapable. Fourth, if you succeed in constructing a model in mathematical form, then all the power of mathematics developed over several thousand years is instantly available to help you use it.

It turns out that we can make great progress by using the simple fact that our ideas about probability are closely linked to the familiar mathematical ideas of proportion and ratio.

1. Basic Concepts

1.1. Random experiments. Events.

Creativity is the ability
to introduce order into the randomness of nature.
Eric Hoffer

We need to develop a language for construction the model. Let's start.

An **experiment (random experiment)** is a process leading to distinct, well-defined possibilities called **outcomes** (with uncertainty as to which outcome will occur).

For example, if our experiment is to roll one die, then there are six outcomes corresponding to the number that shows on the top. The set of all outcomes in this case is $\{1, 2, 3, 4, 5, 6\}$. It is called the **sample space** and is usually denoted by Ω.

Things get a little more interesting when we roll two dice. We have a decision to make: do we want to consider the dice as indistinguishable (to us, they usually are) and have the sample space consist of unordered pairs of numbers, or should we mark the dice (red and green say) and consider an ordered pair of numbers as an outcome of the experiment (the first number for red, the second one for green die)? The choice is ours; we are allowed to consider as much or as little detail about the experiment as we need, but there are two constraints:

(a) We have to make sure that our sample space has enough information to answer the questions at hand (if the question is: what is the probability that the red die shows a higher number than the green die, we obviously need the ordered pairs).

(b) Subsequently, we learn how to assign probabilities to individual outcomes of a sample space. This task can quite often be greatly simplified by a convenient design of the sample space. It just happens that, when rolling two dice, the simple events (pairs of numbers) of the sample space have the same probability of 1/36 when they are ordered; assigning correct probabilities to the unordered list would be difficult.

That is why, for this kind of experiment (rolling a die any fixed number of times or rolling a few dice), we always choose the sample space to consist of an ordered set of numbers (whether the question requires it or not).

We can write the outcomes of this experiment as (m, n), where m is the number on the red die and n is the number on the green die. To visualize the set of outcomes it is useful to make a little table (the table represents the sample space Ω):

(1,1) (2,1) (3,1) (4,1) (5,1) (6,1)
(1,2) (2,2) (3,2) (4,2) (5,2) (6,2)
(1,3) (2,3) (3,3) (4,3) (5,3) (6,3)
(1,4) (2,4) (3,4) (4,4) (5,4) (6,4)
(1,5) (2,5) (3,5) (4,5) (5,5) (6,5)
(1,6) (2,6) (3,6) (4,6) (5,6) (6,6)

There are $36 = 6\times6$ outcomes since there are 6 possible numbers to write in the first slot and for each number written in the first slot there are 6 possibilities for the second.

The goal of probability theory is to compute the probability of various events of interest. Intuitively, an event is a statement about the outcome of an experiment. The formal definition is: an **event** is a subset of the sample space. We will denote events by capital letters A, B, C, ...

For example, the event "The sum is 8" for the experiment of rolling two dice $A = \{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}$. This event contains 5 of the 36 possible outcomes. For a second example, consider B = "There is at least one six". B consists of the last row and last column of the table so it contains 11 outcomes.

Correspondence between set terminology and probabilistic terminology:

Set terminology	*Probabilistic terminology*
Set	Event (random event)
Universal set Ω	Sample space Ω (the widest possible event, a sure (or certain) event)
Elements of Ω	Outcomes, or elementary events (the simplest possible events)
Empty set ϕ	Impossible event
Complement of A	Opposite (complementary, or contrary) event
Disjoint sets	Mutually exclusive events.

I recommend you to look at Appendix 2 (Sets).

1.2. Probability

All models are wrong but some are useful.
George Box

Many occurrences of probability appear in everyday statements such as:

(1) The probability of a spade on cutting a pack of cards is 25 %.

(2) The probability of red in (American) roulette is 18/38.

(3) The probability of a head when you flip a coin is 50 %.

Many other superficially different statements about probability can be reformulated to appear in the above format. This type of statement is in fact so frequent and fundamental that we use a standard abbreviation and notation for it. Anything of the form “the probability of A is p” will be written as: $P(A) = p$.

In the examples above, A and p were, respectively,

$A = \{\text{spade}\}, p = 25\ \%$;

$A = \{\text{red}\}, p = 18/38$;

$A = \{\text{head}\}, p = 50\ \%$.

We can express probability using percentages, for everyday language it doesn’t matter, but in mathematics the probability is a number between 0 and 1 inclusively.

Abstractly, a **probability** is a function that assigns numbers to events, which satisfies:

(i) For any event A, $0 \le P(A) \le 1$.

(ii) If Ω is the sample space then

$P(\Omega) = 1$ (completeness axiom or normalization requirement).

(iii) If A and B are disjoint, i.e., $P(A \cap B) = \phi$; then

$P(A \cup B) = P(A) + P(B)$ (finite additivity axiom).

(iv) If $A_1, A_2, \ldots$ is an infinite sequence of pairwise disjoint events (i.e., $A_i \cap A_j = \phi$; when $i \neq j$) then

$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ (countable additivity axiom).

These assumptions are motivated by the **frequency interpretation of probability**, which states that if we repeat an experiment a large number of times (provided the experiment was made under similar conditions) then the fraction of times the event A occurs will be close to $P(A)$. To be precise, if we let $N(A, n)$ be the number of times A occurs in the n trials then

$$P(A) = \lim_{n\to\infty} \frac{N(A,n)}{n}. \qquad (1.2.1)$$

We will see this result as a theorem called the law of large numbers.

This (the existence of the limit above for some kinds of events) is an empirical fact (so called **the fact of frequency stability**). If it does not occur our world would be quite different.

Thus, evaluating the ratio $N(A, n)/n$ (relative frequency) offers a way of estimating the probability $P(A)$ of the event A. The larger the sample size n, the better will be the approximation to the truth.

A similar empirical observation was recorded by John Graunt in his book "Natural and Political Observations Made Upon the Bills of Mortality" published in 1662. He found that in a large number of births, the proportion of boys born was approximately 14/27. This came as something of a surprise at the time, leading to an extensive debate. We simply interpret the observation in this statement: the probability of a yet unborn child being male is approximately 14/27. Note that this empirical ratio varies slightly from place to place and time to time, but it always exceeds 1/2 in the long run.

For the moment we will use the interpretation (1.2.1) of $P(A)$ to motivate the definition above.

Given (1.2.1), (i) and (ii) are clear. The fraction of times that a given event A occurs must be between 0 and 1, and if Ω has been defined properly (recall that it is the set of all possible outcomes), the fraction of times something in Ω happens is 1. To explain (iii), note that if the A and B are disjoint then $N(A \cup B, n) = N(A, n) + N(B, n)$ since $A \cup B$ occurs if either A or B occurs but it is impossible for both to happen.

Dividing by n and letting $n \to \infty$, we arrive at (iii).

Assumption (iv) is a little controversial. Some have argues passionately that it should not be imposed but we need the assumption for construction a mathematical model of chance (Kolmogorov's axiomatics). In many cases the sample space is finite so (iv) is not relevant.

In discrete case (countable or finite sample space) you can formulate the definition in short-and easy-to-remember way:

a probability is a nonnegative additive function defined on some sample space Ω and satisfied the normalization requirement ($P(\Omega) = 1$).

Note, that condition (i) will be follow from this definition.

When the sample space is uncountable you need more complex construction using subsets of Ω (events), called σ-field, to define probabilities.

We can also assign probabilities speculatively. Return to (1) − (3) statements.

The question is where did those values for the probability p come from? Are they from monitoring the processes? Or, may be, we can receive the results from some theoretical considerations? Let's try.

(1) Let us consider what happens when we pick a card at random from a conventional pack. There are 52 cards, of which 13 are spades. The implica-

tion of the words "at random" is that any card is equally likely to be selected, and the proportion of the pack comprising the spade suit is 13/52 = 1/4. Our intuitive feelings about symmetry suggest that the probability of picking a spade is directly proportional to this fraction, and by convention we choose the constant of proportionality to be unity. Hence, P(spade) = 1/4 = 25 %.

Exactly the same intuitive interpretation comes into play for any random procedure having this kind of symmetry.

(2) *American Roulette*: the wheels have 38 compartments, of which 18 are red, 18 are black, and two are green (the zeros). If the wheel has been made with equal size compartments (and no hidden magnets, or subtle asymmetries), then the ball has 18 chances to land in red out of the 38 available.

This suggests P(red) = 18/38.

(3) In the case of a fair coin, of course, there are only two equally likely chances to P(head) = 50 % and P(tail) = 50 %. This particular case of equal probabilities has passed into the language in the expression a "fifty-fifty" chance.

Return to experiment of rolling one (or two) die (dice). Symmetry dictates that all outcomes are equally likely so each has probability 1/6 (or 1/36). In general the probability of an event C concerning the roll of one die (or two dice) is the number of outcomes in C divided by 6 (or 36). For example, for two dice P(The sum is 8) = 5/36, P(There is at least one six) = 11/36.

In general, this argument (or expression of our intuition) leads to the following definition of probability. Suppose that an experiment has n distinct possible outcomes, and suppose further that by symmetry (or by construction or supposition) these outcomes are equally likely. Then if A is any collection of r of these outcomes ($|A| = r$), we define

$$P(A) = \frac{r}{n} = \frac{|A|}{|\Omega|} = \frac{\text{number of outcomes in } A}{\text{total number of outcomes}}.$$

This formula is called **classical probability**. Many students remember the formula but don't think about the conditions under which it is true.

Note that in this case conditions (i)–(iii) of the definition above are fulfilled.

This idea or interpretation of probability is very appealing to our common intuition. But it is crucial a finite number of outcomes there, it is not always the case (see examples 4 and 5 below).

Also there are plenty of random procedures with no discernible symmetry in the outcomes and there are many non-repeatable random events. Classic examples include horse races, football matches, and elections. When the theoretical and frequency approaches are not applicable, one can use subjective probabilities assigned by experts. For example, performance of economic projects depends on overall macroeconomic conditions. In project evaluation the resulting statements sound like this: with probability 0.6 the economy will be on the rise, and then the profitability of the project will be this much and so on.

There are indeed many other theories devised to explain uncertain phenomena. Attempts have been made to construct more general theories with a smaller set of rules. So far, none of these is in general use.

Note this remark of William Feller: "All definitions of probability fall short of the actual practice". We deal with the model only, but this model can help us to make right decisions in everyday life and business. Most people's intuition about problems in chance will often lead them grossly astray, even with very simple concepts. We mention a few examples here.

1) *The base rate fallacy.* This appears in many contexts, but it is convenient to display it in the framework of a medical test for a disease that affects statistically one person in 100,000. You have a test for the disease that is 99 % accurate. (That is to say, when applied to a sufferer, it shows positive with probability 99 %; when applied to a non sufferer, it shows negative with probability 99 %.) What is the probability that you have the disease if your test shows a positive result? Most people's untutored intuition would lead them to think the chance is high, or at least not small.

In fact, the chance of having the disease, given the positive result, is less than one in a 1,000; indeed, it is more likely that the test was wrong.

2) *The Monty Hall Problem*. You are a contestant in a game show. A nice car and two feral goats are randomly disposed behind three doors, one to each door. You choose a door to obtain the object it conceals. The presenter does not open your chosen door, but opens another door that turns out to reveal a goat.

Then the presenter offers you the chance to switch your choice to the final door.

Do you gain by so doing? That is to say, what is the probability that the final door conceals the car? Many people's intuition tells them that, given the open door, the car is equally likely to be behind the remaining two; so there is nothing to be gained by switching. In fact, this is wrong; you should switch.

3) *Coincidences.* For many people, the famous birthday problem is another example of contradiction with their own intuitions. Twenty-three randomly selected people are listening to a lecture on chance. What is the probability that at least two of them were born on the same day of a year? Untutored intuition leads most people to guess that the chances of this are rather small. In fact, in a random group of 23 people, it is more likely than not that at least two of them were born on the same day of a year. (See

http://www.newton.cam.ac.uk/wmy2kposters/july/ − poster in the London Underground, July 2000.)

This list of counterintuitive results could be extended, but this should at least be enough to demonstrate that only mathematics can save people from their flawed intuition with regard to chance events.

1.3. Historical Remarks

It is remarkable that a science which began
with the consideration of games of chance
should have become the most important object of human knowledge.
Laplace, Pierre Simon

An interesting question in the history of science is: Why was probability not developed until the sixteenth century? We know that in the sixteenth century problems in games of chance made people start to think about probability. But gambling is almost as old as civilization itself. In ancient Egypt a game now called "Hounds and Jackals" was played. In this game the movement of the hounds and jackals was based on the outcome of the roll of four-sided dice made out of animal bones called astragali (the ankle bones of sheep.). Six-sided dice made of a variety of materials date back to the sixteenth century B.C. Gambling was widespread in ancient Greece and Rome. Indeed, in the Roman Empire it was sometimes found necessary to invoke laws against gambling.

Why, then, were probabilities not calculated until the sixteenth century?

Several explanations have been advanced for this late development.

One is that the relevant mathematics was not developed and was not easy to develop. The ancient mathematical notation made numerical calculation complicated, and our familiar algebraic notation was not developed until the sixteenth century. However, many of the combinatorial ideas needed to calculate probabilities were discussed long before the sixteenth century.

Since many of the chance events of those times had to do with religious affairs, it has been suggested that there may have been religious barriers to the study of chance and gambling.

Another suggestion is that a stronger incentive, such as the development of commerce, was necessary.

However, none of these explanations seems completely satisfactory, and people still wonder why it took so long for probability to be studied seriously.

The first person to calculate probabilities systematically was Gerolamo Cardano (1501–1576) in his book "Liber de Ludo Aleae".

In the book Cardano dealt only with the special case that we have called the uniform distribution function. This restriction to *equiprobable outcomes* was to continue for a long time. In this case Cardano realized that the probability that an event occurs is *the ratio of the number of favorable outcomes to the total number of outcomes*.

Many of Cardano's examples dealt with rolling dice. Here he realized that the outcomes for two rolls should be taken to be the 36 ordered pairs (i,j) rather than the 21 unordered pairs. This is a subtle point that was still causing problems much later for other writers on probability. For example, in the eighteenth century the famous French mathematician d'Alembert, author of several works on probability, claimed that when a coin is tossed twice the number of heads that turn up would be 0, 1, or 2, and hence we should assign equal probabilities for these three possible outcomes. Cardano chose the correct sample space for his dice problems and calculated the correct probabilities for a variety of events.

Cardano's mathematical work is interspersed with a lot of advice to the potential gambler in short paragraphs, entitled, for example: "Who Should Play and When," "Why Gambling Was Condemned by Aristotle," "Do Those Who Teach Also Play Well?" and so forth. In a paragraph entitled "The Fundamental Principle of Gambling" Cardano writes: "The most fundamental principle of all in gambling is simply equal conditions, e.g., of opponents, of bystanders, of money, of situation, of the dice box, and of the die itself. To the extent to which you depart from that equality, if it is in your opponent's favor, you are a fool, and if in your own, you are unjust."

Cardano did make mistakes, and if he realized it later he did not go back and change his error. For example, for an event that is favorable in three out of four cases, Cardano assigned the correct odds 3 : 1 that the event will occur. But then he assigned odds by squaring these numbers (i.e., 9 : 1) for the event to happen twice in a row. Later, by considering the case where the odds are 1 : 1, he realized that this cannot be correct and was led to the correct result that when m out of n outcomes are favorable, the odds for a favorable outcome twice in a row are $m^2 : (n^2 - m^2)$. This is equivalent to the realization that if the probability that an event happens in one experiment is p, the probability that it happens twice is p^2.

Cardano proceeded to establish that for three successes the formula should be p^3 and for four successes p^4, making it clear that he understood that the probability is p^n for n successes in n independent repetitions of such an experiment. This will follow from the concept of independence that we introduce later (or product rule for independent events).

Cardano's work was a remarkable first attempt at writing down the laws of probability, but it was not the spark that started a systematic study of the subject. This came from a famous series of letters between Pascal and Fermat. This correspondence was initiated by Pascal to consult Fermat about problems he had been given by Chevalier de Méré, a well-known writer, a prominent figure at the court of Louis XIV, and an ardent gambler.

The first problem de Méré posed was a dice problem. The story goes that he had been betting that at least one six would turn up in four rolls of a die and winning too often, so he then bet that a pair of sixes would turn up in 24 rolls of a pair of dice. The probability of a "six" with one die is 1/6 and, by the product rule for independent experiments, the probability of two "sixes" when a pair of dice is thrown is $(1/6)\times(1/6) = 1/36$. A gambling rule of this time suggested that, since four repetitions was favorable for the occurrence of an event with probability 1/6, for an event six times as unlikely, $6\times4 = 24$ repetitions would be sufficient for a favorable bet. Pascal showed, by exact calculation, that 25 rolls are required for a favorable bet for a pair of "sixes".

The second problem was a much harder one: it was an old problem and concerned the determination of a fair division of the stakes in a tournament when the series, for some reason, is interrupted before it is completed. This problem is now referred to as the problem of points. The problem had been a standard problem in mathematical texts; it appeared in Fra Luca Paccioli's book "Summa de Arithmetica, Geometria, Proportioni et Proportionalitá", printed in Venice in 1494, in the form: "A team plays ball such that a total of 60 points are required to win the game, and each inning counts 10 points. The stakes are 10 ducats. By some incident they cannot finish the game and one side has 50 points and the other 20. One wants to know what share of the prize money belongs to each side. In this case I have found that opinions differ from one to another but all seem to me insufficient in their arguments, but I shall state the truth and give the correct way."

Reasonable solutions, such as dividing the stakes according to the ratio of games won by each player, had been proposed, but no correct solution had been found at the time of the Pascal-Fermat correspondence. The letters deal mainly with the attempts of Pascal and Fermat to solve this problem.

Blaise Pascal (1623–1662) was a child prodigy, having published his treatise on conic sections at age sixteen, and having invented a calculating machine at age eighteen. At the time of the letters, his demonstration of the weight of the atmosphere had already established his position at the forefront of contemporary physicists.

Pierre de Fermat (1601–1665) was a learned jurist in Toulouse, who studied mathematics in his spare time. He has been called by some the prince of amateurs and one of the greatest pure mathematicians of all times.

In a letter dated Wednesday, 29th July, 1654, Pascal writes to Fermat:

"Sir,

Like you, I am equally impatient, and although I am again ill in bed, I cannot help telling you that yesterday evening I received from M. de Carcavi your letter on the problem of points, which I admire more than I can possibly say. I have not the leisure to write at length, but, in a word, you have solved the two problems of points, one with dice and the other with sets of games with perfect justness; I am entirely satisfied with it for I do not doubt that I was in the wrong, seeing the admirable agreement in which I find myself with you now . . . .

Your method is very sound and is the one which first came to my mind in this research; but because the labor of the combination is excessive, I have found a short cut and indeed another method which is much quicker and neater, which I would like to tell you here in a few words: for henceforth I would like to open my heart to you, if I may, as I am so overjoyed with our agreement. I see that truth is the same in Toulouse as in Paris.

Here, more or less, is what I do to show the fair value of each game, when two opponents play, for example, in three games and each person has staked 32 pistoles.

Let us say that the first man had won twice and the other once; now they play an other game, in which the conditions are that, if the first wins, he takes all the stakes; that is 64 pistoles; if the other wins it, then they have each won two games, and therefore, if they wish to stop playing, they must each take back their own stake, that is, 32 pistoles each.

Then consider, Sir, if the first man wins, he gets 64 pistoles; if he loses he gets 32. Thus if they do not wish to risk this last game but wish to separate without playing it, the first man must say: "I am certain to get 32 pistoles, even if I lost I still get them; but as for the other 32, perhaps I will get them, perhaps you will get them, the chances are equal. Let us then divide these 32 pistoles in half and give one half to me as well as my 32 which are mine for sure." He will then have 48 pistoles and the other 16 . . . ."

Number of games *B* has won

3	0	0	0	
2	8	16	32	64
1	20	32	48	64
0	32	44	56	64
	0	1	2	3

Number of games *A* has won

Figure 1.3.1. Pascal's table

Pascal's argument produces the table illustrated in Figure 1.3.1 for the amount due player *A* at any quitting point. Each entry in the table is the average of the numbers just above and to the right of the number. This fact, together with the known values when the tournament is completed, determines all the values in this table. If player *A* wins the first game, then he needs two games to win and *B* needs three games to win; and so, if the tournament is called off, *A* should receive 44 pistoles.

The letter in which Fermat presented his solution has been lost; but fortunately, Pascal describes Fermat's method in a letter dated Monday, 24th August, 1654.

From Pascal's letter:

"This is your procedure when there are two players: If two players, playing several games, find themselves in that position when the first man needs two games and second needs three, then to find the fair division of stakes, you say that one must know in how many games the play will be absolutely decided.

It is easy to calculate that this will be in four games, from which you can conclude that it is necessary to see in how many ways four games can be arranged between two players, and one must see how many combinations would make the first man win and how many the second and to share out the stakes in this proportion. I would have found it difficult to understand this if I had not known it myself already; in fact you had explained it with this idea in mind."

Fermat realized that the number of ways that the game might be finished may not be equally likely. For example, if *A* needs two more games and *B* needs three to win, two possible ways that the tournament might go for *A* to win are WLW and LWLW. These two sequences do not have the same chance of occurring. To avoid this difficulty, Fermat extended the play, adding fictitious plays, so that all the ways that the games might go have the same length, namely four. He was shrewd enough to realize that this extension would not change the winner and that he now could simply count the number of sequences favorable to each player since he had made them all equally likely. If we list all possible ways that the extended game of four plays might go, we obtain the following 16 possible outcomes of the play:

<u>WWWW</u> <u>WLWW</u> <u>LWWW</u> <u>LLWW</u>
<u>WWWL</u> <u>WLWL</u> <u>LWWL</u> LLWL
<u>WWLW</u> <u>WLLW</u> <u>LWLW</u> LLLW
<u>WWLL</u> WLLL LWLL LLLL.

Player *A* wins in the cases where there are at least two wins (the 11 underlined cases), and *B* wins in the cases where there are at least three losses (the other 5 cases). Since *A* wins in 11 of the 16 possible cases Fermat argued that the probability that A wins is 11/16. If the stakes are 64 pistoles, A should receive 44 pistoles in agreement with Pascal's result. Pascal and Fermat developed more systematic methods for counting the number of favorable outcomes for problems like this. We see that these two mathematicians arrived at two very different ways to solve the problem of points.

On the correspondence between Fermat and Pascal Huygens based a widely read textbook "On Calculating in Games of Luck" (1657), followed by the books of James Bernoulli (1713), Pierre de Montmort (1708), and Abraham de Moivre (1718, 1738, 1756). The probabilistic framework remains in use today, much as in Huygens's book.

The modern period of probability theory is connected with names like S.N. Bernstein (1880–1968), E. Borel (1871–1956), and A.N. Kolmogorov (1903–1987). The rigorous structure of the theory that is most in use today was codified by A. Kolmogorov in his book "Grundbegriffe der Wahrscheinlichkeitsrechnung" (1933).

We note, however, that other systems have been, are being, and will be used to model probability.

Historical information about mathematicians can be found in the MacTutor History of Mathematics Archive (http://www-history.mcs.st-andrews.ac.uk/history/index.html).

1.4. Examples

1. Suppose we pick a letter at random from the word TENNESSEE. What is the sample space and what probabilities should be assigned to the outcomes?

The sample space $\Omega = \{T, E, N, S\}$. To describe the probability it is enough to give the values for the individual outcomes, since (iii) implies that $P(A)$ is the sum of the probabilities of the outcomes in A. Since there are nine letters in TENNESSEE the probabilities are $P(\{T\}) = 1/9$, $P(\{E\}) = 4/9$, $P(\{N\}) = 2/9$, and $P(\{S\}) = 2/9$.

2. *Astragali*. Board games involving chance were known in Egypt, 3000 years before Christ. The element of chance needed for these games was at first provided by tossing astragali, the ankle bones of sheep. These bones could come to rest on only four sides, the other two sides being rounded. The upper side of the bone, broad and slightly convex counted four; the opposite side broad and slightly concave counted three; the lateral side flat and narrow, one, and the opposite narrow lateral side, which is slightly hollow, six.

The outcomes of this experiment are $\{1, 3, 4, 6\}$. There is no reason to suppose that all four sides have the same probability so our model will have probabilities for the four outcomes $p_1, p_3, p_4, p_6 > 0$ that have $p_1 + p_3 + p_4$

$+ p_6 = 1$. To define the probability of an event A we let $P(A) = \sum_{i \in A} p_i$. In words we add up the probabilities of the outcomes in A. With a little thought we see that any probability with a finite set of outcomes has this form.

3. In English language text the 26 letters in the alphabet occur with the following frequencies:

E	13.0 %	H	3.5 %	W	1.6 %
T	9.3 %	L	3.5 %	V	1.3 %
N	7.8 %	C	3.0 %	B	0.9 %
R	7.7 %	F	2.8 %	X	0.5 %
O	7.4 %	P	2.7 %	K	0.3 %
I	7.4 %	U	2.7 %	Q	0.3 %
A	7.3 %	M	2.5 %	J	0.2 %
S	6.3 %	Y	1.9 %	Z	0.1 %
D	4.4 %	G	1.6 %.		

From this it follows that vowels (A, E, I, O, U) are used (7.3+13.0+7.4+7.4+2.7) = 37.8 % of the time.

3 (a)* (*Riddle*). This is an unusual paragraph. I'm curious as to just how quickly you can find out what is so unusual about it. It looks so ordinary and plain that you would think nothing was wrong with it. In fact, nothing is wrong with it! It is highly unusual though. Study it and think about it, but you still may not find anything odd. But if you work at it a bit, you might find out. Try to do so without any coaching!

4. Design a sample space for an experiment: flipping a coin until a head appears.

The new feature of this example (waiting for the first head (H)) is that the sample space is infinite: {H, TH, TTH, TTTH, TTTTH, TTTTTH, . . .}. Eventually, we must learn to differentiate between the discrete (countable) infinity, where the individual simple events can be labeled 1st, 2nd, 3rd, 4th, 5th, . . . in an exhaustive manner, and the continuous infinity (real numbers in any interval). The current example is obviously a case of discrete infinity, which implies that the simple events cannot be equally likely (they would all have the probability of $1/\infty = 0$, implying that their sum is 0, an obvious contradiction). But we can easily manage to assign correct and meaningful probabilities even in this case (try to do it or wait for part 6).

5. Design a sample space for an experiment: rotating a wheel with a pointer.

The rotating wheel has also an infinite sample space (an outcome is identified with the final position − angle − of the pointer, measured from some fixed direction), this time being represented by all real numbers from the interval $[0, 2\pi)$ (assuming that angles are measured in radians). This infinity of simple events is of the continuous type, with some interesting consequences. Firstly, from the symmetry of the experiment, all of its outcomes must be equally likely. But this implies that the probability of each single outcome is zero!

Isn't this a contradiction as well? The answer is no; in this case the number of outcomes is no longer countable, and therefore the infinite sum (actually, an integral) of their zero probabilities can become nonzero (we need them to add up to 1). The puzzle is: how do we put all these zero probabilities together to answer a simple question such as: what is the probability that the pointer will stop in the $[0, \pi/2]$ interval? This can be simply solved by the concept of **geometric probability** (try to find it in the Internet). In general such kinds of problems will require introducing a new concept of the so called probability density. We will postpone this until the second part of the probability course (continuous probabilities).

Note that in this case we don't assign probability for every simple outcome (it is zero), but we can calculate probabilities of some subsets (events) of the sample space $\Omega = [0, 2\pi)$, this set of subsets satisfied some rules is called *an event space* (it is σ-field). For finite sample space we can consider a set of all subsets of Ω as an event space.

1.5. Exercises

N.B. *Unless otherwise stated, coins are fair, dice are regular cubes and packs of cards are well shuffled with four suits of 13 cards for all exercises.*

*All exercises marked * have answers.*

1. In New York City, the leading cause of death on the job is not construction accidents, machinery malfunctions, or car crashes – it is homicide! A federal Bureau of labor Statistics study revealed that of 177 New York City workers who died of injuries sustained on the job last year, 122 were homicide victims. Use this information to estimate the probability that an on-the-job death of a New York City worker is the result of a homicide.
2. Give a possible sample space for each of the following experiments:

(a) An election decides between two candidates A and B.

(b) A two-sided coin is tossed.

(c) A student is asked for the month of the year and the day of the week on which her birthday falls.

(d) A student is chosen at random from a class of ten students.

(e) You receive a grade in this course.

For which of the cases would it be reasonable to assign the uniform distribution (equally likely outcomes)?

3. A man receives presents from his three children, Allison, Betty, and Chelsea. To avoid disputes he opens the presents in a random order. What are the possible outcomes?

4. Let $\Omega = \{a, b, c\}$ be a sample space. Let $P(a) = 1/2$, $P(b) = 1/3$, and $P(c) = 1/6$. Find the probabilities for all eight subsets of Ω.

5. Suppose we pick a number at random from the phone book and look at the last digit.

(a) What is the set of outcomes and what probability should be assigned to each outcome?

(b) Would this model be appropriate if we were looking at the first digit?

6. Consider an experiment in which a coin is tossed three times. Describe in words the events specified by the following subsets of Ω= {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT} (H (heads) and T (tails)):

E = {HHH,HHT,HTH,HTT},
E = {HHH,TTT},
E = {HHT,HTH,THH},
E = {HHT,HTH,HTT,THH,THT,TTH,TTT}.

What are the probabilities of the events?

7. Suppose we roll a red die and a green die. What is the probability the number on the red die is larger than the number on the green die?

8. Two dice are rolled. What is the probability the two numbers will

(a) differ by 1 or less,

(b) the maximum of the two numbers will be 5 or larger?

9. If we flip a coin 5 times, what is the probability that the number of heads is an even number (i.e., divisible by 2)?

10. In Las Vegas, a roulette wheel has 38 slots numbered 0, 00, 1, 2, . . . , 36. The 0 and 00 slots are green and half of the remaining 36 slots are red and half are black. A croupier spins the wheel and throws in an ivory ball. If you bet 1 dollar on red, you win 1 dollar if the ball stops in a red slot and otherwise you lose 1 dollar. Give a sample space for a player who makes 3 bets (1 dollar each) on red. What probabilities should be assigned to the outcomes?

11. A die is loaded in such a way that the probability of each face turning up is proportional to the number of dots on that face. (For example, a six is three times as probable as a two.) What is the probability of getting an even number in one throw?

12. Two red cards and two black cards are lying face down on the table. You pick two cards and turn them over. What is the probability that the two cards are different colors?

13. 20 families live in a neighborhood, 4 have 1 child, 8 have 2 children, 5 have 3 children, and 3 have 4 children. If we pick a child at random what is the probability she comes from a family with 1, 2, 3, 4 children?

14. In Galileo's time people thought that when three dice were rolled, a sum of 9 and a sum of 10 had the same probability since each could be obtained in 6 ways:
9: 1 + 2 + 6, 1 + 3 + 5, 1 + 4 + 4, 2 + 2 + 5, 2 + 3 + 4, 3 + 3 + 3,
10: 1 + 3 + 6, 1 + 4 + 5, 2 + 4 + 4, 2 + 3 + 5, 2 + 4 + 4, 3 + 3 + 4.

Compute the probabilities of these sums and show that 10 is a more likely than 9.

15. Suppose we roll three dice. Compute the probability that the sum is (a) 3, (b) 4, (c) 5, (d) 6, (e) 7, (f) 8.

16. An urn contains fifty balls numbered 1 to 50. Relating to the experiment of drawing a ball, what can you state about the following events (elementary, compound, relations between them):

A – the number of the drawn ball is even;
B – the number of the drawn ball is a multiple of 4;
C – the number of the drawn ball is 5;
D – the number of the drawn ball is a multiple of 5;
E – the number of the drawn ball is a power of 5;
F – the number of the drawn ball is a multiple of 10;
G – the number of the drawn ball is a multiple of 3;
H – the number of the drawn ball is a power of 3;
I – the number of the drawn ball is even.

To do this exercise you need to read Appendix 2 (Recall that an event is a set).

17. Two cards are drawn from a 52-card deck. Consider the events:

A – two clubs are drawn;
B – two cards having a value less than 5 each are drawn;
C – a 7 and a Q are drawn.

Decompose these events in elementary events and specify their numbers.

18. Write the sample space for the following experiments:

(a) drawing of a ball from an urn containing seven balls;
(b) drawing of two balls from two urns (one ball from each), the first containing three green balls and the second two red balls;
(c) drawing of a card from a 24-card deck (from the 9 card upward);
(d) rolling two dice;
(e) choosing three numbers from the numbers 1, 2, 3, 4, 5;
(f) choosing seven letters from the letters a, b, c, d, e, f, g, h.

19. An urn contains nine white balls and four black balls. Find the probability of the following events:

(a) A – drawing a white ball;
(b) B – drawing a black ball.

20. Find the number of all possible outcomes for the following experiments:

(a) rolling three dice; generalization: rolling n dice;

(b) spinning a slot machine with four reels having eight symbols each; generalization: spinning a slot machine with n reels of m symbols each;

(c) dealing a player three cards from a 52-card deck;

(d) dealing two players two cards each from 50 cards;

(e) a race with nine competitors.

If the last exercise is difficult to you please read the next section.

2. Combinatorics

Did you know that there are three kinds
of statisticians − those that can count and those that can't.

Many problems in classical probability can be solved by counting the number of outcomes in an event. Such counting often turns out to also be useful in more general contexts. Now we set out some simple methods of dealing with the commonest counting problems.

The basic rules are easy and are perfectly illustrated in the following examples.

Rule 1 (addition rule). If I have m garden forks and n fish forks, then I have $m + n$ forks altogether.

Rule 2 (multiplication rule). If I have m different spoons and n different forks, then there are $m \times n$ distinct ways of taking a spoon and fork.

These rules can be rephrased in general terms, but the idea is already obvious. The important points are that in Rule 1, the two sets in question are disjoint; that is a fork cannot be both a garden fork and a fish fork. In Rule 2, my choice of spoon in no way alters my freedom to choose any fork (and vice versa).

Real problems involve, for example, catching different varieties of fish, drawing various balls from a number of urns, and dealing hands at numerous types of card games. In the standard terminology for such problems, we say that a number n of objects or things are to be divided or distributed into r classes or groups.

The number of ways in which this distribution can take place depends on whether

(i) The objects can be distinguished or not.

(ii) The classes can be distinguished or not.

(iii) The order of objects in a class is relevant or not.

(iv) The order of classes is relevant or not.

(v) The objects can be used more than once (repetition) or not at all.

(vi) Empty classes are allowed or not.

We generally consider only the cases having frequently applications in probability problems. Other aspects are explored in books devoted to combinatorial theory.

Permutations, Arrangements and Combinations. In how many possible ways can we range n distinct objects (such as a, b, c, d, . . ., z) in a row?

When the first letter is freely chosen in 26 (the amount of letters in English alphabet) ways, the next letter can be freely chosen in 25 ways. The next may be freely chosen in 24 ways, and so on. Hence, by Rule 2, the answer is

$26\times25\times24\times$ . . . . $\times3\times2\times1$ = 26! – the number of **permutations** of 26 objects. The permutation is often called a rearrangement of the *n* objects and denote P_n (read "*n* down *P*," or "*n* lower *P*").

Note that *n*! grows very quickly since $n! = n\times(n-1)!$:

1!=1	7! =5,040
2! =2	8! =40,320
3! =6	9! =362,880
4! =24	10! =3,628,800
5! =120	11! =39,916,800
6! =720	12! =479,001,600.

What if some of these objects are indistinguishable, such as, for example, a, a, a, b, b, c? How many distinct permutations of these letters are there, i.e. how many distinct words can we create by permuting aaabbc?

We can start by listing all 6! permutations, and then establishing how many times each distinct word appears on this list (the amount of its repetitions). Luckily enough, the repetition of each distinct word proves to be the same. We can thus simply divide 6! by this common repetition to get the final answer. To get the repetition of a particular word, such as, for example baacba we first attach a unique index to each letter $b_1a_1a_2c_1b_2a_3$ and then try to figure out the number of permutations of these, now fully distinct, symbols, which keeps the actual word (baacba) intact. This is obviously achieved by permuting the a's among themselves, the b's among themselves, etc. We can thus create 3! (number of ways of permuting the a's) times 2! (permuting the b's) combinations which are distinct in the original 6!-item list, but represent the same word now. (We have multiplied 3! by 2! since Rule 2). The answer is thus 6!/(3!2!1!) = 60 (we have included 1! to indicate that there is only one permutation of the single c, to make the formula complete). The resulting expression is so important to us that we introduce a new symbol $\binom{6}{3,2,1} = \frac{6!}{3!2!1!}$, which we read: 6 choose 3 choose 2 choose 1 (note that the bottom numbers must add up to the top number). It is obvious that the same argument holds for any other unique word of the aaabbc type.

It should now be obvious that, in the case of permuting n_1 a's, n_2 b's, n_3 c's, . . ., n_k z's, we will get $M_n(n_1,\ldots,n_k) = \binom{n!}{n_1!n_2!n_3!\ldots\ldots n_k!}$ distinct words (where $n = n_1 + n_2 + \ldots + n_k$, which is the total word length). These numbers are called **multinomial coefficients** (see Multinomial Theorem in Appendix3 for $k = 3$: $(x + y + z)^n = \sum_{\substack{a+b+c=n \\ a,b,c>0}} M_n(a,b,c)x^a y^b z^c$ **).**

A particularly important case that arises frequently is when $k = 2$. This is **binomial coefficients**, and they have their own special notation: $M_n(r, n-r) = \frac{n!}{r!(n-r)!} = \binom{n}{r, n-r} = \binom{n}{r} = C_n^r$ (Binomial Theorem: $(a+b)^n = \sum_{r=0}^{n} C_n^r a^r b^{n-r}$, see Appendix 3). Binomial coefficients C_n^r (or $\binom{n}{r}$, we will use both notations) is also called the number of **combinations** (combinations of n things taken k at a time**)**. They give us the number of ways of choosing a set of r symbols (the order of objects is irrelevant) from a set of n distinct symbols without repetition.

For example, consider a problem of selecting a committee of three out of ten members of a club. We can treat the task the following way: dividing ten men arranged in a row into two groups (3 selected – denote them a's and 7 nonselected – denote them b's). Thus, the task is equivalent the following − how many permutations can we create by permuting aaabbbbbbb? The answer is $C_{10}^3 = \frac{10!}{3!7!} = \frac{10 \cdot 9 \cdot 8 \cdot 7!}{3!7!} = 120$. Note the symmetry of this formula ($C_{10}^3 = C_{10}^7$): selecting 3 people out of 10 can be done in the same number of ways as selecting 7 (and telling them: you did not make it).

But what if you select a chair, treasurer and secretary out of ten members of a club? Now the order of objects in the a's class is important − every of three men holds different posts. Thus, the task is how many permutations can we create by permuting $a_1a_2a_3$bbbbbbb? The answer is $3!C_{10}^3 = \frac{10!}{7!} = \frac{10 \cdot 9 \cdot 8 \cdot 7!}{7!} = 720$. In general, given n distinct symbols, the number of permutations (without repetition) of length $r \le n$ is $A_n^r = P_n^r = \frac{n!}{(n-r)!} = n(n-1)...(n-r+1)$ − the number of **arrangements** r objects over n places (permutations of n things taken r at a time).

We nave considered the cases: (Ordered Selection, No Repetition) − The number of arrangements (permutation), (Unordered Selection, No Repetition) − The number of combinations.

Consider the case (Ordered Selection, Allowing Repetition).

In how many possible ways can we build a five-letter word, using an alphabet of 26 letters? When the first letter is freely chosen in 26 ways, the next letter can be freely chosen in 26 ways again, and so on. This time we have a choice of 26 letters every time. So the answer is 26×26× . . . ×26 (5 times), namely, 26^5.

In general, we can make an ordered selection with repetition *r* objects from *n* distinct objects in n^r different ways.

The case (Unordered Selection, Allowing Repetition), for example, choosing ten pieces of fruit from shelf full of apples, pears, oranges and bananas, is the most difficult one.

We start with 10 empty boxes which we fill by fruit from the shelf (to contain one piece of fruit each), one by one, from left to right. To assure an unordered selection, we insist that apples go first, pears second, and so on. To determine how many boxes get an apple, we place a bar after the last "apple" box, similarly with pears, etc. For example: □□|□|□□□□□|□□ means getting 2 apples, 1 pear, 5 oranges and 2 bananas. Note that we can place the bars anywhere (with respect to the boxes), such as: □□□| |□□□□□□□| (we don't like pears and bananas). Also note that it will take exactly 3 = *n* − 1 bars to complete our 'shopping list'. Thus any permutation of 3 = *n* − 1 bars and 10 = *r* boxes corresponds to a particular selection (at the same time, a distinct permutation represents a distinct choice, so there is a one-to-one correspondence between these permutations and a complete list of fruit selections). We have already solved the problem of permutations (the answer is C_{13}^{3} = 286), so that is the number of options we have now. The general formula is obviously

$$C_{n+r-1}^{n-1}. \tag{2.1}$$

Knowing the general formulas for permutations (arrangements and combinations) is very important but not sufficient: the practice has a decisive role in framing a combinatorial problem correctly and in the proper application of the formulas. Your main task is to be able to correctly decide which of the formulas to use in each particular situation.

Poker game (from Bărboianu C. Understanding and calculations the odds). There are situations where a probability-based decision must be made − if wanted − in a relatively short time; these situations do not allow for thorough calculus even for a person with a mathematical background.

Assume you are playing a classical poker game with a 52-card deck. The cards have been dealt and you hold four suited cards (four cards with same symbol), but also a pair (two cards with same value).

For example, you hold 3♣ 5♣ 8♣ Q♣ Q♦. You must now discard and you ask yourself which combination of cards it is better to keep and which to replace.

To achieve a valuable formation, you will probably choose from the following two variants:

keep the four suited cards and replace one card (so that you have a flush);

keep the pair and replace three cards (so that you have "three of a kind or better").

In this gaming situation, many players intuit that, by keeping the pair (which is a high pair in the current example), the chances for a Q (queen) to be drawn or even for all three replaced cards to have same value, are bigger than the chance for one single drawn card to be ♣ (clubs).

And so, they choose the "safety" and play for "three of a kind or better". Other players may choose to play the flush, owing to the psychological impact of those four suited cards they hold.

In fact, the probability of getting a flush is about 19 % and the probability of getting three of a kind or better is about 6.3 %, which is three times lower.

In case you are aware of these figures beforehand, they may influence your decision and you may choose a specific gaming variant which you consider to have a better chance of winning.

But if you are in a similar gaming situation (you hold four suited cards and a pair) in a 24-card deck draw poker game, the order of those probabilities is reversed: the probability of getting a flush is about 10.5 % and the probability of getting three of a kind or better is almost 50 percent and may help you to determine to keep the pair.

This is a typical example of a decision based on probabilities in a relatively short time. It is obvious that, even assuming you have probability calculus skills, it is impossible to calculate all those figures in the middle of the game.

But you may use results memorized in anticipation obtained through your own calculations or picked from tables of guides containing collections of applied probabilities.

In games of chance, most players make probability-based decisions as part of their strategy, especially regular players.

In the game of poker the following hands are possible; they are listed in increasing order of desirability. In the definitions the word *value* refers to A, K, Q, J, 10, 9, 8, 7, 6, 5, 4, 3, or 2. This sequence also describes the relative ranks of the cards, with one exception: an Ace may be regarded as a 1 for the purposes of making a straight. (See the example in (d) below.)

(a) *one pair:* two cards of equal value plus three cards with different values

J♠ *J*♦ 9♥ *Q*♣ 3♠

(b) *two pair:* two pairs plus another card with a different value

J♠ *J*♦ 9♥ 9♣ 3♠

(c) *three of a kind:* three cards of the same value and two with different values

J♠ *J*♦ *J*♥ 9♣ 3♠

(d) *straight:* five cards with consecutive values

5♥ 4♠ 3♠ 2♥ *A*♣

(e) *flush:* five cards of the same suit

K♣ 9♣ 7♣ 6♣ 3♣

(f) *full house:* a three of a kind and a pair

J♠ *J*♦ *J*♥ 9♣ 9♠

(g) *four of a kind:* four cards of the same value plus another card

J♠ *J*♦ *J*♥ *J*♣ 9♠

(h) *straight flush:* five cards of the same suit with consecutive values

A♣ *K*♣ *Q*♣ *J*♣ 10♣

This example is called a *royal flush*.

To compute the probabilities of these poker hands we begin by observing that there are C_{52}^{5} = 52×51×50×49×48/(1×2×3×4×5) = 2,598,960 ways of picking 5 cards out of a deck of 52, so it suffices to compute the number of ways each hand can occur. We will do three cases to illustrate the main ideas and then leave the rest to the reader.

(d) *straight:* 10×4^5.

A straight must start with a card that is 5 or higher, 10 possibilities. Once the values are decided on, suits can be assigned in 4^5 ways. This counting regards a straight flush as a straight. If you want to exclude straight flushes, suits can be assigned in $4^5 - 4$ ways.

(f) *full house:* $13\times C_4^3\times 12\times C_4^2$.

We start with choosing the value for three of a kind, 13 possibilities, then assign suits to those three cards (C_4^3 ways), then pick the value for the pair (12 ways), then we assign suits to the last two cards (C_4^2 ways).

(a) *one pair:* $13\times C_4^2\times C_{12}^3\times 4^3$.

We first pick the value for the pair (13 ways), next pick the suits for the pair (C_4^2 ways), then pick three values for the other cards (C_{12}^3 ways) and assign suits to those cards (in 4^3 ways).

A common incorrect answer to this question is $13\times C_4^2\times 48\times 44\times 40$.

The faulty reasoning underlying this answer is that the third card must not have the same value as the cards in the pair (48 choices), the fourth must be different from the third and the pair (44 choices),.... However, this reasoning is flawed since it counts each outcome 3! = 6 times. (Note that $48\times44\times40/3! = C_4^2\times4^3$.)

The numerical values of the probabilities of all poker hands are given below:

(a) *one pair* .422569,
(b) *two pair* .047539,
(c) *three of a kind* .021128,
(d) *straight* .003940,
(e) *flush* .001981,
(f) *full house* .001441,
(g) *four of a kind* .000240,
(h) *straight flush* .000015.

The probability of getting none of these hands can be computed by summing the values for (a) through (g) (recall that (d) includes (h)) and subtracting the result from 1. However, it is much simpler to observe that we have nothing if we have five different values that do not make a straight or a flush. So the number of nothing hands is $(C_{13}^5 - 10)\times(4^5 - 4)$ and the probability of a nothing hand is 0.501177.

2.1. Exercises

When dealing with combinations of large numbers, it is useful not to unfold the combinatorial calculus until the end, to allow for factoring out and eventual reductions.

1. How many possible batting orders are there for nine baseball players?
2. A tire manufacturer wants to test four different types of tires on three different types of roads at five different speeds. How many tests are required?
3. A school gives awards in five subjects to a class of 30 students but no one is allowed to win more than one award. How many outcomes are possible?
4. A tourist wants to visit six of America's ten largest cities. In how many ways can she do this if the order of her visits is (a) important, (b) not important?
5. Five businessmen meet at a convention. How many handshakes are exchanged if each shakes hands with all the others?
6. In a class of 19 students, 7 will get *A*'s. In how many ways can this set of students be chosen?

7. How many license plates are possible if the first three places are occupied by letters and the last three by numbers? Assuming all combinations are equally likely, what is the probability the three letters and the three numbers are different?

8. How many four-letter "words" can you make if no letter is used twice and each word must contain at least one vowel (A, E, I, O or U)?

9. Assuming all phone numbers are equally likely, what is the probability that all the numbers in a seven-digit phone number are different?

10. A domino is an ordered pair (m, n) with $0 \le m \le n \le 6$. How many dominoes are in a set if there is only one of each?

11. A person has 12 friends and will invite 7 to a party.

(a) How many choices are possible if Al and Bob are feuding and will not both go to the party?

(b) How many choices are possible if Al and Betty insist that they both go or neither one goes?

12. A basketball team has 5 players over six feet tall and 6 who are under six feet. How many ways can they have their picture taken if the 5 taller players stand in a row behind the 6 shorter players who are sitting on a row of chairs?

13. Six students, three boys and three girls, lineup in a random order for a photograph. What is the probability that the boys and girls alternate?

14. Seven people sit at a round table. How many ways can this be done if Mr. Jones and Miss Smith

(a) must sit next to each other,

(b) must not sit next to each other?

(Two seating patterns that differ only by a rotation of the table are considered the same).

15. How many ways can four rooks be put on a chessboard so that no rook can capture any other rook?

The next tasks (16*–23*) have answers (see Answers and Solutions).

16.* A college team plays a series of 10 games which they can either win (W), lose (L) or tie (T).

(a) How many possible outcomes can the series have (differentiating between WL and LW, i.e. order is important).

(b) How many of these have exactly 5 wins, 4 losses and 1 tie?

(c) Same as (a) if we don't care about the order of wins, losses and ties?

17.* A student has to answer 20 true-false questions.

(a) In how many distinct ways can this be done?

(b) How many of these will have exactly 7 correct answers?

(c) At least 17 correct answers?

(d) Fewer than 3? (exclude 3).

18.* In how many ways can 3 Americans, 4 Frenchmen, 4 Danes and 2 Canadians be seated (here we are particular about nationalities, but not about individuals)

(a) in a row.

(b) In how many of these will people of the same nationality sit together?

(c) Repeat (a) with circular arrangement.

(d) Repeat (b) with circular arrangement.

19.* In how many ways can we put 12 books into 3 shelves?

Remark. This question is somehow ambiguous: do we want to treat the books as distinct or identical, and if we do treat them as distinct, do we care about the order in which they are placed within a shelf? The choice is ours, let's try it each way (the shelves are obviously distinct, and large enough to accommodate all 12 books if necessary).

20.* Twelve men can be seated in a row in 12! = 479001600 number of ways.

(a) How many of these will have Mr. A and Mr. B sit next to each other?

(b) How many of the original arrangements will have Mr. A and Mr. B sitting apart?

(c) How many of the original arrangements will have exactly 4 people sit between Mr. A and Mr. B?

21.* Consider the standard deck of 52 cards (4 suits: hearts, diamonds, spades and clubs, 13 "values": 2, 3, 4, . . ., 10, Jack, Queen, King, Ace). Deal 5 cards from this deck. This can be done in 2598960 distinct ways.

(a) How many of these will have exactly 3 diamonds?

(b) Exactly 2 aces?

(c) Exactly 2 aces and 2 diamonds?

22.* In how many ways can we deal 5 cards each to 4 players?

(b) So that each gets exactly one ace?

(c) None gets any ace.

(d) Mr. A gets 2 aces, the rest get none.

(e) (Any) one player gets 2 aces, the other players get none.

(f) Mr. A gets 2 aces.

(g) Mr. C gets 2 aces.

23.* Roll a die five times. The number of possible (ordered) outcomes is 7776. How many of these will have:

(a) One pair of identical values (and no other duplicates).

(b) Two pairs.

(c) A triplet.

(d) “Full house” (a triplet and a pair).
(e) “Four of a kind”.
(f) “Five of a kind”.
(g) Nothing.

3. Properties of Probability

Probability theory is nothing
but common sense reduced to calculation.
Laplace, Pierre Simon

The true logic of this world is the calculus of probabilities.
James Clerk Maxwell

From the definition we can derive many important and useful relationships, for example, for any sets A and B

(I) $P(A^c) = 1 - P(A)$ − the **complement rule**,

(II) $P(A \backslash B) = P(A) - P(A \cap B)$,

(III) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$) − the **sum rule**.

Formula (I) follows from $1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$ (we use the additivity of probability). For an example consider A = "at least one six" in the experiment of rolling two dice. In this case A^c = "no six". There are 5×5 outcomes with "no six" so $P(A^c) = 25/36$ and $P(A) = 1 - 25/36 = 11/36$ as we computed before. Now, setting $A = \Omega$, establishes $P(\phi) = 0$.

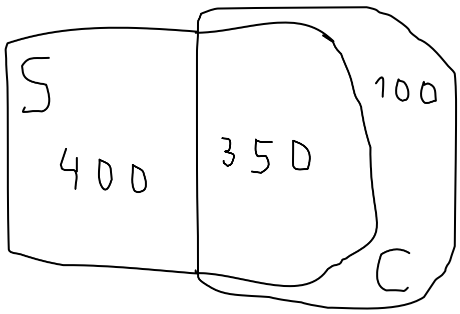
Finally, using additivity repeatedly, we obtain $P(B) = P(B \cap A) + P(B \cap A^c)$, and $P(A \cup B) = P(A \cup (B \cap A^c)) = P(A) + P(B \cap A^c) = P(A) + P(B) - P(B \cap A)$, which proves (3). (Intuitively, $P(A) + P(B)$ counts $A \cap B$ twice so we have to subtract $P(A \cap B)$ to make the net number of times $A \cap B$ is counted equal to 1.) To illustrate this rule let A = "red die shows six", B = "green die shows six" in the experiment of rolling two dice. In this case $A \cup B$ = "at least one 6", and $A \cap B = \{(6, 6)\}$ so we have $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 1/6 + 1/6 - 1/36 = 11/36$.

The same principle applies to counting outcomes in sets. (See formula (A3.1) in Appendix 3)

Example. A survey of 1000 students revealed that 750 owned iPads, 450 owned cars, and 350 owned both. How many own either a car or a stereo?

Letting $|S|$ denote the number of students with iPads, and $|C|$ the number with cars, the reasoning that led to (III) tells us that $|S \cup C| = |S| + |C| - |S \cap C| = 750 + 450 - 350 = 850$.

We can confirm this by drawing a picture:



We can extend (III) to: $P(A\cup B\cup C)$ = $P(A)$ + $P(B\cup C)$ − $P(A\cap(B\cup C))$ = $P(A)$ + $P(B)$ + $P(C)$ − $P(B\cap C)$ − $P((A\cap B)\cup(A\cap C))$ = $P(A)$ + $P(B)$ + $P(C)$ − $P(A\cap B)$ − $P(A\cap C)$ − $P(B\cap C)$ + $P(A\cap B\cap C)$.

And, by induction, we can get the fully general

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k) - \sum_{k\neq j}^{n} P(A_k \cap A_j) + \sum_{k\neq j\neq i}^{n} P(A_k \cap A_j \cap A_i) - \ldots \pm P(A_1 \cap A_2 \cap \ldots \cap A_n)$$

(the plus sign for n odd, the minus sign for n even). The formula computes the probability that at least one of the A_i events happens.

You should now prove (II) as an exercise. From (II) it follows that if $B\subset A$ then $P(B)\leq P(A)$ (**monotonicity of probability**).

Note, that probability of any (Boolean) expression involving events *A, B, C, ...* can be always converted into probabilities involving the individual events and their simple (non-complemented) intersections ($A\cap B$, $A\cap B\cap C$, etc.) only. For example,

1) $P\{(A\cap B)\cup(B\cap C)^c\}$ = $P(A\cap B)$ + $P\{(B\cap C)^c\}$ − $P\{A\cap B\cap(B\cap C)^c\}$ = $P(A\cap B)$ +1− $P(B\cap C)$ − $P(A\cap B)$ + $P(A\cap B\cap B\cap C)$ =1− $P(B\cap C)$ + $P(A\cap B\cap C)$. This can be also deduced from the corresponding Venn diagram, bypassing the algebra.

2) $P\{(A\cap B)\cup(C\cup D)^c\}$ = $P(A\cap B)$ + $P\{(C\cup D)^c\}$ − $P\{(A\cap B)\cap(C\cup D)^c\}$ = $P(A\cap B)$ +1 − $P(C\cup D)$ − $P(A\cap B)$ + $P\{(A\cap B)\cap(C\cup D)\}$ =1 − $P(C\cup D)$ + $P\{(A\cap B\cap C)\cup(A\cap B\cup D)\}$ =1− $P(C)$ − $P(D)$ + $P(C\cap D)$ + $P(A\cap B\cap C)$ + $P(A\cap B\cap D)$ − $P(A\cap B\cap C\cap D)$.

Note that intersections are usually easy to deal with, unions are hard but can be converted to intersections as shown above.

3.1. Examples

1. Four players are dealt 5 cards each. What is the probability that at least one player gets exactly 2 aces.

Solution: let A_1 be the event that the first player gets exactly 2 aces, A_2 means that the second player has exactly 2 aces, etc. The question amounts to finding $P(A_1 \cup A_2 \cup A_3 \cup A_4)$. By our formula, this equals $P\left(\bigcup_{k=1}^{4} A_k\right) = \sum_{k=1}^{4} P(A_k) - \sum_{k \neq j}^{4} P\left(A_k \cap A_j\right) + 0$ (the intersection of 3 or more of these events is empty − there are only 4 aces). For $P(A_1)$ we get $\frac{C_4^2 C_{48}^3}{C_{52}^5}$ = 3.993 % (the denominator counts the total number of five-card hands, the numerator counts only those with exactly two aces) with the same answer for $P(A_2)$, $P(A_3)$, $P(A_4)$ (the four players must have equal chances). Similarly $P(A_1 \cap A_2) = \binom{4}{2,2,0}\binom{48}{3,3,42} / \binom{52}{5,5,42}$ = 0.037 % (the denominator represents the number of ways of dealing 5 cards each to two players, the numerator counts only those with 2 aces each), and the same probability for any other pair of players.

Final answer: $4P(A_1) - 6P(A_1 \cap A_2)$ = 15.75 %.

2. There are 100,000 lottery tickets marked 00000 to 99999. One of these is selected at random. What is the probability that the number on it contains 84 (consecutive, in that order) at least once.

Solution: let's introduce four events: A means that the first two digits of the ticket are 84 (regardless of what follows), B: 84 is found in the second and third position, C: 84 in position three and four, and D: 84 in the last two positions. Obviously we need $P(A \cup B \cup C \cup D) = P(A) + P(B) + P(C) + P(D) - P(A \cap C) - P(A \cap D) - P(B \cap D) + 0$ (the remaining possibilities are all impossible events − the corresponding conditions are incompatible).

The answer is 4×1000/100,000 − 3×10/100,000 = 0.04 − 0.0003 = 3.97 % (the logic of each fraction should be obvious − there are 1000 tickets which belong to A, 10 tickets which meet conditions A and C, etc.).

3. Suppose that k distinct letters (to different friends) have been written, each with a corresponding (uniquely addressed) envelope. Then, for some strange reason, the letters are placed in the envelopes purely randomly (after a thorough shuffling).

(a) What is the probability of all letters being placed correctly?

(b) What is the probability that none of the k letters are placed correctly?

(c) What is the probability of exactly one letter being placed correctly?

The sample space of this experiment is thus a list of all permutations of k objects (123, 132, 213, 231, 312, 321), when $k = 3$ (we will assume that 123 represents the correct placement of all three letters). In general, there are $k!$ of these, all of them equally likely (due to symmetry, i.e. none of these arrangements should be more likely than any other).

(a) *Solution* (fairly trivial): only one out of $k!$ random arrangements meets the criterion, thus the answer is $1/k!$ (very small for k beyond 10).

(b) *Solution* is this time a lot more difficult. First we have to realize that it is relatively easy to figure out the probability of any given letter being placed correctly, and also the probability of any combination (intersection) of these, i.e. two specific letters correctly placed, three letters correct . . ., etc.

We use the following notation: A_1 means that the first letter is placed correctly (regardless of what happens to the rest of them), A_2 means the second letter is placed correctly, etc.

Probability $P(A_1)$ is computed by counting the number of permutations which have 1 in the correct first position, and dividing this by $k!$. The number of permutations which have 1 fixed is obviously $(k-1)!$ (we are permuting 2, 3, ... ,k, altogether $k-1$ objects). $P(A_1)$ is thus equal to $(k-1)!/k!$.

The probability of A_2, A_3, etc. can be computed similarly, but it should be clear from the symmetry of the experiment that all these probabilities must be the same, and equal to $P(A_1) = 1/k$ (why should any letter have a better chance of being placed correctly than any other?).

Similarly, let us compute $P(A_1 \cap A_2)$, i.e. probability of the first and second letter being placed correctly (regardless of the rest). By again counting the corresponding number of permutations (with 1 and 2 fixed), we arrive at $(k-2)!/k! = \frac{1}{k(k-1)}$.

This must be the same for any other pair of letters, e.g. $P(A_3 \cap A_7) = \frac{1}{k(k-1)}$, etc.

In this manner we also get $P(A_1 \cap A_2 \cap A_3) = P(A_3 \cap A_7 \cap A_{11}) = \frac{1}{k(k-1)(k-2)}$, etc.

So now we know how to deal with any intersection. All we need to do is to express the event “all letters misplaced” using intersections only, and evaluate the answer, thus: $P(\overline{A_1} \cap \overline{A_2} \cap \ldots \cap \overline{A_k})$ (all letters misplaced) = (De Morgan’s law)$P(\overline{A_1 \cup A_2 \cup \ldots \cup A_k}) = 1 - P(A_1 \cup A_2 \cup \ldots \cup A_k) = 1 - \sum_{i=1}^{k} P(A_i)$ $+ \sum_{i<j}^{k} \mathrm{P}(A_i \cap A_j) + \ldots + (-1)^k P(A_1 \cap A_2 \cap \ldots \cap A_k) = 1 - k\frac{1}{k} + \binom{k}{2}\frac{1}{k(k-1)}$

$$-\binom{k}{3}\frac{1}{k(k-1)(k-2)}+\ .\ .\ .\ +(-1)^k\frac{1}{k!}=\ 1-1+\frac{1}{2!}-\frac{1}{3!}+...+(-1)^k\frac{1}{k!}.$$

For k = 3 this implies 1 − 1 + 1/2−1/6 = 1/3 (check, only 231 and 312 out of six permutations). For k = 1, 2, 4, 5, 6, and 7 we get: 0 (one letter cannot be misplaced), 50 % for two letters (check), 37.5 % (four letters), 36.67 % (five), 36.81 % (six), 36.79 % (seven), after which the probabilities do not change practically (i.e., surprisingly, we get practically the same answer for 100 letters, a million letters, etc.).

Can we identify the limit of the 1 −1+ 1/2! − 1/3! + . . . sequence? Yes, of course, this is the expansion of $e^{-1}\approx 0.36788$ (See Appendix 3).

(c) Similarly, the probability of exactly one letter being placed correctly is $1-1+\frac{1}{2!}-\frac{1}{3!}+\ ...\mp\frac{1}{(k-1)!}$ (the previous answer short of its last term!). This equals to 1, 0, 50 %, 37.5 %, . . . for k = 1, 2, 3, 4, . . . respectively, and has the same limit.

3.2. Exercises

1. An experiment has 5 possible outcomes (simple events) with the following probabilities:

Simple Event	Probability
S_1	.15
S_2	.20
S_3	.20
S_4	.25
S_5	.20

a) Find the probability of each of the following events:

A: Outcome S_1, S_2, or S_4 occurs.

B: Outcome S_2, S_3, or S_5 occurs.

C: Outcome S_4 does not occur.

b) List the simple events in the complements of events A, B and C.

c) Find the probabilities of A^c, $B^c \cap C^c$.

2. In the freshman class, 62 % of the students take math, 49 % take science, and 38 % take both science and math. What percentage takes at least one science or math course?

3. 24 % of people have American Express Cards, 61 % have VISA cards and 8 % have both. What percentage of people have at least one abovementioned credit card?

4. Tversky and Kahneman asked a group of subjects to carry out the following task. They are told that: Linda is 31, single, outspoken, and very

bright. She majored in philosophy in college. As a student, she was deeply concerned with racial discrimination and other social issues, and participated in anti-nuclear demonstrations.

The subjects are then asked to rank the likelihood of various alternatives, such as:

(1) Linda is active in the feminist movement.

(2) Linda is a bank teller.

(3) Linda is a bank teller and active in the feminist movement.

Tversky and Kahneman found that between 85 and 90 percent of the subjects rated alternative (1) most likely, but alternative (3) more likely than alternative (2). Is it? They call this phenomenon the *conjunction fallacy*, and note that it appears to be unaffected by prior training in probability or statistics. Is this phenomenon a fallacy? If so, why?

5. Suppose $\Omega = \{a, b, c\}$, $P(\{a, b\}) = 0.7$, and $P(\{b, c\}) = 0.6$. Compute the probabilities of $\{a\}$, $\{b\}$, and $\{c\}$.

6. Suppose A and B are disjoint with $P(A) = 0.3$ and $P(B) = 0.5$. What is $P(A^c \cap B^c)$?

7. Given two events A and B with $P(A) = 0.4$ and $P(B) = 0.7$. What are the maximum and minimum possible values for $P(A \cap B)$?

8. Two cards are drawn successively from a deck of 52 cards. Find the probability that the second card is higher in rank than the first card.

Hint: Show that 1 = P(higher)+P(lower)+P(same) and use the fact that P(higher) = P(lower).

9.* Within the next hour 4 people in a certain town will call for a cab. They will choose, randomly, out of 3 existing (equally popular) taxi companies. What is the probability that no company is left out (each gets at least one job)?

10.* Consider a 10 floor government building with all floors being equally likely to be visited. If six people enter the elevator (individually, i.e. independently) what is the probability that they are all going to (six) different floors?

11.* Jim, Joe, Tom and six other boys are randomly seated in a row. What is the probability that at least two of the three friends will sit next to each other?

12.* There are 10 people at a party (no twins). Assuming that all 365 days of a year are equally likely to be someone's birth date (not quite, say the statistics, but we will ignore that) and also ignoring leap years, what is the probability of:

(a) All these ten people having different birth dates?

(b) Exactly two people having the same birth date (and no other duplication).

13.* (Extension of the exercise 10*). What if the floors are not equally likely (they issue licenses on the 4th floor, which has therefore a higher probability of 1/2 to be visited by a "random" arrival − the other floors remain equally likely (with the probability of 1/16 each).

14.* (Quality control). A shipment of 50 precision parts including 4 that are defective is sent to an assembly plant. The quality control division selects 10 at random for testing and rejects the entire shipment if 1 or more are found defective. What is the probability this shipment passes inspection?

15. Probability of which event is higher: that tomorrow's temperature at noon will not exceed 19°C or that it will not exceed 25°C?

4. Conditional Probability, Independence

Patient: Will I survive this risky operation?
Surgeon: Yes, I'm absolutely sure that you will survive the operation.
Patient: How can you be so sure?
Surgeon: Well, 9 out of 10 patients die in this operation,
and yesterday my ninth patient died.

Mathematical knowledge adds vigor to the minds,
frees it from prejudice, credulity, and superstition.
John Arbuthnot

Suppose you have a well-shuffled conventional pack of cards. Obviously (by symmetry), the probability $P(T)$ of the event T that the top card is an ace is $P(T) = 4/52 = 1/13$.

However, suppose you notice that the bottom card is the ace of spades (event S_A). What now is the probability that the top card is an ace? There are 51 possibilities and three of them are aces, so by symmetry again the required probability is 3/51. To distinguish this from the original probability, we denote it by $P(T/S_A)$ and call it the conditional probability of T given S_A.

Similarly, had you observed that the bottom card was the king of spades S_K, you would conclude that the probability that the top card is an ace is $P(T/S_K) = 4/51$.

Here is a less trivial example.

Example (*Poker*). Suppose you are playing poker. As the hand is dealt, you calculate the chance of being dealt a royal flush R, assuming that all hands of five cards are equally likely. (A *royal flush* comprises 10, J, Q, K, A in a single suit.) The probability $P(R) = 4/C_{52}^{5} = 1/649{,}740$. The dealer deals your last card face up. It is the ace of spades, S_A. If you accept the card, what now is your chance of picking up a royal flush?

Intuitively, it seems unlikely still to be $P(R)$ above, as the conditions for getting one have changed. Now you need your first four cards to be the ten to king of spades precisely. (Also, had your last card been the two of spades, S_2, your chance of a royal flush would definitely be zero.) As above, to distinguish this new probability, we call it the conditional probability of R given S_A and denote it by $P(R/S_A)$.

Is it larger or smaller than $P(R)$? At least you do have an ace, which is a start, so it might be greater. But you cannot now get a flush in any suit but spades, so it might be smaller. To resolve the uncertainty, you assume that any set of four cards from the remaining 51 cards is equally likely to complete your hand and calculate that $P(R/S_A) = 1/C_{51}^4 = \frac{13}{5}$ $P(R)$.

Your chances of a royal flush have more than doubled.

Let us investigate these ideas in a more general setting. As usual we are given a sample space, and a probability function $P(\ )$. We suppose that some event B occurs, and denote the conditional probability of any event A, given B, by $P(A/B)$. As we did for $P(\ )$, we observe that $P(\ |B)$ is a function of events, which takes values in [0, 1]. But what function is it?

Clearly, $P(A)$ and $P(A/B)$ are not equal in general, because even when $P(B^c) \neq 0$ we always have $P(B^c/B) = 0$.

Second, we note that given the occurrence of B, the event A can occur if and only if $A \cap B$ occurs. This makes it natural to require that $P(A/B)$ is proportional to $P(A \cap B)$. Finally, it is obvious that $P(B/B) = 1$.

After a thought about these three observations, it appears that a candidate to play the role of $P(A/B)$ is $P(A \cap B)/P(B)$. Note that this follows also from classical definition of probability. Let the event B occur. This is as if the whole sample space has shrunk to B only, and we can consider B as the new "reduced" sample space. Let $|B| = m$, and $|AB| = r$, than $P(A/B) = \frac{r}{m} = \frac{r/n}{m/n} = \frac{P(A \cap B)}{P(B)}$. Not surprisingly, all formulas which hold true in the original sample space are still valid in the new sample space B, i.e. conditionally, e.g.: $P(A/B) = 1 - P(A^c/B)$, $P(A \cup C/B) = P(A/B) + P(C/B) - P(A \cap C/B)$, etc. (make all probabilities of any old formula conditional on B).

In formal (constructed from axioms) theory of probability conditional probability introduced by definition:

Let A and B be events with $P(B) > 0$. *Given that B occurs, the* ***conditional probability*** *that A occurs is denoted by* $P(A/B)$ *and defined by*

$$P(A/B) = P(A \cap B)/P(B). \tag{4.1}$$

When $P(B) = 0$, the conditional probability $P(A/B)$ is not defined by (4.1). However, it is convenient to adopt the convention that, even when $P(B) = 0$, we may still write $P(A \cap B) = P(A/B)\ P(B)$, both sides having the value zero. Thus, whether $P(B) > 0$ or not, it is true that $P(A \cap B) = P(A/B)\ P(B)$.

Try to see the probabilistic meaning of this equation. Denote M the event that you obtain a passing grade in Math and by S the event that you obtain a passing grade in Stats. Suppose that Math is a prerequisite for Stats. Then the equation $P(S \cap M) = P(S/M)P(M)$ basically tells us that to obtain a passing grade in both Math and Stats you have to pass Math first and, with that prerequisite satisfied, to obtain a passing grade in Stats. The fact that you have not failed Math tells something about your abilities and increases the chance of passing Stats.

The rule which was used to compute the probability of the intersection is called the **product rule** and it can be generalized to any three, four etc. events: $P(A \cap B) = P(A/B)\,P(B)$, $P(A \cap B \cap C) = P(A)P(B/A)P(C/A \cap B)$, $P(A \cap B \cap C \cap D) = P(A)P(B/A)P(C/A \cap B)P(D/A \cap B \cap C)$.

Example. (*Coincidence*). If you randomly choose twenty four persons, what do you think of the probability of two or more of them having the same birthday (this means the same month and the same day of the year)?

Solution. A simple method of calculus to use here is the step by step one (in fact, we use the product rule): the probability for the birthday of two arbitrary persons not to be the same is 364/365 (because we have one single chance from 365 for the birthday of the first person to match the birthday of the second). The probability for the birthday of a third person to be different from those of the other two is 363/365; for the birthday of a fourth person is 362/365, and so on, until we get to the last person, the 24th, with a 342/365 probability.

We have obtained twenty three fractions, which all must be multiplied to get the probability of all twenty four birthdays to be different. The product is a fraction that remains as 23/50 after reduction.

The probability we are looking for is the probability of the contrary event, and this is $1 - 23/50 = 27/50$.

This calculus does not take February 29 into account, or that birthdays (according to statistics) have a tendency to concentrate higher in certain months rather than in others. The first circumstance diminishes the probability, while the second increases it.

If you bet on the coincidence of birthdays of twenty four persons, on average you would loose twenty three and win twenty seven of each fifty bets over time.

Of course, the more persons considered, the higher the probability. With over sixty persons, probability gets very close to certitude. For 100 persons, the chance of a bet on a coincidence is about 3,000,000 : 1. Obviously, absolute certitude can be achieved only with 366 persons or more.

Intuitively, two events A and B are independent if the occurrence of A has no influence on the probability of occurrence of B and vice versa, i.e. $P(A/B) = P(A)$ and $P(B/A) = P(B)$. From this and the product rule it follows for independent events A and B that $P(A \cap B) = P(A)P(B)$ (and vice versa: $P(A \cap B) = P(A)P(B) \Rightarrow P(A/B) = P(A)$ and $P(B/A) = P(B)$). Thus, $P(A \cap B) = P(A)P(B) \Leftrightarrow P(A/B) = P(A)$ and $P(B/A) = P(B)$.

The formal definition is: A and B are ***independent*** if $P(A \cap B) = P(A)P(B)$.

We now give three classic examples of independent events. In each case it should be clear that the events are intuitively independent, and we will check the formal definition occurring.

1) Flip two coins. A = "The first coin shows Head", B = "The second coin shows Head". $P(A) = 1/2$, $P(B) = 1/2$, $P(A \cap B) = 1/4$.

2) Roll two dice. A = "The first die shows 5", B = "The second die shows 2". $P(A) = 1/6$, $P(B) = 1/6$, $P(A \cap B) = 1/36$.

3) Pick a card from a deck of 52. A = "The card is an ace", B = "The card is a spade", $P(A) = 1/13$, $P(B) = 1/4$, $P(A \cap B) = 1/52$.

Two examples of events that are not independent are

1) Draw two cards from a deck. A = "The first card is a spade", B = "The second card is a spade". $P(A) = 1/4$, $P(B) = 1/4$, but $P(A \cap B) = C_{13}^2 / C_{52}^2$ $= \frac{13 \cdot 12}{52 \cdot 51} < \left(\frac{1}{4}\right)^2$. Please, explain why $P(B) = 1/4$.

Intuitively, these two events are not independent, since getting a spade the first time reduces the fraction of spades in the deck and makes it harder to get a spade the second time.

2) Roll two dice. A = "The sum of the two dice is 9", B = "The first die is 2". $A = \{(6,3), (5,4), (4,5), (3,6)\}$, so $P(A) = 4/36$. $P(B) = 1/6$, but $P(A \cap B) = 0$ since (2,7) is impossible.

In general if A and B are disjoint events that have positive probability, they are not independent since $P(A)P(B) > 0 = P(A \cap B)$.

The events $A_1, \ldots, A_n$ are said to be **independent** if for any $1 \le i_1 < i_2 \ldots < i_k \le n$ we have $P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \times \ldots \times P(A_{i_k})$.

Likewise (4.1), $P(A \cap B^c) = P(A|B^c)P(B^c)$ and hence, for any events A and B, we have proved the following **partition rule**:

$$P(A) = P(A \cap B) + P(A \cap B^c) = P(A/B)P(B) + P(A/B^c)P(B^c).$$

We often have occasion to use the following elementary generalization of the rule (the extended partition rule or **formula of total probability**):

$$P(A) = \sum_{i=1}^{n} P(A/B_i)P(B_i) \qquad (4.2)$$

whenever $A \subset \bigcup_{i=1}^{n} B_i$ and $B_i \cap B_j = \phi$ for $i \neq j$.

Example. Two players are dealt 5 cards each. What is the probability that they will have the same number of aces?

Solution. We partition the sample space according to how many aces the first player gets, calling the events $A_0, A_1, \ldots, A_4$. Let B be the event of our question (both players having the same number of aces). Then, by the formula of total probability: $P(B) = P(A_0)P(B/A_0) + P(A_1)P(B/A_1) + P(A_2)P(B/A_2) + P(A_3)P(B/A_3) + P(A_4)P(B/A_4) =$

$$\binom{4}{0}\binom{48}{5}\Big/\binom{52}{5}\times\binom{4}{0}\binom{43}{5}\Big/\binom{47}{5}+\binom{4}{1}\binom{48}{4}\Big/\binom{52}{5}\times\binom{3}{1}\binom{44}{4}\Big/\binom{47}{5}+$$

$$\binom{4}{2}\binom{48}{3}\Big/\binom{52}{5}\times\binom{2}{2}\binom{45}{3}\Big/\binom{47}{5}+0=0.4933.$$

We may write $P(B_j/A) = P(B_j \cap A)/P(A) = P(A/B_j)P(B_j)/P(A)$, and expanding the denominator using (4.2), we have proved the following celebrated result; also known as Bayes's Rule (or Theorem):

Bayes's Theorem. If $A \subset \bigcup_{i=1}^{n} B_i$ and $B_i \cap B_j = \phi$ *for* $i \neq j$, then

$$P(B_j/A) = P(A/B_j)P(B_j)/P(A) = P(A/B_j)P(B_j)/\sum_{i=1}^{n} P(A/B_i)P(B_i).$$

Interpretation Bayes's Rule in terms of subjective probabilities. Before occurrence of the event A, an expert forms an opinion about likelihood of the event B. That opinion, embodied in $P(B)$, is called a *prior probability*. After occurrence of A, the expert's belief is updated to obtain $P(B/A)$, called a *posterior probability*. By the Bayes's Theorem, updating is accomplished through multiplication of the prior probability by the factor $P(A/B)/P(A)$.

The following is a typical example of the rule application in practice.

Example. You have a blood test for some rare disease that occurs by chance in 1 in every 100,000 people. The test is fairly reliable; if you have the disease, it will correctly say so with probability 0.95; if you do not have the disease, the test will wrongly say you do with probability 0.005. If the test says you do have the disease, what is the probability that this is a correct diagnosis?

Solution. Let D be the event that you have the disease and T the event that the test says you do. Then, we require $P(D/T)$, which is given by $P(D/T) = P(T/D)P(D)/\{P(T/D)P(D) + P(T/D^c)P(D^c)\} =$
$(0.95)(0.00001)/\{(0.95)(0.00001) + (0.99999)(0.005)\} = 0.002$.

Despite appearing to be a pretty good test, for a disease as rare as this the test is almost useless.

It will be useful and, hope, interesting to read following articles concerning the topic http://plus.maths.org/issue9/news/banks/index.html and http://plus.maths.org/issue21/features/clark/index.html.

4.1. Exercises

1. Prove that if A and B are independent, then their complements are also independent.
2. You are going to meet a friend at the airport. Your experience tells you that the plane is late 70 % of the time when it rains, but is late only 20 % of the time when it does not rain. The weather forecast that morning calls for a 40 % chance of rain. What is the probability the plane will be late?
3. A student is taking a multiple-choice test in which each question has four possible answers. She knows the answers to 50 % of the questions, can narrow the choices down to two 30 % of the time, and does not know anything about 20 % of the questions. What is the probability she will correctly answer a question chosen at random from the test?
4. Two hunters shoot at a deer, which is hit by exactly one bullet. If the first hunter hits his targets with probability 0.3 and the second with probability 0.6, what is the probability the second hunter killed the deer?
5. 5 % of men and 0.25 % of women are colorblind. What is the probability that a colorblind person is a man?
6. A reader of Marilyn vos Savant's column wrote in with the following question:

"My dad heard this story on the radio. At Duke University, two students had received A's in chemistry all semester. But on the night before the final exam, they were partying in another state and didn't get back to Duke until it was over. Their excuse to the professor was that they had a flat tire, and they asked if they could take a make-up test. The professor agreed, wrote out a test and sent the two to separate rooms to take it. The first question (on one side of the paper) was worth 5 points, and they answered it easily. Then they flipped the paper over and found the second question, worth 95 points: 'Which tire was it?' What was the probability that both students would say the same thing? My dad and I think it's 1 in 16. Is that right?"

(a) Is the answer 1/16?

(b) The following question was asked of a class of students. "I was driving to school today, and one of my tires went flat. Which tire do you think it was?" The responses were as follows: right front, 58 %, left front, 11 %, right rear, 18 %, left rear, 13 %. Suppose that this distribution holds in the general population, and assume that the two test-takers are randomly chosen from the general population. What is the probability that they will give the same answer to the question?

7. Suppose the experiment consists of rolling two dice (red and green), the event A is: "the total number of dots equals 6", B is: "the red die shows an even number". Compute $P(B/A)$.

8. Consider the following data on traffic accidents:

age group	% of drivers	accident probability
16 to 25	15	0.10
26 to 45	35	0.04
46 to 65	35	0.06
over 65	15	0.08

Calculate

(a) the probability a randomly chosen driver will have an accident this year,

(b) the probability a driver is between 46 and 65 given that she had an accident.

9.* Let $P(A) = 0.1$, $P(B) = 0.2$, $P(C) = 0.3$ and $P(D) = 0.4$; A, B, C, D – independent events. Compute $P\{(A \cup B) \cap \overline{C \cup D}\}$.

10.* Out of 10 dice, 9 of which are regular but one is "crooked" (6 has a probability of 0.5), a die is selected at random (we cannot tell which one, they all look identical). Then, we roll it twice. We will answer three questions:

(a) Given that the first roll resulted in a six (event S_1), what is the (conditional) probability of getting a six again in the second roll (event S_2)?

(b) Are S_1 and S_2 independent?

(c) Given that both rolls resulted in a six, what is the (conditional) probability of having selected the crooked die?

11.* Let us return to the second example in Examples 3 (lottery with 100,000 tickets) and compute the probability that a randomly selected ticket has an 8 and a 4 on it (each at least once, in any order, and not necessarily consecutive).

12.* The same question, but this time we want at least one 8 followed (sooner or later) by a 4 (at least once). What makes this different from the original question is that 8 and 4 now don't have to be consecutive.

13.* Ten people have been arrested as suspects in a crime one of them must have committed. A lie detector will (incorrectly) incriminate an innocent person with a 5 % probability, it can (correctly) detect a guilty person with a 90 % probability.

(a) One person has been tested so far and the lie detector has its red light flashing (implying: "that's him"). What is the probability that he is the criminal?

(b) All 10 people have been tested and exactly one incriminated. What is the probability of having the criminal now?

14.* Two men take one shot each at a target. Mr. A can hit it with the probability of 1/4, Mr. B's chances are 2/5 (he is a better shot). What is the probability that the target is hit (at least once)? Here, we have to (on our own) assume independence of the two shots.

15.* A, B, C are mutually independent, having (the individual) probabilities of 0.25, 0.35 and 0.45, respectively. Compute $P\{(A \cap \overline{B}) \cup C\}$.

16.* Two coins are flipped, followed by rolling a die as many times as the number of heads shown.

(a) What is the probability of getting fewer than 5 dots in total?

(b) Given that there were exactly 3 dots in total, what is the conditional probability that the coins showed exactly one head?

17.* What is more likely, getting at least one 6 in four rolls of a die, or getting at least one double 6 in twenty four rolls of a pair of dice?

18.* Four people are dealt 13 cards each. You (one of the players) got one ace. What is the probability that your partner has the other three aces?

19.* A simple padlock is made with only ten distinct keys (all equally likely). A thief steals, independently, 5 of such keys, and tries these to open your lock. What is the probability that he will succeed?

20. A man has five coins in his pocket. Two are double-headed, one is double-tailed, and two are normal. They can be distinguished only by looking at them.

(a) The man shuts his eyes, chooses a coin at random, and tosses it. What is the probability that the lower face of the coin is a head?

(b) He opens his eyes and sees that the upper face is a head. What is the probability that the lower face is a head?

21. Binary digits, i.e., 0's and 1's, are sent down a noisy communications channel. They are received as sent with probability 0.9 but errors occur with probability 0.1. Assuming that 0's and 1's are equally likely, what is the probability that a 1 was sent given that we received a 1?

22. To improve the reliability of the channel described in the last exercise, we repeat each digit in the message three times. What is the probability that 111 was sent given that

(a) we received 101?

(b) we received 000?

23. A cab was involved in a hit and run accident at night. Two cab companies green and blue operate 85 % and 15 % of the cabs in the city respectively. A witness identified the cab as blue. However, in a test 80 % of witnesses were able to correctly identify the cab color. Given this what is the probability that the cab involved in the accident was blue?

24. A student goes to class on a snowy day with probability 0.4, but on a no snowy day attends with probability 0.7. Suppose that 20 % of the days in February are snowy. What is the probability it snowed on February 7th given that the student was in class on that day?

25. You are a serious student who studies on Friday nights but your roommate goes out and has a good time. 40 % of the time he goes out with his girlfriend; 60 % of the time he goes to a bar. 30 % of the times when he goes out with his girlfriend he spends the night at her apartment. 40 % of the times when he goes to a bar he gets in a fight and gets thrown in jail. You wake up on Saturday morning and your roommate is not home. What is the probability he is in jail?

26. Two masked robbers try to rob a crowded bank during the lunch hour but the teller presses a button that sets off an alarm and locks the front door. The robbers, realizing they are trapped, throw away their masks and disappear into the chaotic crowd. Confronted with 40 people claiming they are innocent, the police give everyone a lie detector test. Suppose that guilty people are detected with probability 0.95, and innocent people appear to be guilty with probability 0.01. What is the probability Mr. Jones is guilty given that the lie detector says he is?

27. At a slot machine with three reels and seven symbols, find the probability of getting the same symbol on all reels in one spin.

28. Two shooters are shooting two shots each at a target. The probability of the event "the first shooter hits the target" is 0.7 for each shot and the probability for the second shooter is 0.8. What is the probability of both shooters missing the target?

29.* Probability theory was used in a famous court case: *People vs. Collins*. In this case a purse was snatched from an elderly person in a Los Angeles suburb.

A couple seen running from the scene were described as a black man with a beard and a mustache and a blond girl with hair in a ponytail. Witnesses said they drove off in a partly yellow car. Malcolm and Janet Collins were arrested.

He was black and though clean shaven when arrested had evidence of recently having had a beard and a mustache. She was blond and usually wore her hair in a ponytail. They drove a partly yellow Lincoln. The prosecution called a professor of mathematics as a witness who suggested that a conservative set of probabilities for the characteristics noted by the witnesses would be as shown below

man with mustache	1/4,
girl with blond hair	1/3,
girl with ponytail	1/10,
black man with beard	1/10,
interracial couple in a car	1/1000,
partly yellow car	1/10.

The prosecution then argued that the probability that all of these characteristics are met by a randomly chosen couple is the product of the probabilities or 1/12,000,000, which is very small. He claimed this was proof beyond a reasonable doubt that the defendants were guilty. The jury agreed and handed down a verdict of guilty of second-degree robbery.

If you were the lawyer for the Collins couple how would you have countered the above argument?

30. A person *A* receives information and transmits it to another person *B*. Person *B* transmits it to a third person *C*, who transmits it to a fourth person *D*. Knowing that each person tells the truth in one case out of three, find the probability that *A* told the truth, given that *D* told the truth.

5. The Monty Hall problem and other puzzles

The mind has its illusions as the sense of sight;
and in the same manner that the sense of feeling
corrects the latter, reflection and calculation correct the former.
Pierre-Simon Laplace

Math is more than computations and numbers;
it's about being able to solve problems.

Conditional probabilities are the sources of many “paradoxes” in probability. One must be very careful in dealing with problems involving conditional probability.

The Monty Hall problem. We consider now a problem called the Monty Hall problem. This has long been a favorite problem but was revived by a letter from Craig Whitaker to Marilyn vos Savant for consideration in her column in Parade Magazine. Craig wrote: Suppose you’re on Monty Hall’s Let’s Make a Deal! You are given the choice of three doors, behind one door there is a car, the others, goats. You pick a door, say 1, Monty opens another door, say 3, which has a goat. Monty says to you “Do you want to pick door 2?” Is it to your advantage to switch your choice of doors?

Marilyn gave a solution concluding that you should switch, and if you do, your probability of winning is 2/3. Several irate readers, some of whom identified themselves as having a PhD in mathematics, said that this is absurd since after Monty has ruled out one door there are only two possible doors and they should still each have the same probability 1/2 so there is no advantage to switching. Marilyn stuck to her solution and encouraged her readers to simulate the game and draw their own conclusions from this. Other readers complained that Marilyn had not described the problem completely.

In particular, the way in which certain decisions were made during a play of the game was not specified. We will assume that the car was put behind a door by rolling a three-sided die which made all three choices equally likely. Monty knows where the car is, and always opens a door with a goat behind it. Finally, we assume that if Monty has a choice of doors (i.e., the contestant has picked the door with the car behind it), he chooses each door with probability 1/2. Marilyn clearly expected her readers to assume that the game was played in this manner.

As in the case with most apparent paradoxes, this one can be resolved through careful analysis.

We begin by describing a simpler related question. We say that a contestant is using the “stay” strategy if he picks a door, and, if offered a chance to switch to another door, declines to do so (i.e., he stays with his original choice). Similarly, we say that the contestant is using the “switch” strategy if he picks a door, and, if offered a chance to switch to another door, takes the offer.

Now suppose that a contestant decides in advance to play the “stay” strategy. His only action in this case is to pick a door (and decline an invitation to switch, if one is offered).

What is the probability that he wins a car? The same question can be asked about the “switch” strategy.

Using the “stay” strategy, a contestant will win the car with probability 1/3. On the other hand, if a contestant plays the “switch” strategy, then he will win whenever the door he originally picked does not have the car behind it, which happens with probability 2/3.

This very simple analysis, though correct, does not quite solve the problem that Craig posed. Craig asked for the conditional probability that you win if you switch, given that you have chosen door 1 and that Monty has chosen door 3. To solve this problem, we compute the conditional probability given the information.

This is a process that takes place in several stages: the car is put behind a door, the contestant picks a door, and finally Monty opens a door.

Thus it is natural to analyze this using a probability tree. The resulting tree and the probabilities are shown in the Figure 5.1.

It is tempting to reduce the tree’s size by making certain assumptions such as: “Without loss of generality, we will assume that the contestant always picks door 1.” We have chosen not to make any such assumptions, in the interest of clarity.

Now the given information, namely that the contestant chose door 1 and Monty chose door 3, means only two paths through the tree are possible (see the Figure 5.2).

For one of these paths, the car is behind door 1 and for the other it is behind door 2. The path with the car behind door 2 is twice as likely as the one with the car behind door 1. Thus the conditional probability is 2/3 that the car is behind door 2 and 1/3 that it is behind door 1, so if you switch you have a 2/3 chance of winning the car, as Marilyn claimed.

At this point, you may think that the two problems above are the same, since they have the same answers.

Recall that we assumed in the original problem if the contestant chooses the door with the car, so that Monty has a choice of two doors, he chooses each of them with probability 1/2. Now suppose instead that in the case that

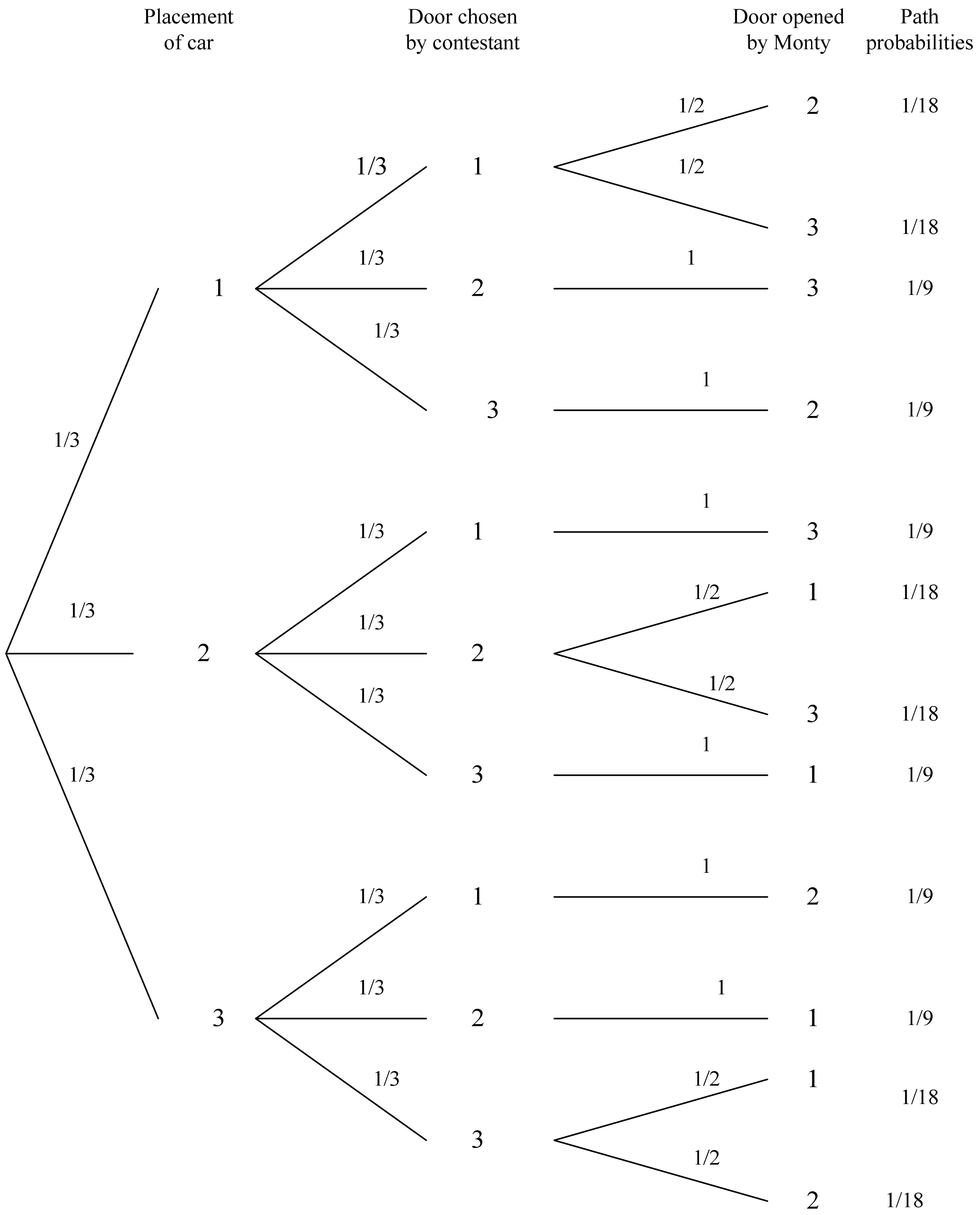


Figure 5.1. The Monty Hall problem

he has a choice, he chooses the door with the larger number with probability 3/4. In the "switch" vs. "stay" problem, the probability of winning with the "switch" strategy is still 2/3. However, in the original problem, if the contestant switches, he wins with probability 4/7. You can check this by noting that the same two paths as before are the only two possible paths in the tree. The path leading to a win, if the contestant switches, has probability 1/9, while the path which leads to a loss, if the contestant switches (i.e. a win, if the contestant doesn't switch), has probability 1/12.

You can read the article devoted to the theme http://plus.maths.org/issue32/features/wilson/index.html)

You might have the impression at this stage that the first step towards the solution of a probability problem is always a specification of a sample space. In fact one seldom needs an explicit listing of the sample space; an assignment of (conditional) probabilities to well chosen events is usually enough to set the probability machine in action. Only in cases of possible confusion, or great mathematical precision, do I find a list of possible outcomes worthwhile to contemplate. *Construction of a sample space is a non-trivial exercise as a rule.*

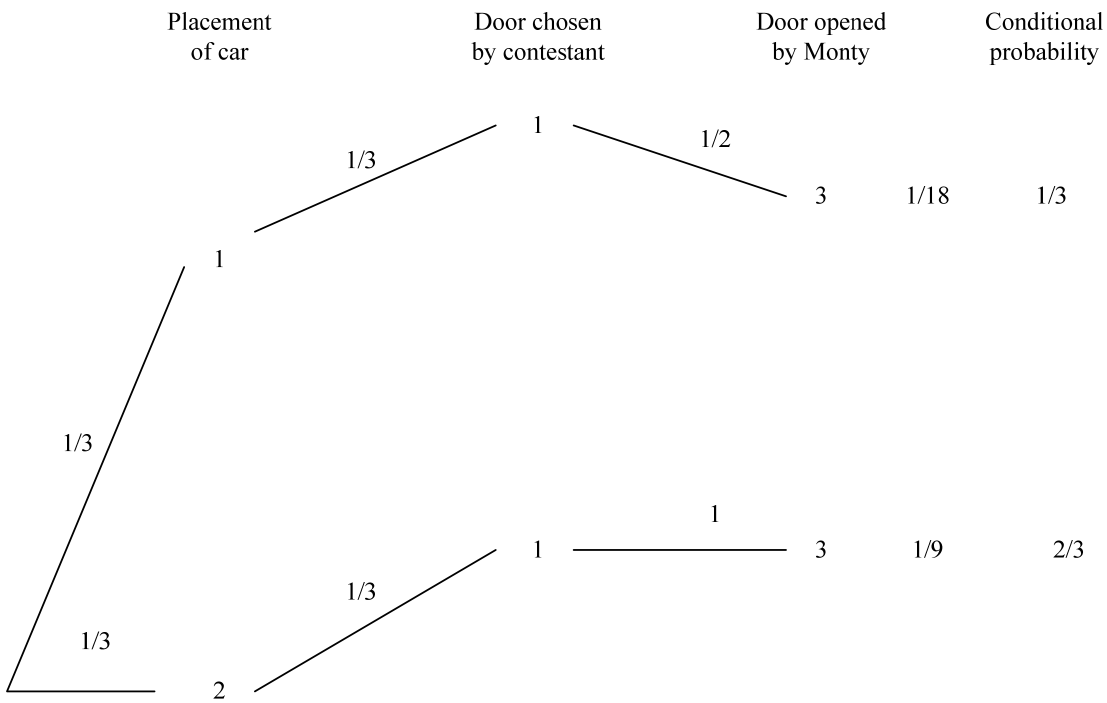


Figure 5.2. Conditional probabilities for the Monty Hall problem

Study more puzzles concerning conditional probability.

1) Consider a family with two children. Given that one of the children is a boy, what is the probability that both children are boys?

One way to approach this problem is to say that the other child is equally likely to be a boy or a girl, so the probability that both children are boys is 1/2.

The “textbook” solution would be to draw the tree diagram and then form the conditional tree by deleting paths to leave only those paths that are consistent with the given information. The result is shown in the Figure 5.3.

We see that the probability of two boys given a boy in the family is not 1/2 but rather 1/3.

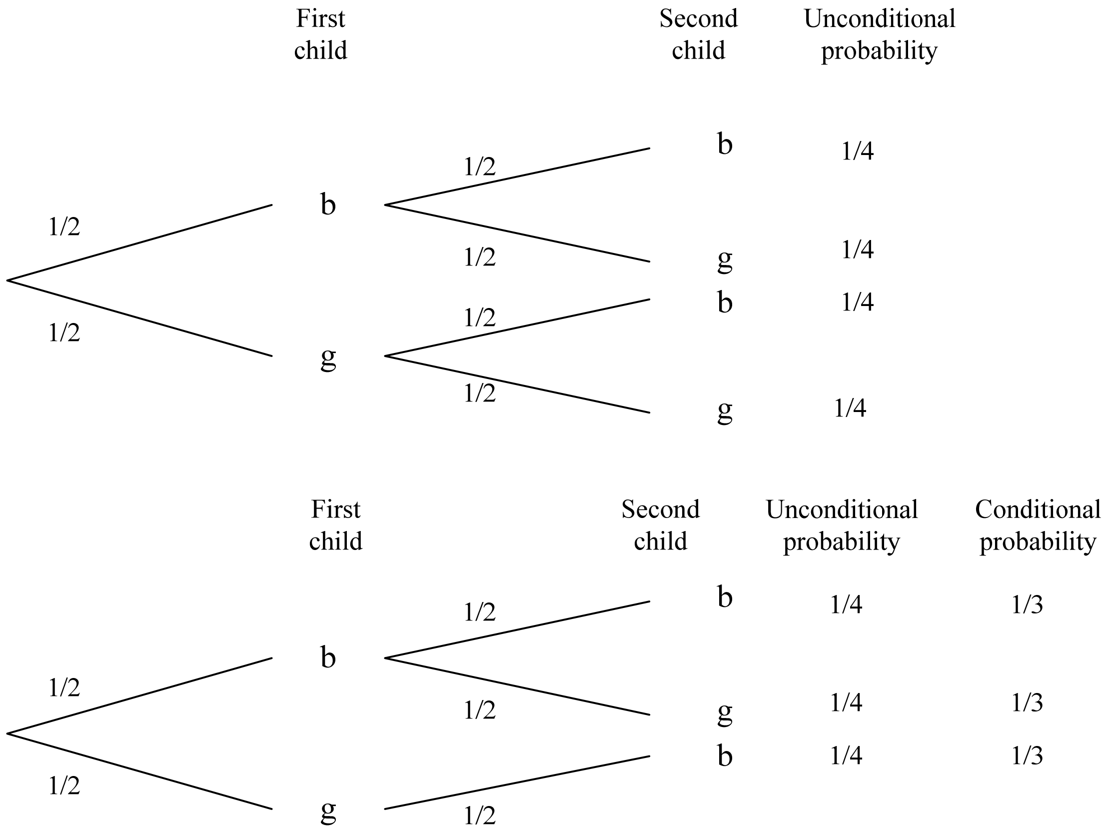


Figure 5.3. A family with two children (“textbook” solution)

The answer to conditional probabilities of this kind can change depending upon how the information given was actually obtained. For example, 1/2 is the correct answer for the following scenario.

Mr. Smith is the father of two. We meet him walking along the street with a young boy whom he proudly introduces as his son. What is the probability that Mr. Smith’s other child is also a boy?

As usual we have to make some additional assumptions. For example, we will assume that if Mr. Smith has a boy and a girl, he is equally likely to choose either one to accompany him on his walk. In the Figure 5.4 we show the tree analysis of this problem and we see that 1/2 is, indeed, the correct answer.

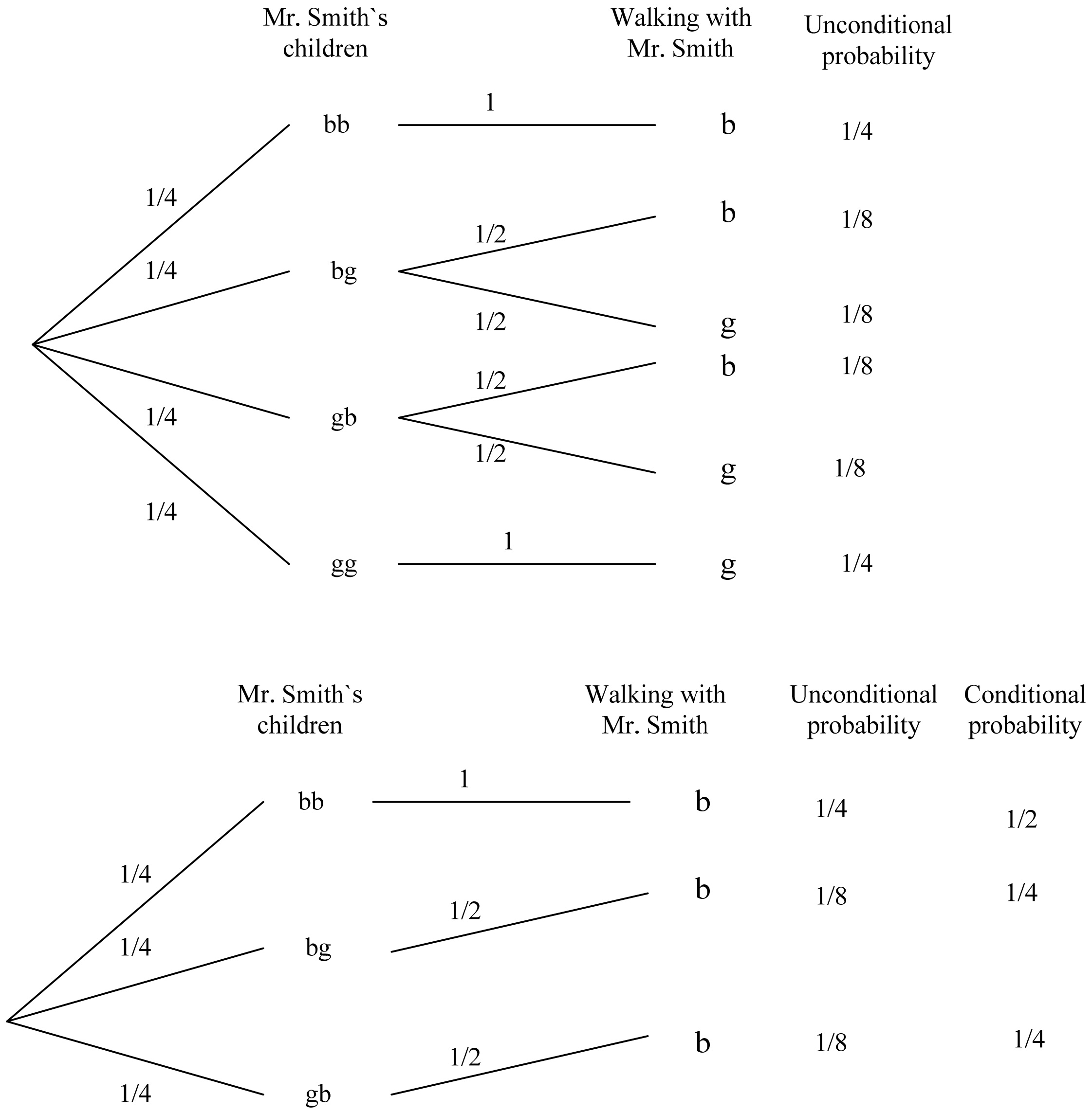


Figure 5.4. A family with two children

It is not so easy to think of reasonable scenarios that would lead to the classical 1/3 answer. An attempt was made by Stephen Geller in proposing this problem to Marilyn vos Savant. Geller's problem is as follows:

A shopkeeper says she has two new baby beagles to show you, but she doesn't know whether they're both male, both female, or one of each sex. You tell her that you want only a male, and she telephones the fellow who's giving them a bath. "Is at least one a male?" she asks. "Yes," she informs you with a smile.

What is the probability that the other one is male? The reader is asked to decide whether the model which gives an answer of 1/3 is a reasonable one to use in this case.

In these examples, the apparent paradoxes could easily be resolved by clearly stating the model that is being used and the assumptions that are being made.

2) Two envelopes each contain a certain amount of money. One envelope is given to Ali and the other to Baba and they are told that one envelope contains twice as much money as the other. However, neither knows who has the larger prize. Before anyone has opened their envelope, Ali is asked if she would like to trade her envelope with Baba. She reasons as follows: Assume that the amount in my envelope is x. If I switch, I will end up with $x/2$ with probability 1/2, and $2x$ with probability 1/2. If I were given the opportunity to play this game many times, and if I were to switch each time, I would, on average, get $0.5 \times (x/2) + 0.5 \times (2x) = 1.25x$.

This is greater than my average winnings if I didn't switch.

Of course, Baba is presented with the same opportunity and reasons in the same way to conclude that he too would like to switch. So they switch and each thinks that his/her net worth just went up by 25 %.

Since neither has yet opened any envelope, this process can be repeated and so again they switch. Now they are back with their original envelopes and yet they think that their fortune has increased 25 % twice. By this reasoning, they could convince themselves that by repeatedly switching the envelopes, they could become arbitrarily wealthy. Clearly, something is wrong with the above reasoning, but where is the mistake?

One of the tricks of making paradoxes is to make them slightly more difficult than is necessary to further befuddle us. In this paradox we could just have well started with a simpler problem. Suppose Ali and Baba know that I am going to give then either an envelope with \$5 or one with \$10 and I am going to toss a coin to decide which to give to Ali, and then give the other to Baba.

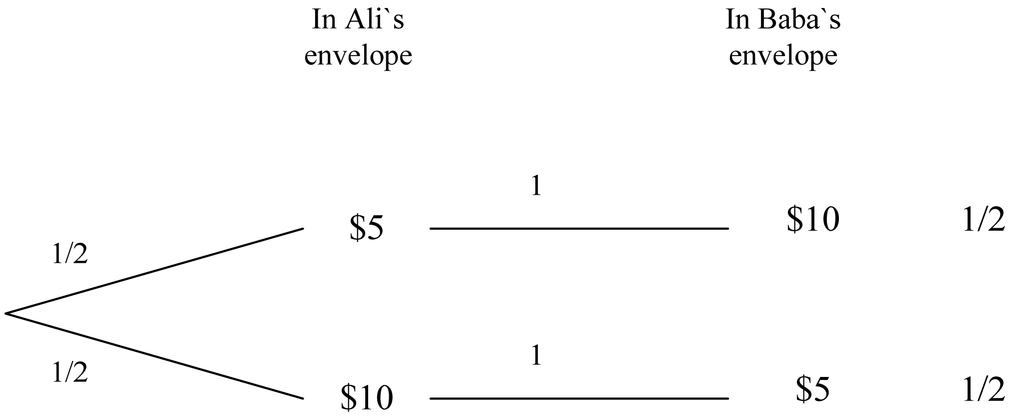


Figure 5.5. Two envelopes problem

Then Ali can argue that Baba has $2x$ with probability 1/2 and $x/2$ with probability 1/2. This leads Ali to the same conclusion as before. But now it is clear that this is nonsense, since if Ali has the envelope containing $5, Baba cannot possibility have half of this, namely $2.50, since that was not even one of the choices. Similarly, if Ali has $10, Baba cannot have twice as much, namely $20. In fact, in this simpler problem the possibly outcomes are given by the tree diagram in the Figure 5.5.

From the diagram, it is clear that neither is made better off by switching.

In the above example, Ali's reasoning is incorrect because he infers that if the amount in his envelope is x, then the probability that his envelope contains the smaller amount is 1/2, and the probability that her envelope contains the larger amount is also 1/2. In fact, these conditional probabilities depend upon the distribution of the amounts that are placed in the envelopes.

3) *Simpson's paradox* refers to the phenomenon whereby an event C increases the probability of E in a given population p and, at the same time, decreases the probability of E in every subpopulation of p. In other words,

$$P(E/C) > P(E/C^c), \qquad (5.1)$$

$$P(E/C,F) < P(E/C^c,F), \qquad (5.2)$$

$$P(E/C,F) < P(E/C^c,F^c). \qquad (5.3)$$

if F and F^c are two complementary properties describing two subpopulations.

Although such order reversal might not surprise students of probability, it is paradoxical when given causal interpretation. For example, if we associate C (connoting *cause)* with taking a certain drug, E (connoting *effect)* with recovery, and F with being a female then − under the causal interpretation of (5.2)−(5.3) the drug seems to be harmful to both males and females yet beneficial to the population as a whole (equation (5.1)). Intuition deems such a result impossible, and correctly so.

The tables in Figure 5.6 represent Simpson's reversal numerically. We see that, overall, the recovery rate for patients receiving the drug (C) at 50 % exceeds that of the control (C^c) at 40 % and so the drug treatment is apparently to be preferred. However, when we inspect the separate tables for males and females, the recovery rate for the untreated patients is 10 % higher than that for the treated ones, for males and females both.

(a)

Combined	E	E^c		Recovery Rate
Drug (C)	20	20	40	50 %
No Drug (C^c)	16	24	40	40 %
	36	44	80	

(b)

Males	E	E^c		Recovery Rate
Drug (C)	18	12	30	60 %
No Drug (C^c)	7	3	10	70 %
	25	15	40	

(c)

Females	E	E^c		Recovery Rate
Drug (C)	2	8	10	20 %
No Drug (C^c)	9	21	30	30 %
	11	29	40	

Figure 5.6. Recovery rates under treatment (C) and control (C^c) for males, females, and combined

The conditioning operator in probability calculus stands for the evidential conditional “given that we see”. Accordingly, the inequality $P(E/C) > P(E/C^c)$ is not a statement about C being a positive causal factor for E, but rather about C being positive *evidence* for E, which may be due to spurious confounding factors that cause both C and E. In our example, the drug appears beneficial overall because the males, who recover (regardless of the drug) more often than the females, are also more likely than the females to use the drug. Indeed, finding a drug-using patient (C) of unknown gender, we would do well inferring that the patient is more likely to be a male and hence more likely to recover, in perfect harmony with (5.1) − (5.3).

The standard method for dealing with potential confounders of this kind is to “hold them fixed,” namely, to condition the probabilities on any factor that might cause both C and E. In our example, if being a male (F^c) is perceived to be a cause for both recovery (E) and drug usage (C), then the effect of the drug needs to be evaluated separately for men and women (as in (5.2)−(5.3)) and averaged accordingly.

Thus, assuming F is the only confounding factor, (5.2) − (5.3) properly represent the efficacy of the drug in the respective populations while (5.1) represents merely its evidential weight in the absence of gender information, and the paradox dissolves.

5.1. Exercises

1. *The Prisoner's Dilemma*. Three prisoners, Al, Bob, and Charlie, with apparently equally good records have applied for parole. The parole board has decided to release two of the three, and the prisoners know this but not which two. A warder friend of prisoner Al knows who are to be released. Prisoner Al realizes that it would be unethical to ask the warder if he, Al, is to be released, but thinks of asking for the name of one prisoner other than himself who is to be released.

He thinks that before he asks, his chances of release are 2/3. He thinks that if the warder says "Bob will be released," his own chances have now gone down to 1/2, because either Al and Bob or B and Charlie are to be released. And so Al decides not to reduce his chances by asking. However, Al is mistaken in his calculations. Please, criticize Al's reasoning.

Assume the warder says that Bob to be released and prisoner Charlie overhears the conversation between Al and the warder. The warder might have nominated Charlie as a prisoner to be released. The fact that he didn't do so conveys some information to Charlie. Do you see why Al and Charlie can infer different information from the warder's reply? What is the conditional probability for Charlie to be released?

2. Here are three variations of the Monty Hall problem.

(a) Suppose that everything is the same except that Monty forgot to find out in advance which door has the car behind it. In the spirit of "the show must go on," he makes a guess at which of the two doors to open and gets lucky, opening a door behind which stands a goat. Now should the contestant switch?

(b) You have observed the show for a long time and found that the car is put behind door 1 45 % of the time, behind door 2 40 % of the time and behind door 3 15 % of the time. Assume that everything else about the show is the same. Again you pick door 1. Monty opens a door with a goat and offers to let you switch. Should you? Suppose you knew in advance that Monty was going to give you a chance to switch. Should you have initially chosen door 1?

(c)* Devise a variant game in which the contestant is presented with 10 doors. Again behind one of them is a car, behind the others booby prizes. After the contestant picks a door, Monty (or his avatar) opens just seven of the remaining nine unopened doors, but is careful never to open the door hiding the car. There are now three unopened doors − the one that the contestant originally picked and two others. Which strategy works best, switching to one of the other two unopened doors or sticking with the original pick? Furthermore, what is the probability of winning by following these two strategies?

One more question: Can you think of any real-world situations − crime mysteries, world politics, administrative deceptions − which might be modeled on some close variant of the Monty Hall problem? That is, are there situations in which the "contestant," say, a reporter, must choose among various alternatives and the "host," say, an official, knows the true answer, but is evasive about it and instead answers a question different from the one the contestant asks?

3. *Box Paradox*. A cabinet has three drawers. In the first drawer there are two gold balls, in the second drawer there are two silver balls, and in the third drawer there is one silver and one gold ball. A drawer is picked at random and a ball chosen at random from the two balls in the drawer. Given that a gold ball was drawn, what is the probability that the drawer with the two gold balls was chosen?

4. The following problem is called the *two aces problem*. This problem was also submitted to Marilyn vos Savant by the master of mathematical puzzles Martin Gardner, who remarks that it is one of his favorites.

A bridge hand has been dealt, i.e. thirteen cards are dealt to each of four player.

Given that your partner has at least one ace, what is the probability that he has at least two aces? Given that your partner has the ace of hearts, what is the probability that he has at least two aces? Answer these questions for a version of bridge in which there are eight cards, namely four aces and four kings, and each player is dealt two cards. (You may wish to solve the problem with a 52-card deck.)

It is natural to ask “How do we get the information that the given hand has an ace?” Consider two different ways that we might get this information. (Again, assume the deck consists of eight cards.)

(a) Assume that the person holding the hand is asked to “Name an ace in your hand” and answers “The ace of hearts.” What is the probability that he has a second ace?

(b) Suppose the person holding the hand is asked the more direct question “Do you have the ace of hearts?” and the answer is “yes”. What is the probability that he has a second ace?

5. For a real life example, consider the average SAT (Standardized Aptitude Test) verbal score. It was 504 in 1981 and 21 years later in 2002 it was again 504. However when we break things down by ethnic groups, we see that all of them increased their scores:

	1981	2002
non-Hispanic whites	519	527
African Americans	412	431
Mexican Americans	438	446
Asian Americans	474	501

Can you explain the result? What additional information you need to take into consideration?

6. Discrete random variables

The excitement that a gambler feels
when making a bet is equal to the amount
he might win times the probability of winning it.
Blaise Pascal

In many experiments, outcomes are defined in terms of numbers or may be associated with numbers, if we so choose. In either case, we want to assign probabilities directly to these numbers, as well as to the underlying events.

A **random variable** is a numerical value determined by the outcome of a random experiment.

This is important to be able to use arithmetic operations.

A random variable is a variable whose values cannot be predicted for sure. On the other hand, to know the value of sin x it suffices to plug the argument x in sin x, or the number of days in a year are not random variables.

Capital X is used for the variable and lower-case x's for its values. Formally, $X(\ )$ is a function with domain Ω and range $D \subset R$, and so for each $\omega \in \Omega, X(\omega) = x \in D$.

We have seen a number of examples of random variables.

- Roll two dice and let X = the sum of the two numbers that appear.
- Roll a die until a 4 appears and let X = the number of rolls we need.
- Flip a coin 10 times and let X = the number of Heads we get.
- Draw 13 cards out of a deck of 52 and let X = the number of Hearts we get.
- (Darts). You throw one dart at a conventional dartboard. A natural sample space is the set of all possible points of impact. This is of course uncountable because it includes every point of the dartboard, much of the wall, and even parts of the floor or ceiling if you are not especially adroit. However, your score $X(\omega)$ is one of a finite set of integers lying between 0 and 60, inclusive.

In these five cases X is a **discrete random variable**. That is, there is a finite or countable sequence of possible values. In contrast, the height of a randomly chosen person or the time they spent waiting for the bus this morning are **continuous random variables**.

6.1. Distributions

The **distribution** (or **probability mass function**) of a discrete random variable, is described by giving the value of $P(X = x)$ for all values of x. In each case, we will only give the values of $P(X = x)$ when $P(X = x) > 0$. The other values we do not mention are 0. The distribution of a discrete random variable is defined by following way:

Discrete random variable with n values

Values of X (all possible)	Probabilities
x_1	p_1
x_2	p_2
……………………………	
x_n	p_n

Note that $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} p(X = x_i) = P(\Omega) = 1$.

It is obvious that if X and Y are discrete random variables such that $Y = g(X)$, where $g(\ )$ is a real valued function defined on R, then Y has probability mass function (p.m.f.) given by $P(Y = g(X) = y) = \sum_{i:g(x_i)=y} P(X = x_i)$.

Example 1 (**Discrete Uniform Distribution**). Roll a die and let X = the number that appear.

X	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

If X is a random variable which represents the outcome of an experiment of this type, we say that X is uniformly distributed.

In general, if the sample space Ω is of size n, where $0 < n < \infty$, then the distribution is defined to be $1/n$ for all $\omega \in \Omega$.

Consider five examples above.

Example 2. Roll two dice and let X = the sum of the two numbers that appear.

We have computed the distribution before:

X	2	3	4	5	6	7	8	9	10	11	12
$P(X = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Example 3 (**Geometric distribution**). If we repeat an experiment with probability p of success until a success occurs then the number of trials required N has the following distribution $P(N = n) = (1 - p)^{n-1}p$ for $n = 1, 2, \ldots$

In words, N has a geometric distribution with parameter p, a phrase we will abbreviate as $N = geometric\ (p)$.

To check the formula note that in order to first have success on trial n, we must have $n - 1$ failures followed by a success, which has probability $(1 - p)^{n-1}p$. In the example at the beginning of the section, success is rolling a 4, so $p = 1/6$.

For an example of the use of the geometric distribution, we consider

Birthday problem (*II*). How large must the group be so that there is a probability > 0.5 that someone will have the same birthday as you do?

In our first encounter with the birthday problem it was surprising that the size needed to have two people with the same birthday was so small. This time the surprise goes in the other direction. Assuming 365 equally likely birthdays, a naive guess is that 183 people will be enough. However in a group of n people the probability all will fail to have your birthday is $(364/365)^n$. Setting this equal to 0.5 and solving $n = \ln(0.5)/\ln(364/365) = -0.69314/(-0.0027435) = 252.7$.

So we need 253 people. The "problem" is that many people in the group will have the same birthday so the number of different birthdays is smaller than the size of the group.

Consider an experiment with only two possible outcomes (we call them success and failure) which happen with the probability of p and $q \equiv 1 - p$ respectively (examples: flipping a coin, rolling a die and being concerned only with obtaining a six versus any other number, a team winning or losing a game, drawing a marble from a box with red and blue marbles, shooting against a target to either hit or miss, etc.)

We define a random variable X as the number of successes one gets in one round (or trial) of this experiment (also called an indicator variable of the success). Its distribution is obviously

X	0	1
$P(X = x)$	q	p .

In this case we say that X has **Bernoulli Distribution** with parameter p. Consider a generalization of this distribution.

Example 4 (**Binomial distribution**).

If we perform an experiment n times independently (the results of any previous trials don't reflect to the results of following ones) and on each trial there is a probability p of success then the number of successes S has the distribution $P(S = k) = p_n(k) = C_n^k p^k (1 - p)^{n-k}$ for $k = 0, \ldots, n$.

In words, S has a binomial distribution with parameters n and p, a phrase we will abbreviate as $S = binomial\ (n, p)$.

The third example mentioned at the beginning is the special case: $n = 10$ and $p = 1/2$.

Example 5 (**Hypergeometric distribution**). Consider an urn with M red balls and N black balls. If we draw out n balls then the number of red balls we get, R, has $P(R = r) = \dfrac{C_M^r C_N^{n-r}}{C_{M+N}^n}$ for $r = 0, \ldots, n$.

Here the denominator gives the number of ways of picking n of the $M+N$ balls and the numerator gives the number of ways of picking r of the M red balls and $n - r$ of the N black balls. (By convention $C_n^k = 0$ if $k > n$ or $k < 0$.)

For the forth example above $M = 13$ Hearts, $N = 39$ other cards and we draw $n = 13$.

Our new example is:

Example 6 (**Poisson distribution**). X is said to have a Poisson distribution with parameter λ *if* $P(X = k) = e^{-\lambda}\lambda^k/k!$ for $k = 0, 1, 2, \ldots$.

Here $\lambda > 0$ is a parameter. To see that this is a distribution we recall $\sum_{k=0}^{\infty} \lambda^k / k! = e^{\lambda}$, so the proposed probabilities are nonnegative and sum to 1.

The Poisson distribution arises in a number of situations because of the next result.

Poisson approximation to the binomial. Suppose S_n has a binomial distribution with parameters n and p_n. If $p_n \to 0$ and $np_n \to \lambda$ as $n \to \infty$ then

$$P(S_n = k) \to e^{-\lambda}\lambda^k/k! \tag{6.1}$$

In words, if we have a large number of independent events with small probability then the number that occur has approximately a Poisson distribution. The key to the proof is the following fact: If $p_n \to 0$ and $np_n \to \lambda$ *as* $n \to \infty$ then $(1 - p_n)^n \to e^{-\lambda}$ *as* $n \to \infty$.

To illustrate the use of this we return to the probability in People vs. Collins (see 29* in Exercises 4). There $n = 1{,}000{,}000$ and $p_n = 1/12{,}000{,}000$ so $(1 - 1/12{,}000{,}000)^{1{,}000{,}000} \to e^{-1/12} = 0.9200$.

When we apply (6.1) we think, "If S_n = binomial (n, p) and p is small then S_n is approximately Poisson (np)." The next example also illustrates the use of this approximation and shows that the number of trials does not have to be very large for us to get accurate answers.

Example. Suppose we roll two dice 12 times and we let D be the number of times a double 6 appears. Here $n = 12$ and $p = 1/36$, so $np = 1/3$. We will now compare $P(D = k)$ with the Poisson approximation for $k = 0, 1, 2$.

$k = 0$ exact answer: $P(D = 0) = (1 - 1/36)^{12} = 0.7132$,

Poisson approximation: $P(D = 0) = e^{-1/3} = 0.7165$,

$k = 1$ exact answer: $P(D = 1) = C_{12}^1 (1/36)(1 - 1/36)^{11} = 0.2445$, Poisson approximation: $P(D = 1) = e^{-1/3}$ (1/3) = 0.2388,

$k = 2$ exact answer: $P(D = 2) = C_{12}^{2}(1/36)^2(1 - (1/36))^{10} = 0.0384$,

Poisson approximation: $P(D = 2) = e^{-1/3}(1/3)^2/2! = 0.0398$.

Some early data showing a close approximation to the Poisson distribution was the number of German soldiers kicked to death by cavalry horses between1875 and 1894.

The Poisson distribution can be used for births as well as for deaths. There were 63 births in Ithaca, NY between March 1 and April 8, 2005, a total of 39 days, or 1.615 per day. The next table gives the observed number of births per day and compares with the prediction from the Poisson distribution

	0	1	2	3	4	5	6
observed	9	12	9	5	3	0	1
Poisson	7.75	12.52	10.11	5.44	2.19	.71	.19

The Poisson distribution is often used as a model for the number of people who go to a fast-food restaurant between 12 and 1, the number of people who make a cell phone call between 1:45 and 1:50, or the number of traffic accidents in a day. To explain the reasoning in the last case we note that any one person has a small probability of having an accident on a given day, and it is reasonable to assume that the events A_i = "The ith person has an accident" are independent. Now it is not reasonable to assume that the probabilities of having an accident $p_i = P(A_i)$ are all the same, nor is it reasonable to assume that all women have the same probability of giving birth, but fortunately the Poisson approximation does not require this.

Example (**Lottery Double Winner**). The following item was reported in the February 14, 1986 edition of the New York Times: A New Jersey woman won the lottery twice within a span of four months. She won the jackpot for the first time on October 23, 1985 in the Lotto 6/39. Then she won the jackpot in the new Lotto 6/42 on February 13, 1986. Lottery officials calculated the probability of this as roughly one in 17.1 trillion. What do you think of this statement?

It is easy to see where they get this from. The probability of a person picked in advance of the lottery getting all six numbers right both times is $(1/C_{39}^{6})\times(1/C_{42}^{6}) = 1/(17.1\times1012)$.

One can immediately reduce this number by noting that the first lottery had some winner, who if they played only one ticket in the second lottery had a $1/C_{42}^{6}$ chance.

The odds drop even further when you consider that there are a large number of people who submit more than one entry for each weekly draw and that wins on October 23, 1985 and February 13, 1986 is not the only combination. Suppose for concreteness that each week 50 million people play the lottery and buy five tickets. The probability of one person winning on a given week is $p_1 = 5/C_{42}^6 = 9.915\times10^{-7}$.

The number of times one person will win a jackpot in the next 200 drawings is roughly Poisson with mean $\lambda_1 = 200p_1 = 1.983\times10^{-4}$. (Below we will see that the parameter λ_1 represents the average number of times one person wins a jackpot in the next 200 drawings.)

The probability that a given player wins the jackpot two or more times is $p_0 = 1 - e^{-\lambda_1} - \lambda_1 e^{-\lambda_1} = 1.965\times10^{-8}$.

The number of double winners in a population of 50 million players is Poisson with mean $\lambda_0 = 50{,}000{,}000p_0 = 0.9825$ so the probability of no double winner is $e^{-0.9825} = 0.374$.

6.2. Moments, Mean, Variance

Climate is what you expect, weather is what you get.
Robert Heinlein

The **expected value** of X is defined to be

$$EX = \sum_{x_i \in D} x_i p_i\ .$$

In words, we multiply each possible value by its probability and sum.

This is also known as the **expectation**, or **mean**, or **average** or first moment of X.

Note that the series (if D is infinite) above is absolutely convergent, otherwise the expected value doesn't exist (by definition).

Example. Roll one die and let X be the number that appears. $P(X = x) = 1/6$ for $x = 1, 2, 3, 4, 5, 6$ so

$$EX = 1\times\frac{1}{6} + 2\times\frac{1}{6} + 3\times\frac{1}{6} + 4\times\frac{1}{6} + 5\times\frac{1}{6} + 6\times\frac{1}{6} = \frac{21}{6} = 3\frac{1}{2}.$$

In this case the expected value is just the "usual" average of the six possible values.

Example (Roulette). If you play roulette and bet \$1 on black then you win \$1 with probability 18/38 and you lose \$1 with probability 20/38, so the expected value of your winnings X is $EX = 1\times18/38 + (-1)\times20/38 = -0.0526$.

If you play n times and let X_i be your winnings on the ith play then the *law of large numbers* implies that $(X_1 + \ldots + X_n)/n$ will be close to -0.0526. In words, in the long run you will lose about 5.26 cents per play.

Example (**Bernoulli distribution**). Suppose $P(X = 1) = p$ and $P(X = 0) = 1 - p$. Compute EX.

$$EX = pP(X = 1) + (1 - p)P(X = 0) = p \qquad (6.2)$$

It is obvious that if X and Y are discrete random variables such that $Y = g(X)$ then $EY = \sum_{x_i \in D} g(x_i) p_i$.

Let $g(x) = x^k$, $E(X^k)$ is the k**th moment of** X. When $k = 1$ this is the first moment or mean of X.

Example. Suppose X is the result of rolling one die. Compute EX^2 – the second moment of X.

$EX^2 = (1 + 4+9 + 16+ 25 + 36)/6 = 91/6 = 15.1666.$

Now we establish some important properties of the expected value.

Let X be a random variable with the mean $E(X)$, and let a and b be constants. Then:

$$E(X + b) = EX + b$$
$$E(aX) = aEX \qquad (6.3)$$

In words, if we add 5, for example, to a random variable then we add 5 to its expected value. If we multiply a random variable by 3, for example, we multiply its expected value by 3.

If X and Y are two discrete random variables, and if $E(X)$ and $E(Y)$ exist, then the mean $E(X + Y)$ exists and we have

$$E(X + Y) = E(X) + E(Y). \qquad (6.4)$$

From (6.3) and (6.4) it follows

$$E(c_1X_1 + \cdot\cdot\cdot + c_nX_n) = c_1EX_1 + \ldots + c_nEX_n \qquad (6.5)$$

(*linearity* of the expected value).

Example (*Binomial distribution*). When $P(S = k) = C_n^k p^k (1 - p)^{n-k}$ for $k = 0, \ldots, n$ we have $ES = np$.

This answer is intuitive clear – each trial brings a probability p of success so the expected number of successes in n trials is np. It is not easy to get the answer directly from the definition of the distribution so we will take a different approach.

Let $X_i = 1$ if there is a success on the ith trial of n independent trials and 0 otherwise, so that

$$S = X_1 + \cdot \ldots + X_n. \qquad (6.5^*)$$

The fact that $ES = np$ now follows from (6.4).

Example (*Hypergeometric distribution*). If we draw out n balls from an urn with M red balls and N black balls, then the number of red balls, R, has $P(R = r) = \dfrac{C_M^r C_N^{n-r}}{C_{M+N}^n}$ for $r = 0, \ldots, n$.

In this case $ER = nM/(M + N)$.

Let $X_i = 1$ if a red ball is drawn on the ith trial and 0 otherwise, so that $R = X_1 + . . . + X_n$. Each ball drawn has probability $M/(M + N)$ of being red, so $EX_i = M/(M + N)$, and it follows from (6.4) that $ER = nM/(M + N)$.

Example (*Poisson distribution*). If $P(X = k) = e^{-\lambda}\lambda^k/k!$ for $k = 0, 1, 2, . . .$ then $EX = \lambda$.

To see that this is the right answer, remember that Poisson implies Binomial $(n, \lambda/n)$ approaches Poisson (λ) as $n \to \infty$ and the Binomial have mean $n(\lambda/n) = \lambda$. To get this directly from the formula, note that since the $k = 0$ term makes no contribution to the sum, $EX = \sum_{k=o}^{\infty} ke^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} ke^{-\lambda}\frac{\lambda^k}{k!} = \lambda\sum_{k=o}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = \lambda$, since $\sum_{k=o}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = 1$.

Consider two more examples concerning to Poisson distribution.

(a) The distribution of V-2 rocket hits in south London during World War II. The area under study was divided into 576 area of equal size. There were a total of 537 hits or an average of 0.9323 per subdivision. Using the Poisson distribution the probability a subdivision is not hit is $e^{-.9323} = .3936$. Multiplying by 576 we see that the expected number not hit was 226.71 which agrees well with the 229 that were observed not to be hit.

(b) A typesetter makes, on the average, one mistake per 1000 words. Assume that he is setting a book with 100 words to a page. Let S_{100} be the number of mistakes that he makes on a single page. Then the exact probability distribution for S_{100} would be obtained by considering S_{100} as a result of 100 Bernoulli trials with $p = 1/1000$. The expected value of S_{100} is $100(1/1000) = 0.1$. The exact probability that $S_{100} = j$ is $p_{100}(j)$ and the Poisson approximation is $\frac{0.1^j}{j!}e^{-0.1}$.

Example (*Geometric distribution*). When $P(N = n) = (1 - p)^{n-1}p$, for $n = 1, 2, 3, . . .$ we have $EN = 1/p$.

This answer is intuitive. We have an average of p successes per trial, so in n trials we have an average of np successes and if we want $np = 1$ we need $n = 1/p$.

To get this from the definition, we begin with the sum of the geometric series $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$ and differentiate with respect to x to get $\sum_{k=0}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$. Dropping the $k = 0$ term from the left since it is 0 and setting $x = 1 - p$ we receive $\sum_{k=0}^{\infty} k(1-p)^{k-1} = \frac{1}{p^2}$. Multiplying each side by p we have $\sum_{k=0}^{\infty} kp(1-p)^{k-1} = \frac{1}{p}$.

Example (*China's one child policy*). In order to limit the growth of its population, the Chinese government decided to a limit a family to having just one child. An alternative that was suggested was the "one son" policy: as long a woman has only female children she is allowed to have more children. One concern voiced about this policy was that no family would have more than one son but many families would have several girls. This concern leads to our question: How would the one son policy affect the ratio of male to female?

To simplify the problem we assume that a family will keep having children until it has a male child. Assuming that male and female children are equally likely and the sexes of successive children are independent, he total number of children has a geometric distribution with success probability 1/2, so by the previous example the expected number of children is 2. There is always one male child, so the expected number of female children is 2 −1 = 1.

Does this continue to hold if some families stop before they have a male child? Consider for simplicity the case in which a family will stop when they have a male child or a total of three children. There are four outcomes

$P(M) = 1/2$

$P(FM) = 1/4$

$P(FFM) = 1/8$

$P(FFF) = 1/8$.

The average number of male children is 1/2+1/4+1/8 = 7/8 while the average number of female children is 1(1/4) + 2(1/8) + 3(1/8) = 7/8.

The last calculation makes the equality of the expected values look like a miracle, but it is not and the claim holds true if a family with k female children continues with probability p_k and stops with probability $1-p_k$. To explain this intuitively, if we replace M by +1 and F by −1, then childbirth is a *fair game*.

For the stopping rules under consideration the average winnings when we stop have mean 0, i.e., the expected number of male children equals the expected number of female children.

Fair and unfair games. One of the most valuable uses to which a gambler can put his knowledge of probabilities is to decide whether a game or proposition is fair, or equitable. To do this a gambler must calculate his "expectation". A gambler's expectation is the amount he stands to win multiplied by the probability of his winning it.

A game is a **fair game** if the gambler's expectation equals his stake.

If a gambler is offered 10 units each time he tosses a head with a true coin, and pays 5 units for each toss is this a fair game? The gambler's expectation is 10 units multiplied by the probability of throwing a head, which is 1/2. His expectation is 10 units×(1/2) = 5 units, which is what he stakes on the game, so the game is fair.

Suppose a gambler stakes 2 units on the throw of a die. On throws of 1, 2, 3 and 6 he is paid the number of units shown on the die. If he throws 4 or 5 he loses. Is this fair? This can be calculated as above. The probability of throwing any number is 1/6. His expectation is therefore 6/6+3/6+2/6+1/6, which equals 2, the stake for a throw, so the game is fair.

Petersburg paradox (*D. Bernoulli*). In the Petersburg casino, you pay an entrance fee c and you get the prize 2^T , where T is the number of times, the casino flips a coin until "head" appears (inclusive). For example, if the sequence of coin experiments would give "tail, tail, tail, head", you would win $2^4 - c = 16 - c$, the win minus the entrance fee. Fair would be an entrance fee which is equal to the expectation of the win, which is $\sum_{k=1}^{\infty} 2^k P(T = k) = \sum_{k=1}^{\infty} 1 = \infty$.

The paradox is that nobody would agree to pay even an entrance fee c = 10. What do you think about? Some discussion of the problem you can see in Solutions and Answers, but in general the problem is rather complicated one.

Now we will be interested in the expected values of various functions of random variables. The most important of these are the variance and the standard deviation which give an idea about how spread out the distribution is.

If $EX^2 < \infty$ then the **variance** of X is defined to be $\text{var}(X) = E(X - EX)^2$.

To illustrate this concept we will consider some examples. But first, we need a formula that enables us to more easily compute var (X):

$$\text{var}(X) = EX^2 - (EX)^2. \tag{6.6}$$

Proof. Letting $\mu = EX$ to make the computations easier to see, we have $\text{var}(X) = E(X - \mu)^2 = E\{X^2 - 2\mu X + \mu^2\} = EX^2 - 2\mu EX + \mu^2$ by (6.5) and the facts that $E(-2\mu X) = -2\mu EX$, $E(\mu^2) = \mu^2$. Substituting $\mu = EX$ now gives the result.

The reader should note that EX^2 means the expected value of X^2 and in the proof $E(X - \mu)^2$ means the expected value of $(X - \mu)^2$. When we want the square of the expected value we will write $(EX)^2$. This convention is designed to cut down on parentheses.

The variance measures how spread-out the distribution of X is.

We will show that

$$\text{var}(X + b) = \text{var}(X)$$
$$\text{var}(aX) = a^2 \text{ var } (X). \tag{6.7}$$

In words, the variance is not changed by adding a constant to X, but multiplying X by a multiplies the variance by a^2.

Proof. If $Y = X + b$ then the mean of Y, $EY = EX + b$ by (6.3) so

$\text{var } (X + b) = E\{(X + b) - (EX + b)\}^2 = E\{X - EX\}^2 = \text{var}(X)$.

If $Y = aX$ then $EY = aEX$ by (6.3) so $\text{var } (aX) = E\{(aX - aEX)^2\} = a^2E(X - \mu X)^2 = a^2\text{var } (X)$.

The scaling relationship (6.7) shows that if X is measured in feet then the variance is measured in feet^2. This motivates the definition of the **standard deviation** $\sigma(X) = \sqrt{\text{var}(X)}$, which is measured in the same units as X and has a nicer scaling property.

$$\sigma(aX) = |a|\,\sigma(X) \tag{6.8}$$

We get the absolute value here since $\sqrt{a^2} = |a|$.

Example. Roll one die and let X be the resulting number. Find the variance and standard deviation of X.

$EX = 7/2$ and $EX^2 = 91/6$ so $\text{var}(X) = EX^2 - (EX)^2 = 91/6 - 49/4 = 105/36 = 2.9166$ and $\sigma(X) = \sqrt{\text{var}(X)} = 1.7078$. The standard deviation $\sigma(X)$ gives the size of the "typical deviation from the mean".

Example (*New York Yankees 2004 salaries*). Salaries are in units of *M,* millions of dollars per year and for convenience have been truncated at the thousands place.

A. Rodriguez	21.726	D. Jeter	18.6
M. Mussina	16	K. Brown	15.714
J. Giambi	12.428	B. Williams	12.357
G. Sheffield	12.029	M. Rivera	10.89
J. Posada	9	J. Vazquez	9
J. Contreras	9	J. Olerud	7.7
H. Matsui	7	S. Karsay	6
E. Loazia	4	T. Gordon	3.5
P. Quantrill	3	K. Lofton	2.985
J. Lieber	2.7	T. lee	2
G. White	1.925	F. Heredia	1.8
R. Sierra	1	M. Cairo .	9
J. Falherty	.775	T. Clark	.75
E. Wilson	.7	O. Hernandez	.5
D. Osborne	.45	C.J. Nitowski	.35
J. DePaula	.302	B. Crosby	.301

The total team salary is 183,355,513. Dividing by 32 players gives a mean of 6.149 M dollars. The second moment is 73.778 M^2 so the variance is $73.778 - (6.149)^2 = 35.961M^2$ and the standard deviation is 5.996 M.

Example (*Geometric distribution*). Suppose $P(N = n) = (1-p)^{n-1}p$ for $n = 1, 2, \ldots$ and 0 otherwise. Compute the variance and standard deviation of X.

We learned that $EN = 1/p$. By formulae (6.6) variance var(N) $=\sum_{k=1}^{\infty} k^2(1-p)^{k-1}p - \frac{1}{p^2}$. To compute the first term we begin by observing that $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$. Differentiating this identity twice gives $\sum_{k=2}^{\infty} k(k-1)x^{k-2} = \frac{2}{(1-x)^3}$. Setting $x = 1-p$ gives $\sum_{k=2}^{\infty} k^2(1-p)^{k-2} - \sum_{k=2}^{\infty} k(1-p)^{k-2} = (1-p)^{-1}\left[\sum_{k=1}^{\infty} k^2(1-p)^{k-1} - 1 - EN/p + 1\right] = (1-p)^{-1}\left[\sum_{k=1}^{\infty} k^2(1-p)^{k-1} - 1/p^2\right] = \frac{2}{p^3}$.

From this it follows that $\sum_{k=1}^{\infty} k^2(1-p)^{k-1} = \frac{2(1-p)}{p^3} + \frac{1}{p^2}$ and var (N) = $\left(\frac{2(1-p)}{p^3} + \frac{1}{p^2}\right)p - \frac{1}{p^2} = \frac{1}{p^2} - \frac{1}{p} = \frac{1-p}{p^2}$. Taking the square root we see that $\sigma(X) = \sqrt{1-p}/p$.

Example (*Poisson distribution*). Suppose $P(X = k) = e^{-\lambda}\lambda^k/k!$ for $k = 0, 1, 2, \ldots$ and 0 otherwise. Compute the variance and standard deviation of X.

We learned that $EX = \lambda$. To compute EX^2 we begin by observing that the k = 0 and 1 terms make no contribution to the sum, so $E\{X(X-1)\} = \sum_{k=2}^{\infty} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!} = \lambda^2 \sum_{k=2}^{\infty} e^{-\lambda}\frac{\lambda^{k-2}}{(k-2)!} = \lambda^2 \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = \lambda^2$ since $\sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k=0}^{\infty} P(X = k) = 1$. From this it follows that $EX^2 = E\{X(X-1)\} + EX = \lambda^2 + \lambda$, var $(X) = EX^2 - (EX)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

Example (*Bernoulli distribution*). Suppose $X = 1$ with probability p and 0 with probability $(1-p)$. Compute the variance of X.

As we observed $EX = p$. To compute var $(X) = EX^2 - (EX)^2$ we note that $EX^2 = p\times 1^2 + (1-p)\times 0^2 = p$ so var $(X) = p - p^2 = p(1-p)$.

To receive the variance of the Binomial distribution by simple way we need the following definition that is designed analogous to the case of random events.

Random variables X and Y being **independent** means that so called *joint probability* $P(X = I \cap Y = j) = P(X = i)P(Y = j)$ for every possible combination of i and j (in words, each joint probability is a product of the two corresponding *marginal probabilities*).

Example (*Binomial distribution*). It is not difficult to prove that the variance of the sum of n independent random variables is the sum of the variances. Combining this with the previous example and (6.5^*) we see that the variance of the Binomial (n, p) is $np(1 - p)$.

6.3. Exercises

1. Suppose we roll three tetrahedral dice that have 1, 2, 3, and 4 on their four sides. Find the distribution for the sum of the three numbers.
2. Suppose we draw 3 balls out of an urn with 5 red and 4 black balls. Find the distribution for the number of red balls drawn.
3. How many children should a family plan to have so that the probability of having at least one child of each sex is at least 0.95?
4. Use the Poisson approximation to compute the probability that you will roll at least one double 6 in 24 trials of rolling two dice. How does this compare with the exact answer?
5. The probability of a three of a kind in poker is approximately 1/50. Use the Poisson approximation to compute the probability you will get at least one three of a kind if you play 20 hands of poker.
6. In one of the New York state lottery games, a number is chosen at random between 0 and 999. Suppose you play this game 250 times. Use the Poisson approximation to estimate the probability you will never win and compare this with the exact answer.
7. Suppose 1 % of a certain brand of Christmas lights is defective. Use the Poisson approximation to compute the probability that in a box of 25 there will be at most one defective bulb.
8. In February 2000, 2.8 % of Colorado's labor force was unemployed. Calculate the probability that in a group of 50 workers exactly one is unemployed.
9. An insurance company insures 3000 people, each of whom has a 1/1000 chance of an accident in one year. Use the Poisson approximation to compute the probability there will be at most 2 accidents.
10. Books from a certain publisher contain an average of 1 misprint per page. What is the probability that on at least one page in a 300 page book there are five misprints?
11. You want to invent a gambling game in which a person rolls two dice and is paid some money if the sum is 7, but otherwise he loses his money. How much should you pay him for winning a \$1 bet if you want this to be a fair game, that is, to have expected value 0?

12. A bet is said to carry 3 to 1 odds if you win \$3 for each \$1 you bet. What must the probability of winning be for this to be a fair bet?
13. A lottery has one \$100 prize, two \$25 prizes, and five \$10 prizes. What should you be willing to pay for a ticket if 100 tickets are sold?
14. In a popular gambling game, three dice are rolled. For a \$1 bet you win \$1 for each six that appears (plus your dollar back). If no six appears you lose your dollar. What is your expected value?
15. In the Las Vegas game Wheel of Fortune, there are 54 possible outcomes. One is labeled "Joker", one "Flag", two "20", four "10", seven "5", fifteen "2", and twenty-four "1". If you bet \$1 on a number you win that amount of money if the number comes up (plus your dollar back). If you bet \$1 on Flag or Joker you win \$40 if that symbol comes up (plus your dollar back). What bets have the best and worst expected value here?
16. In blackjack the dealer gets two cards, one of which you can see and one of which you cannot. When the dealer's visible card is an Ace, she offers you a chance to take out "insurance". You can bet \$1 that the invisible card is a face card or a 10. If it is, you win \$2, otherwise you lose \$1. What is the expected value of this bet

(a) if we assume that the dealer's other card was chosen at random from a second deck of 52?

(b) if we use the information that the dealer's Ace and our two cards, which are a 6 and an 8, came from the same deck of 52?

17. Twelve ducks fly overhead. Each of 6 hunters picks one duck at random to aim at and kills it with probability 0.6.

(a) What is the mean number of ducks that are killed?

(b) What is the expected number of hunters who hit the duck they aim at?

18. Suppose we pick 3 students at random from a class with 10 boys and 15 girls. Let X be the number of boys selected and Y be the number of girls selected. Find $E(X - Y)$.
19. A random variable has $P(X = x) = x/15$ for $x = 1, 2, 3, 4, 5$ and 0 otherwise. Find the mean and variance of X.
20. In a group of five items, two are defective. Find the distribution of N the number of draws we need to find the first defective item. Find the mean and variance of N.
21. Suppose X and Y are independent with $EX = 1$, $EY = 2$, var $(X) = 3$ and var $(Y) = 1$. Find the mean and variance of $3X + 4Y - 5$.
22. In a class with 18 boys and 12 girls, boys have probability 1/3 of knowing the answer and girls have probability 1/2 of knowing the answer to a typical question the teacher asks. Assuming that whether or not the students

know the answer are independent events, find the mean and variance of the number of students who know the answer.

23. Can we have a random variable with $EX = 3$ and $EX^2 = 8$?

Answers and Solutions

Exercises 1.5

3a.* The letter E, which is the most commonly used letter in the English language, does not appear even once in the paragraph.

Exercises 2.1

16.* (a) *Answer*: $3^{10} = 59049$.

(b) *Answer*: $\binom{10}{5,4,1} = 1260$.

(c) *Answer*: $\binom{12}{2} = 66$ (only one of these will have 5 wins, 4 losses and 1 tie).

17.* (a) *Answer*: $2^{20} = 1048576$.

(b) *Answer*: $\binom{20}{7} = 77520$.

(c) (Here, there is no "shortcut" formula, we have to do this individually, one by one, adding the results): $\binom{20}{17}+\binom{20}{18}+\binom{20}{19}+\binom{20}{20}=\binom{20}{3}+\binom{20}{2}+\binom{20}{1}+\binom{20}{0} = 1351$.

(d) $\binom{20}{2}+\binom{20}{1}+\binom{20}{0} = 211$.

18.* (a) *Answer* (same as the number of permutations of AAAFFFF DDDDCC): $\binom{13}{3,4,4,2} = 900{,}900$.

(b) *Answer*: we just have to arrange the four nationalities, say A, F, D and C: 4! = 24.

(c) *Answer* (13 of the original arrangements are duplicates now, as AAAFFFFDDDDCC, AAFFFFDDDDCCA,...,CAAAFFFFDDDDC are identical): 900,900/13 = 69300.

(d) *Answer* (circular arrangement of nationalities): 3! = 6.

19.* (a) If the books are treated as 12 identical copies of the same novel, then the only decision to be made is: how many books do we place on each shelf.

The answer follows from (Unordered Selection, Allowing Repetition) with $n = 3$ and $r = 12$ − for each book we have to select a shelf, but the order does not matter, and repetition is allowed: $\binom{14}{2} = 91$.

(b) If we treat the books as distinct and their order within each shelf important, we solve this in two stages:

First we decide how many books we place in each shelf, which was done in (a), then we choose a book to fill, one by one, each allocated slot (here we have 12×11×10×. . .×2×1 choices).

Answer: 91×12! = 43,589,145,600.

(c) Finally, if the books are considered all distinct, but their arrangement within each shelf is irrelevant, we simply have to decide which shelf each book will go to (similar to (a), order important now).

This can be done in $3\times3\times3\times...\times3 = 3^{12}$ = 531,441 number of ways.

20.* (a) *Solution*: Mr. A and Mr. B have to be first treated as a single item, for a total of 11 items. These can be permuted in 11! number of ways.

Secondly, we have to place Mr. A and Mr. B in the two chairs already allocate to them (2! ways).

Answer: 2×11! = 79,833,600.

(b) This set consists of those which did not have them sit next to each other, i.e. 12! − 2×11! = 399,168,000.

(c) *Solution*: First, we allocate two chairs for Mr. A and Mr. B, thus: M¤¤¤¤M¤¤¤¤¤¤, ¤M¤¤¤¤M¤¤¤¤¤, . . .,¤¤¤¤¤¤M¤¤¤¤M, altogether in 7 possible ways (here we count using our fingers − no fancy formula). Secondly, we seat the people. We have 10! choices for filling the ¤¤...¤ chairs, times 2 choices for how to fill the two reserved chairs (for Mr. A and Mr. B).

Answer: 7×2×10! = 50,803,200.

21.* (a) *Solution*: first select 3 diamonds and two "non-diamonds", then combine these together, in $\binom{13}{3}\binom{39}{2}$ = 211,926 number of ways.

(b) Same logic: $\binom{4}{2}\binom{48}{3}$ = 103,776.

(c) This is slightly more complicated because there is a card which is both an ace and a diamond. The deck must be first divided into four parts, the ace of diamonds (1 card) the rest of the aces (3), the rest of the diamonds (12), the rest of the cards (36). We then consider two cases, either the ace of diamonds is included, or not. The two individual answers are added, since they are mutually incompatible (no "overlap"): $\binom{1}{1}\binom{3}{1}\binom{12}{1}\binom{36}{2}+\binom{1}{0}\binom{3}{2}\binom{12}{2}\binom{36}{1}$ = 29,808.

22.* *Answer*: $\binom{52}{5}\binom{47}{5}\binom{42}{5}\binom{37}{5} = 1.4783\times10^{24}$.

(b) *Answer* (consider dealing the aces and the non-aces separately): $4!\binom{48}{4}\binom{44}{4}\binom{40}{4}\binom{36}{4} = 3.4127 \times 10^{21}$.

(c) *Answer*: $\binom{48}{5}\binom{43}{5}\binom{38}{5}\binom{33}{5} = 1.9636 \times 10^{23}$.

(d) *Answer*: $\binom{4}{2}\binom{48}{3}\binom{45}{5}\binom{40}{5}\binom{35}{5} = 2.7084\times10^{22}$.

(e) *Solution*: the previous answer is correct whether it is Mr. A, B, C or D who gets the 2 aces (due to symmetry), all we have to do is to add the four (identical) numbers, because the four corresponding sets cannot overlap, i.e. are mutually incompatible or exclusive).

Answer: $4\times2.7084\times10^{22}= 1.0834\times10^{23}$.

(f) *Answer*: $\binom{4}{2}\binom{48}{3}\binom{47}{5}\binom{42}{5}\binom{37}{5} = 5.9027\times10^{22}$. Note that when computing the probability of this happening, the $\binom{47}{5}\binom{42}{5}\binom{37}{5}$ part cancels out (we can effectively deal 5 cards to him and stop).

(g) *Solution*: if he is the third player to be dealt his cards, we can either do this to long and impractical way (taking into account how many aces have been dealt to Mr. A and Mr. B), thus: $=5.9027\times10^{22}$, or be smart and argue that, due to the symmetry of the experiment, the answer must be the same as for Mr. A.

23.* (a) *Solution*: first we choose the value which should be represented twice and the three values to go as singles: $\binom{6}{1}\binom{5}{3}$, then we decide how to place the 5 selected numbers in the five blank boxes, which can be done in $\binom{5}{2,1,1,1}$ ways (equal to the number of aabcd permutations).

Answer: $\binom{6}{1}\binom{5}{3}\binom{5}{2,1,1,1} = 3600$.

(b) The same logic gives: $\binom{6}{2}\binom{4}{1}\binom{5}{2,2,1} = 1800$.

(c) $\binom{6}{1}\binom{5}{2}\binom{5}{3,1,1} = 1200$.

(d) $\binom{6}{1}\binom{5}{1}\binom{5}{3,2}$ = 300.

(e) $\binom{6}{1}\binom{5}{1}\binom{5}{4,1}$ = 150.

(f) $\binom{6}{1}\binom{5}{5}$ = 6.

(g) *Solution*: we again fill the empty boxes, one by one, avoiding any duplication: 6×5×4×3×2 = 720.

Note that all these answers properly add up to 7776 = 6^5 (check).

Exercises 3.2

9.* *Solution*: this is a roll-of-a-die type of experiment (this time we roll 4 times − once for each customer − and the die is 3-sided − one side for each company). The sample space will thus consist of 3^4 equally likely possibilities, each looking like this: 1 3 2 1. How many of these contain all three numbers?

To achieve that, we obviously need one duplicate and two singles. There are 3 ways to decide which company gets two customers. Once this decision has been made (say 1 2 2 3), we simply permute the symbols (getting $\binom{4}{2,1,1}$ distinct "words").

Answer: $3\binom{4}{2,1,1}/3^4$ = 4/9 = 44.44 %.

10.* *Solution*: the experiment is in principle identical to rolling a 9-sided die (there are nine floors to be chosen from, exclude the main floor!) six times (once for each person − this corresponds to selecting his/her floor). The sample space thus consists of 9^6 equally likely outcomes (each looking like this: 2 4 8 6 9 4 − ordered selection, repetition allowed). Out of these, only 9×8 ×7×6×5×4 = P_9^6 consist of all distinct floors.

Answer: $P_9^6/9^6$ = 11.38 %.

11.* *Solution*: let's introduce *A*: "Jim and Joe sit together", *B*: "Jim and Tom sit together", *C*: "Joe and Tom sit together". We need $P(A\cup B\cup C) = P(A) + P(B) + P(C) - P(A\cap B) - P(A\cap C) - P(B\cap C) + P(A\cap B\cap C)$. There are 9! arrangements of the boys, 2×8! meet condition *A* (same with *B* and *C*), 2×7! meet both *A* and *B* (same with $A\cap C$ and $B\cap C$), none will meet all three.

Answer: 3×(2×8!)/9! − 3×(2×7!)/9! = 58.33 %.

12.* (a) *Solution*: This, in principle, is the same as choosing 6 different floors in an elevator.

Answer: $P_{365}^{10}/365^{10}$ = 88.31 %.

(b) *Solution*: this is similar to the exercise 9* where we needed exactly one duplicate. By a similar logic, there are 365 ways to choose the date of the duplication, $\binom{10}{2}$ ways of placing these into 2 of the 10 empty slots, and P_{364}^{8} of filling out the remaining 8 slot with distinct birth dates.

Answer: $365\binom{10}{2}P_{364}^{8}/365^{10}$ = 11.16 %.

These two answers account for 99.47 % of the total probability. Two or three duplicates, and perhaps one triplicate would most likely take care of the rest; try it!

13.* *Solution*: the sample space will be the same, but the individual probabilities will no longer be identical; they will now equal to $(1/2)^i(1/16)^{6-i}$ where i is how many times 4 appears in the selection (2 4 8 6 9 4 will have the probability of $(1/2)^2(1/16)^4$, etc.). We have to single out the outcomes with all six floors different and add their probabilities. Luckily, there are only two types of these outcomes: (i) those without any 4 − we have P_8^6 of these, each having the probability of $(1/16)^6$, and (ii) those with a single 4 − there are $6\times P_5^8$ of these, each having the probability of $(1/2)(1/16)^5$.

Answer: $P_8^6(1/16)^6 + 6P_8^5(1/2)(1/16)^5$ = 2.04 % (the probability is a lot smaller now).

14.* There are C_{50}^{10} ways of choosing the test sample, and C_{46}^{10} ways of choosing all good parts so the probability is $C_{46}^{10}/C_{50}^{10} = \frac{46!40!10!}{36!10!50!} = \frac{40\cdot 39\cdot 38\cdot 37}{50\cdot 49\cdot 48\cdot 47} = 0.396$.

Using almost identical calculations a company can decide on how many bad units they will allow in a shipment and design a testing program with a given probability of success.

Exercises 4.1

9.* *Solution*: . . . = $P(A\cup B)\times(1 - P(C\cup D))$ = (0.1 + 0.2 − 0.02)×(1 − 0.3 − 0.4 + 0.12) = 11.76 %.

10.* The sample space of the complete experiment, including probabilities (c − crooked die, r − regular die), is

$r6\ 6$	0.9(1/6)(1/6)	■○
$r6\ \overline{6}$	0.9(1/6)(5/6)	■
$r\overline{6}\ 6$	0.9(5/6)(1/6)	○
$r\overline{6}\ \overline{6}$	0.9(5/6)(5/6)	
$c6\ 6$	0.1(1/2)(1/2)	■○
$c6\ \overline{6}$	0.1(1/2)(1/2)	■

$c\bar{6}\ 6$ 0.1(1/2)(1/2)   ○
$c\bar{6}\ \bar{6}$ 0.1(1/2)(1/2).

(a) *Solution*: in our sample space we mark off the simple events contributing to S_1 (by ■) and to S_2 (by ○) and compute $P(S_1 \cap S_2)/P(S_1)$ (by adding the corresponding probabilities).

Answer: 25 %.

(b) Let us check it out: $P(S_1 \cap S_2) \overset{?}{=} P(S_1)P(S_2)$.

Solution: 18/360 (=1/20) $\neq$ $(72/360)^2$ (=1/25).

Answer: No.

(c) *Answer*: 50 %.

11.* *Solution*: define A: no 8 at any place, B: no 4. We need $P(\bar{A} \cap \bar{B})$ (at least one 8 and at least one 4) = $P(\overline{A \cup B})$ (De Morgan's law) = $1 - P(A \cup B)$ = $1 - P(A) - P(B) + P(A \cap B)$. Clearly $A = A_1 \cap A_2 \cap \ldots \cap A_5$, where A_1: "no 8 in the first place", A_2: "no 8 in the second place", etc. $A_1, A_2, \ldots, A_5$ are mutually independent (selecting a random 5 digit number is like rolling an 10-sided die five times), thus $P(A) = P(A_1)P(A_2)\ldots P(A_5) = (9/10)^5$.

Similarly, $P(B) = (9/10)^5$. Now, $A \cap B \equiv C_1 \cap C_2 \cap \ldots \cap C_5$ where C_1: not 8 nor 4 in the first spot, C_2: not 8 nor 4 in the second, etc.; these of course are also independent, which implies $P(A \cap B) = (8/10)^5$.

Answer: $1 - 2(9/10)^5 + (8/10)^5 = 14.67$ %.

12.* *Solution*: we partition the sample space according to the position at which 8 appears for the first time: $B_1, B_2, \ldots, B_5$, plus B_0 (which means there is no 8). Verify that this is a partition. Now, if A is the event of our question (8 followed by a 4), we can apply the formula of total probability thus: $P(A) = P(A/B_1)P(B_1) + P(A/B_2)P(B_2) + P(A/B_3)P(B_3) + P(A/B_4)P(B_4) + P(A/B_5)P(B_5) + P(A/B_0)P(B_0)$. Individually, we deal with these in the following manner (we use the third term as an example): $P(B_3) = (9/10)^2\,(1/10)$ (no 8 in the first slot, no 8 in the second, 8 in the third, and anything after that; then multiply due to independence), $P(A/B_3) = 1-(9/10)^2$ (given the first 8 is in the third slot, get at least one 4 after; easier through complement: $1 - P$(no 4 in the last two slots)).

Answer: $P(A) = [1 - (9/10)^4](1/10) + [1 - (9/10)^3](9/10)(1/10) + [1 - (9/10)^2](9/10)^2(1/10) + [1 - (9/10)](9/10)^3(1/10) + 0(9/10)^4(1/10) = 8.146$ %.

13.* (a) *Solution*: using C for "criminal", I for "innocent" R for "red light flashing" and G for "green", we have the following sample space:

CR (1/10)(9/10)=0.090   ⋎ ○
CG (1/10)(1/10)=0.010
IR (9/10)(1/20)=0.045   ⋎
IG (9/10)(19/20)=0.855

Answer: $P(C/R) = 0.090/(0.090+0.045) = 2/3$ (far from certain!).

(b) A simple event consists now of a complete record of these tests (the sample space has of 2^{10} of these), e.g. *RGGRGGRGGG*. Assuming that the first item represents the criminal (the sample space must "know" who the criminal is) we can assign probabilities by simply multiplying since the tests are done independently of each other. Thus, the simple event above will have the probability of $0.9\times0.95^2\times0.05\times0.95^2\times0.05\times0.95^3$, etc. Since only one test resulted in *R*, the only simple events of relevance (the idea of a "reduced" sample space) are:

RGGGGGGGGG 0.9×0.95^9
GRGGGGGGGG $0.1\times0.95^8\times0.05$
....
GGGGGGGGGR $0.1\times0.95^8\times0.05$

Given that it was one of these outcomes, what is the probability it was actually the first one?

Answer: $(0.9\times0.95^9)/(0.9\times0.95^9 + 9\times0.1\times0.95^8\times0.05) = 95$ % (now we are a lot more certain − still not 100 % though!).

14.* *Solution* (using an obvious notation): $P(A\cup B) = P(A) + P(B) - P(A\cap B) = 1/4 + 2/5 - 1/10 = 55$ %.

Alternately: $P(A\cup B) = 1 - P(\overline{A\cup B}) = 1 - P(\overline{A}\cap\overline{B}) = 1 - (3/4)(3/5) = 55$ % (replacing *P*(at least one hit) by 1 − *P*(all misses)).

15.* *Solution*: . . . $= P(A\cap\overline{B}) + P(C) - P(A\cap\overline{B}\cap C) = 0.25\times0.65+0.45 - 0.25\times0.65\times0.45 = 53.94$ %.

16.* *Solution*: (a) Introduce a partition A_0, A_1, A_2 according to how many heads are obtained. If *B* stands for "getting fewer than 5 dots", the total probability formula gives: $P(B) = P(A_0)P(B/A_0)+P(A_1)P(B/A_1)+P(A_2)P(B/A_2) = (1/4)\cdot1+(2/4)\cdot(4/6)+(1/4)\cdot(6/36) = 62.5$ %.

The probabilities of A_0, A_1, and A_2 followed from the sample space of two flips: {HH, HT, TH, TT}; the conditional probabilities are clear for $P(B/A_0)$ and $P(B/A_1)$, $P(B/A_2)$ requires going back to 36 outcomes of two rolls of a die and counting those having a total less than 5: {1 1, 1 2, 1 3, 2 1, 2 2, 3 1}.

(b) We are given the outcome of the second stage to guess at the outcome of the first stage. We need the Bayes's rule, and the following (simplified) sample space:

0 3	(1/4)·0	˅
0 $\overline{3}$	(1/4)·1	
1 3	(1/2)·(1/6)	˅ ○
1 $\overline{3}$	(1/2)·(5/6)	
2 3	(1/4)·(2/36)	˅
2 $\overline{3}$	(1/4)·(34/36),	

where the first entry is the number of heads, and the second one is the result of rolling the die, simplified to tell us only whether the total dots equaled 3, or did not ($\overline{3}$).

Answer:$P(\mathbf{1}/\mathbf{3})$ = (1/12)/(1/12+1/72) = 85.71 %.

17.* *Solution*: let's work it out. The first probability can be computed as 1 − P(no sixes in 4 rolls) = 1 − $(5/6)^4$ (due to independence of the individual rolls) = 51.77 %. The second probability, similarly, as 1 − P(no double six in 24 rolls of a pair) = 1− $(35/36)^{24}$ = 49.14 % (only one outcome out of 36 results in a double six).

Answer: getting at least one 6 in four rolls is more likely.

18.* We can visualize the experiment done sequentially with you being the first player and your partner the second one (even if the cards were actually dealt in a different order, this cannot change probabilities). The answer is a natural conditional probability, i.e. the actual condition (event) is decided in the first stage (consider it completed accordingly). The second stage then consists of dealing 13 cards out of 39, with 3 aces remaining.

Answer: $\binom{3}{3}\binom{36}{10}/\binom{39}{13}$ = 3.129 %.

19.* *Solution*: again, a roll-of-a-die type of experiment (10 sides, 5 rolls). The question is in principle identical to rolling a die to get at least one six. This is easier through the corresponding complement.

Answer: 1 − $(9/10)^5$ = 40.95 %.

29.* The counsel for the defense argued as follows: Suppose, for example, there are 5,000,000 couples in the Los Angeles area and the probability that a randomly chosen couple fits the witnesses' description is 1/12,000,000. Then the probability that there are two such couples, given that there is at least one, is not at all small. Find this probability. (The California Supreme Court overturned the initial guilty verdict.)

Exercises 5.1

2.* (c) *Answer*: The chance the prize is behind the door originally chosen by the contestant is 1/10 and remains 1/10. The chance it's behind one of the nine other unopened doors is 9/10. Since the host opens seven of these nine other unopened doors, the 9/10 probability that it's behind one of them is divided between two of these nine doors. So the contestant should switch to one of these two. Doing so raises his probability of winning from 1/10 to one-half of 9/10 or 45 %.

Section 6

Petersburg paradox. The problem with this Casino is that it is not quite clear, what is "fair". For example, the situation T = 20 is so improbable that practically it never occurs in the life-time of a person. Therefore, for any practical reason, one has not to worry about large values of T. This, as well as the finiteness of money resources is the reason, why Casinos do not have to worry about the following bullet proof *martingale strategy* in roulette. Bet c dollars on red. If you win, stop, if you lose, bet $2c$ dollars on red. If you win, stop. If you lose, bet $4c$ dollars on red. Keep doubling the bet. Eventually after n steps, red will occur and you will win $2^n c - (c + 2c + \ldots + 2^{n-1}c) = c$ dollars. In his book "The Newcomes", W.M. Thackeray remarks "You have not played as yet? Do not do so; above all avoid a martingale if you do." This was a good advice.

"A common gamblers' fallacy called "*the doctrine of the maturity of the chances*" (or "Monte Carlo fallacy") falsely assumes that each play in a game of chance is not independent of the others and that a series of outcomes of one sort should be balanced in the short run by other possibilities. A number of "systems" have been invented by gamblers based largely on this fallacy; casino operators are happy to encourage the use of such systems and to exploit any gambler's neglect of the strict rules of probability and independent plays." (*Encyclopedia Britannica*).

Not only do betting systems fail to beat casino games with a house advantage, they can't even dent it. Roulette balls and dice simply have no memory. Every spin in roulette and every toss in craps are independent of all past events. In the short run you can fool yourself into thinking a betting system works, by risking a lot to win a little. However, in the long run no betting system can withstand the test of time.

"No one can possibly win at roulette unless he steals money from the table while the croupier isn't looking." (Albert Einstein).
(You can read the article http://plus.maths.org/issue29/features/haigh/)

But on the other hand "Certainly the game is rigged. Don't let that stop you; if you don't bet, you can't win." (Robert Heinlein).

The decision is up to you.

Appendix 1

Probability and Statistics Pre-course Survey

(http://www.dartmouth.edu/~chance/course/Evaluation/pre-course_survey.html)

The purpose of this survey is to indicate what you already know and think about probability and statistics.

Part I

There are a series of statements concerning beliefs or attitudes about probability, statistics and mathematics. Following each statement is an "agreement" scale which ranges from 1 to 5, as shown below.

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree, nor Disagree	Agree	Strongly Agree

If you strongly agree with a particular statement, circle the number 5 on the scale. If you strongly disagree with the statement, circle the number 1.

---

1. I often use statistical information in forming my opinions or making decisions. 1 2 3 4 5
2. To be an intelligent consumer, it is necessary to know something about statistics. 1 2 3 4 5
3. Because it is easy to lie with statistics, I don't trust them at all. 1 2 3 4 5
4. Understanding probability and statistics is becoming increasingly important in our society, and may become as essential as being able to add and subtract. 1 2 3 4 5
5. Given the chance, I would like to learn more about probability and statistics. 1 2 3 4 5
6. You must be good at mathematics to understand basic statistical concepts. 1 2 3 4 5
7. When buying a new car, asking a few friends about problems they have had with their cars is preferable to consulting an owner satisfaction survey in a consumer magazine. 1 2 3 4 5
8. Statements about probability (such as what the odds are of winning a lottery) seem very clear to me. 1 2 3 4 5

9. I can understand almost all of the statistical terms that I encounter in newspapers or on television. 1 2 3 4 5

10. I could easily explain how an opinion poll works. 1 2 3 4 5

Part II

1. A small object was weighed on the same scale separately by nine students in a science class. The weights (in grams) recorded by each student are shown below.

6.2 6.0 6.0 15.3 6.1 6.3 6.2 6.15 6.2

The students want to determine as accurately as they can the actual weight of this object. Of the following methods, which would you recommend they use?

_____a. Use the most common number, which is 6.2.

_____b. Use the 6.15 since it is the most accurate weighing.

_____c. Add up the 9 numbers and divide by 9.

_____d. Throw out the 15.3, add up the other 8 numbers and divide by 8.

2. A marketing research company was asked to determine how much money teenagers (ages 13–19) spend on recorded music (cassette tapes, CDs and records). The company randomly selected 80 malls located around the country. A field researcher stood in a central location in the mall and asked passers-by, which appeared to be the appropriate age to fill out a questionnaire. A total of 2,050 questionnaires were completed by teenagers. On the basis of this survey, the research company reported that the average teenager in this country spends $155 each year on recorded music.

Listed below are several statements concerning this survey. Place a check by every statement that you agree with.

_____a. The average is based on teenagers' estimates of what they spend and therefore could be quite different from what teenagers actually spend.

_____b. They should have done the survey at more than 80 malls if they wanted an average based on teenagers throughout the country.

_____c. The sample of 2,050 teenagers is too small to permit drawing conclusions about the entire country.

_____d. They should have asked teenagers coming out of music stores.

_____e. The average could be a poor estimate of the spending of all teenagers given that teenagers were not randomly chosen to fill out the questionnaire.

_____f. The average could be a poor estimate of the spending of all teenagers given that only teenagers in malls were sampled.

_____g. Calculating an average in this case is inappropriate since there is a lot of variation in how much teenagers spend.

_____h. I don't agree with any of these statements.

3. Which of the following sequences is most likely to result from flipping a fair coin 5 times?

_____a. H H H T T

_____b. T H H T H

_____c. T H T T T

_____d. H T H T H

_____e. All four sequences are equally likely.

4. Select the alternative below that is the best explanation for the answer you gave for the item above.

_____a. Since the coin is fair, you ought to get roughly equal numbers of heads and tails.

_____b. Since coin flipping is random, the coin ought to alternate frequently between landing heads and tails.

_____c. Any of the sequences could occur.

_____d. If you repeatedly flipped a coin five times, each of these sequences would occur about as often as any other sequence.

_____e. If you get a couple of heads in a row, the probability of a tails on the next flip increases.

_____f. Every sequence of five flips has exactly the same probability of occurring.

5. Listed below are the same sequences of Heads and Tails that were listed in Item 3. Which of the sequences is least likely to result from flipping a fair coin 5 times?

_____a. H H H T T

_____b. T H H T H

_____c. T H T T T

_____d. H T H T H

_____e. All four sequences are equally unlikely.

6. The Caldwells want to buy a new car, and they have narrowed their choices to a Buick or an Oldsmobile. They first consulted an issue of Consumer Reports, which compared rates of repairs for various cars. Records of repairs done on 400 cars of each type showed somewhat fewer mechanical problems with the Buick than with the Oldsmobile.

The Caldwells then talked to three friends, two Oldsmobile owners, and one former Buick owner. Both Oldsmobile owners reported having a few mechanical problems, but nothing major.

The Buick owner, however, exploded when asked how he likes his car: “First, the fuel injection went out – $250 bucks. Next, I started having trouble with the rear end and had to replace it. I finally decided to sell it after the transmission went. I'd never buy another Buick.”

The Caldwells want to buy the car that is less likely to require major repair work. Given what they currently know, which car would you recommend that they buy?

_____a. I would recommend that they buy the Oldsmobile, primarily because of all the trouble their friend had with his Buick. Since they haven't heard similar horror stories about the Oldsmobile, they should go with it.

_____b. I would recommend that they buy the Buick in spite of their friend's bad experience. That is just one case, while the information reported in Consumer Reports is based on many cases. And according to that data, the Buick is somewhat less likely to require repairs.

_____c. I would tell them that it didn't matter which car they bought. Even though one of the models might be more likely than the other to require repairs, they could still, just by chance, get stuck with a particular car that would need a lot of repairs. They may as well toss a coin to decide.

7. Half of all newborns are girls and half are boys. Hospital A records an average of 50 births a day. Hospital B records an average of 10 births a day. On a particular day, which hospital is more likely to record 80 % or more female births?

_____a. Hospital A (with 50 births a day).

_____b. Hospital B (with 10 births a day).

_____c. The two hospitals are equally likely to record such an event.

8. "Megabucks" is a weekly lottery played in many states. The numbers 1 through 36 are placed into a container. Six numbers are randomly drawn out, without replacement. In order to win, a player must correctly predict all 6 numbers. The drawing is conducted once a week, each time beginning with the numbers 1 through 36.

The following question about the lottery appeared in The New York Times (May 22, 1990). Are your odds of winning the lottery better if you play the same numbers week after week or if you change the numbers every week? What do you think?

______a. I think the odds are better if you play the same numbers week after week.

______b. I think the odds are better if you change the numbers every week.

______c. I think the odds are the same for each strategy.

9. For one month, 500 elementary students kept a daily record of the hours they spent watching television. The average number of hours per week spent watching television was 28. The researchers conducting the study also obtained report cards for each of the students. They found that the students who did well in school spent less time watching television than those students who did poorly.

Listed below are several possible statements concerning the results of this research. Place a check by every statement that you agree with.

_____a. The sample of 500 is too small to permit drawing conclusions.

_____b. If a student decreased the amount of time spent watching television, his or her performance in school would improve.

_____c. Even though students who did well watched less television, this doesn't necessarily mean that watching television hurts school performance.

_____d. One month is not a long enough period of time to estimate how many hours the students really spend watching television.

_____e. The research demonstrates that watching television causes poorer performance in school.

_____f. I don't agree with any of these statements.

10. An experiment is conducted to test the efficiency of a new drug on curing a disease. The experiment is designed so that the number of patients who are cured using the new drug is compared to the number of patients who are cured using the current treatment. The percentage of patients who are cured using the current treatment is 50 % and 65 % are cured who have used the new drug. A *P-value* of 5 % (0.05) is given as an indication of the *statistical significance* of these results.

The *P-value* tells you:

______a. There is a 5 % chance that the new drug is more effective than the current treatment.

______b. If the current treatment and the new drug were equally effective, then 5 % of the times we conducted the experiment we would observe a difference as big, or bigger than the 15 % we observed here.

______c. There is a 5 % chance that the new drug is better than the current treatment by at least 15 %

11. Gallup* reports the results of a poll that shows that 58 % of a random sample of adult Americans approve of President Clinton's performance as president. The report says that the *margin of error* is 3 %.

What does this *margin of error* mean?

______a. One can be 95 % "confident" that between 55 % and 61 % of all adult Americans approve of the President's performance.

______b. One can be sure that between 55 % and 61 % of all adult Americans approve of the President's performance.

______c. The sample percentage of 58 % could be off by 3 % in either direction due to inaccuracies in the survey process.

______d. There is a 3 % chance that the percentage of 58 is an inaccurate estimate of the population of all Americans who approve of President Clinton's performance as president.

*Gallup Poll − a sampling by the American Institute of Public Opinion or its British counterpart of the views of a representative cross section of the population, used as a means of forecasting voting.

Etymology: named after George Horace Gallop (1901-1984), US statistician.

Appendix 2

Sets

A **set** is a collection of things that are called the elements of the set. Recall that the notion of a set is not defined. We just use equivalent names (collection, family, or array) and hope that, with practice, the right intuition will develop. The elements can be any kind of entity: numbers, people, poems, blueberries, points, lines, and so on, endlessly. Upper case letters are usually used to denote sets. If the set S includes some element denoted by x, then we say x belongs to S and write $x \in S$. If x does not belong to S, then we write $x \notin S$.

There are two ways of describing a set, either by a *list* or by a *rule*.

Example. If S is the set of numbers shown by a conventional die, then the *rule* is that S comprises the integers lying between 1 and 6 inclusive. This may be written formally as follows: $S = \{x\colon 1 \le x \le 6$ and x is an integer$\}$. Alternatively, S may be given as a *list:* $S = \{1, 2, 3, 4, 5, 6\}$.

We distinguish a number 2 from a set $\{2\}$. Arithmetic operations apply to numbers, while for sets we use set operations.

One important special case arises when the set is empty (there are no elements belonging to it); for example, consider the set of elephants playing football on Mars. We denote the empty set by ϕ. The **empty set** (or null set) ϕ contains no element.

If S and T are two sets such that every element of S is also an element of T, then we say that T **includes** S and write either $S \subset T$ or $T \supset S$. If $S \subset T$ and $T \subset S$, then S and T are said to be **equal** and we write $S = T$.

Combining Sets. Given any nonempty set, we can divide it up, and given any two sets, we can join them together and so on.

Let A and B be sets. Their ***union****, denoted by $A \cup B$, is the set of elements that are in A or B, or in both. Their* ***intersection****, denoted by $A \cap B$, is the set of elements in both A and B.*

Note that the union may be referred to as the *join* or *sum*; the intersection may be referred to as the *meet* or *product*.

Example (*probabilistic*). In Kazakhstan the profits of wheat producers critically depend on weather conditions during summer (a lot of rain results in an abundant crop; there is always a lot of sun) and the weather during the harvesting time (just three rainy days in September are enough for the quality of wheat to plunge). Let us denote by S the event that the weather is good in summer and by H the event that it is good during the harvesting time. The farmers are happy when $P(S \cap H)$ is high.

We use the obvious notation $\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup ... \cup A_n$ and $\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap ... \cap A_n$.

Clearly, $A \cap B = B \cap A$ and $A \cup B = B \cup A$; $A \cup (B \cup C) = (A \cup B) \cup C$ and $A \cap (B \cap C) = (A \cap B) \cap C$.

But we must be careful when making more intricate combinations of larger numbers of sets. For example, we cannot write down simply $A \cup B \cap C$; this is not well defined because in general $(A \cup B) \cap C \neq A \cup (B \cap C)$.

Intersection is distributive over union: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ (try to prove it through Venn diagrams).

Similarly, union is distributive over intersection: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ (try to proof). This is unlike the regular algebra of adding and multiplying numbers (addition is not distributive over multiplication: $a+(b\times c) \neq (a+b)\times(a+c)$), obviously the two algebras (the second is Boolean Algebra (algebra of sets)) "behave" differently.

Note the following

If $A \cap B = \phi$ (i.e. there is no overlap between the two sets, they have no elements in common)*, then A and B are said to be **disjoint**.*

We can also remove bits of sets, giving rise to set differences, as follows.

*Let A and B be sets. That part of A that is not also in B is denoted by $A \backslash B$, called the **difference** of A and B. Elements that are in A or B but not both, comprise the **symmetric difference**, denoted by $A \Delta B$.*

Finally, we can combine sets in a more complicated way by taking elements in pairs, one from each set.

*Let A and B be sets, and let $C = \{(a, b): a \in A, b \in B\}$ be the set of ordered pairs of elements from A and B. Then C is called the **direct product** of A and B and denoted by $A \times B$.*

Example. Let A be the interval $[0, a]$ of the x-axis, and B the interval $[0, b]$ of the y-axis. Then $C = A \times B$ is the rectangle of base a and height b with its lower left vertex at the origin, when $a, b > 0$.

In probability problems, all sets of interest A lie in a universal set Ω (the universal set Ω contains all needed elements), so that $A \subset \Omega$ for all A. That part of Ω that is not in A is called the ***complement*** of A and denoted by A^c or $\overline{A}$.

Formally, $A^c = \Omega \backslash A = \{x: x \in \Omega, x \notin A\}$.

Obviously, from the diagram or by consideration of the elements

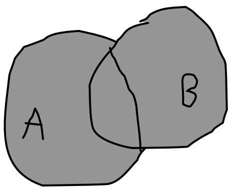
$A \cup A^c = \Omega$, $A \cap A^c = \phi$, $(A^c)^c = A$.

If $A_j \cap A_k = \phi$ for $j \neq k$, $1 \le k \le n$, $1 \le j \le n$, and $\bigcup_{i=1}^{n} A_i = \Omega$, then the collection ($A_i$, $1 \le i \le n$) is said to form a ***partition*** *of* Ω.

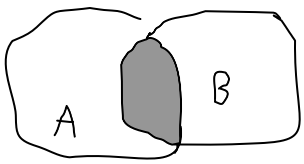
De Morgan's Laws: $\overline{A \cap B} = \overline{A} \cup \overline{B}$, and $\overline{A \cup B} = \overline{A} \cap \overline{B}$.

These can be verified easily by Venn diagrams; both can be extended to any number of events, for example: $\overline{A \cap B \cap C} = \overline{A} \cup \overline{B} \cup \overline{C}$, $\overline{A \cup B \cup C} = \overline{A} \cap \overline{B} \cap \overline{C}$.

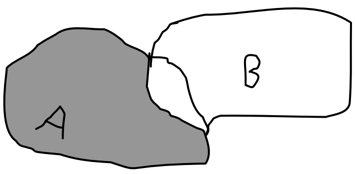
Usually, **Venn diagrams** are used to represent sets. These provide very expressive pictures, which are often so clear that they make algebra redundant:



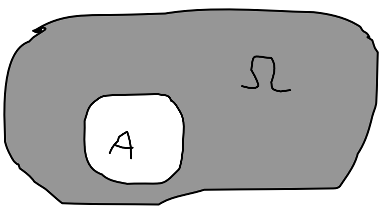
• union: $A \cup B = A$ or B or both;



• intersection: $A \cap B = A$ and B;



• difference of A and B (relative complement): $A \backslash B$ = in A but not in B;



• (absolute) complement: A^c = not A = $\Omega \backslash A$.

http://www.combinatorics.org/Surveys/ds5/VennEJC.html − an article on Venn diagrams with history and many nice pictures.

Appendix 3
Review of Elementary Mathematical Prerequisites

Mathematics is not primarily a matter of plugging numbers into formulae and performing rote computations. It is a way of questioning and thinking that may be unfamiliar to many of us, but is available to almost all of us.
John Allen Paulos

It is difficult to make progress in any branch of mathematics without using the ideas and notation of sets and functions. We therefore give a brief synopsis of what we need here for completeness (as to sets see Appendix 2), although it is likely that the reader will already be familiar with this.

Notation. We use a good deal of familiar standard mathematical notation in this text. The basic notation is set out below. More specialized notation for probability theory is introduced as required throughout the text.

We list some fundamental notation:

e − the base of natural logarithms; Euler's number;
$\ln x$ − the logarithm of x to base e;
$\log_2 x$ − the logarithm of x to base 2;
π − the ratio of circumference to diameter for a circle;
Ω − the universal set;
$n! = n\times(n-1)\times\ldots 3\times 2\times 1$ − factorial n; note that $0! = 1$, by convention;
$|x|$ − modulus or absolute value of x;
$[x]$ − the integer part of x;
R − the real line;
Z − the integers;
$\sum_{i=1}^{n} a_i = a_1 + a_2 + \ldots + a_n$ − summation symbol;
$\prod_{i=1}^{n} a_i = a_1 \times a_2 \times \ldots \times a_n$ − product symbol
(in which the three points mean "and so on").

What do you think grows faster (as n increases): e^n, $n!$ or n^n?

Size. When sets have a finite number of elements, it is often useful to consider the number of elements they contain; this is called their **size** or **cardinality**.

For any set A, we denote its size by $|A|$; it is easy to see that size has the following properties.

If sets A and B are disjoint, then $|A\cup B| = |A| + |B|$, and more generally, when A and B are not necessarily disjoint,

$$|A \cup B| + |A \cap B| = |A| + |B|. \quad \text{(A3.1)}$$

Naturally, $|\phi| = 0$, and if $A \subset B$, then $|A| \le |B|$.

Finally, for the product of two finite sets $A \times B$, we have $|A \times B| = |A| \times |B|$.

When sets are infinite, a great deal more care and subtlety is required in dealing with the idea of size.

However, we intuitively see that we can consider the length of subsets of a line, or areas of sets in a plane, or volumes in space, and so on. It is easy to see that if A and B are two subsets of a line, with lengths $|A|$ and $|B|$, respectively, then in general $|A \cup B| + |A \cap B| = |A| + |B|$. Therefore $|A \cup B| = |A| + |B|$ when $A \cap B = \phi$. We can define the product of two such sets as a set in the plane with area $|A \times B|$, which satisfies the well-known elementary rule for areas and lengths $|A \times B| = |A| \times |B|$ and is thus consistent with the finite case above. Volumes and sets in higher dimensions satisfy similar rules.

Functions. Suppose we have sets A and B, and a rule that assigns to each element a in A a unique element b in B. Then this rule is said to define a **function** from A to B; for the corresponding elements, we write $b = f(a)$.

Here the symbol $f(\cdot)$ denotes the rule or function; often we just call it f. The set A is called the *domain* of f, and the set of elements in B that can be written as $f(a)$ for some a is called the *range* of f; we may denote the range by R.

Anyone who has a calculator is familiar with the idea of a function. For any function key, the calculator will supply $f(x)$ if x is in the domain of the function; otherwise, it says “error”.

If f is a function from A to B, we can look at any b in the range R of f and see how it arose from A. This defines a rule assigning elements of A to each element of R, so if the rule assigns a unique element a to each b this defines a function from R to A. It is called the **inverse function** and is denoted by $f^{-1}(\cdot)$: $a = f^{-1}(b)$.

Example. Let $A \subset \Omega$ and define the following function $I(\cdot)$ on Ω:

$I(\omega) = 1$ if $\omega \in A$,

$I(\omega) = 0$ if $\omega \notin A$.

Then I is a function from Ω to $\{0, 1\}$; it is called the **indicator** of A because by taking the value 1 it indicates that $\omega \in A$. Otherwise, it is zero.

This is about as simple a function as you can imagine, but it is surprisingly useful. For example, note that if A is finite you can find its size by summing $I(\omega)$ over all ω: $|A| = \sum_{\omega \in \Omega} I(\omega)$.

Sums. Consider the sum $\sum_{i=1}^{n} a_i = a_1 + a_2 + ... + a_n$.

The variable i is a *dummy variable* or *index of summation*, so any symbol will suffice: $\sum_{r=1}^{n} a_r = \sum_{i=1}^{n} a_i = a_1 + a_2 + \ldots + a_n$.

In general, $\sum_{i=1}^{n}(ca_i + db_i) = c\sum_{i=1}^{n} a_i + d\sum_{i=1}^{n} b_i$.

In particular, $\sum_{i=1}^{n} 1 = n$; $\sum_{i=1}^{n} i = \frac{(n+1)n}{2}$ – the **arithmetic sum**; $\sum_{i=1}^{n} i^2 = \frac{(2n+1)(n+1)n}{6}$; $\sum_{i=1}^{n} i^3 = \left(\sum_{i=1}^{n} i\right)^2 = \frac{(n+1)^2 n^2}{4}$; $\sum_{i=1}^{n} q^i = \frac{1-q^{n+1}}{1-q}$ – the **geometric sum**.

Binomial expansion. You probably all know that

$(x + y)^2 = x^2 + 2xy + y^2$,

$(x + y)^3 = x^3 + 3x^2y + 3xy^2 + y^3$,

. . . .

$(x + y)^n = C_n^0 y^n + C_n^1 y^{n-1}x + C_n^2 y^{n-2}x^2 + \ldots + C_n^{n-1} y\ x^{n-1} + C_n^n x^n$,

the general (last-line) expansion usually written as $(x + y)^n = \sum_{r=0}^{n} C_n^r x^r y^{n-r}$.

We discuss $C_n^r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$ in Part 2 (Combinatorics) in detail.

We have to remember that 0!= 1 by definition so that $C_n^0 = C_n^n = 1$.

A whole row of these coefficients can be easily constructed as follows (**Pascal's triangle**):

$$
\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1 \\
=\binom{5}{0} & & =\binom{5}{1} & & =\binom{5}{2} & & =\binom{5}{2} & & =\binom{5}{1} & & =\binom{5}{0}.
\end{array}
$$

Each number is the sum of the ones on the row above on its immediate left and right. To get the 1's on the edges to work correctly we consider the blanks to be zeros. In symbols $C_n^k = C_{n-1}^{k-1} + C_{n-1}^k$.

Multinomial expansion is an extension of the binomial expansion, having more than 2 terms inside parentheses. We will derive it using three terms, the extension to any other number of terms is then quite obvious.

We want to generalize the well known: $(x+y+z)^2 = x^2+y^2+z^2+2xy+2xz+2yz$ to: $(x + y + z)^n = (x + y + z)(x + y + z) \ldots (x + y + z)$ (n factors) = (distributive law) $xxx...x + yxx...x + \ldots + zzz...z$ (all 3^n n-letter words built out of x, y and z) = (collecting algebraically identical contributions) $x^n + \binom{n}{1} x^{n-1}y - \binom{n}{n-5,3,2} x^{n-5}y^3z^2 + \ldots + z^n$ (the coefficients representing the number of words with the corresponding number of x's, y's and z's) = $\sum_{\substack{i+j+k=n \\ i,j,k \geq 0}} M_n(i,j,k)x^i y^j z^k$, where the summation is over all possible selections of non-negative exponents which add up to n.

We discuss $M_n(i,j,k) = \binom{n}{i,j,k}$ in Part 2 (Combinatorics) in detail.

How many terms are there in this summation? Our formula (2.1) tells us that it should be $\binom{n+2}{2}$ − the three exponents are chosen in the apple-pear-orange like manner. This is because when choosing the x's, y's and z's we don't care about the order, and we are allowed to repeat.

Limits. Often, we have to deal with "infinite sums" (series). A fundamental concept in this context is that of the limit of a sequence.

Definition. Let $(s_n, n \geq 1)$ *be a sequence of real numbers. If there is a number s such that* $|s_n - s|$ *may ultimately always be as small as we please, then s is said to be the* ***limit*** *of the sequence* s_n. *Formally, we write* $\lim_{n\to\infty} s_n = s$ *if and only if for any* $\varepsilon > 0$, *there is a finite* n_0 *such that* $|s_n - s| < \varepsilon$ *for all* $n > n_0$.

Notice that s_n need never actually take the value s, it must just get closer to it in the long run (e.g., let $x_n = n^{-1}$).

Series. Let $(a_r, r \geq 1)$ be a sequence of terms, with partial sums $s_n = \sum_{i=1}^{n} a_i, \quad n \geq 1$.

If s_n has a finite limit s as $n \to \infty$, then the series $\sum_{i=1}^{\infty} a_i$ is said to *converge* with s (we note this as $\sum_{i=1}^{\infty} a_i = s$). Otherwise, it *diverges*. If $\sum_{i=1}^{\infty} |a_i|$ converges, then $\sum_{i=1}^{\infty} a_i$ is said to be *absolutely convergent*.

For example, in the geometric sum above, if $|q| < 1$, then $|q|^n \to 0$ as $n \to \infty$. Hence $\sum_{i=1}^{\infty} q^i = \frac{1}{1-q}$, $|q| < 1$, and the series is absolutely convergent for $|q| < 1$. For example, $\sum_{i=1}^{\infty} (i+1)q^i = \frac{1}{(1-q)^2}$, $|q| < 1$; $\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}$; $\sum_{i=1}^{\infty} \frac{1}{i^4} = \frac{\pi^4}{90}$.

Also for all x, where e is the base of natural logarithms, $\exp x = e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$ (we can consider this as a definition).

An important property of e^x is the *exponential limit theorem*: $(1 + x/n)^n \to e^x$ as $n \to \infty$.

References

1. Bărboianu C. Understanding and calculations the odds. Probability Theory Basics and Calculus Guide for Beginners, with Applications in Games of Chance and Everyday Life. – 2006. – 420 p.

2. Bar-Hillel M. and Falk R. Some teasers concerning conditional probabilities// Cognition. – 1982. – Vol. 11. – P. 109–122.

3. Elias J. Syllabus //www.montana.edu

4. Grinstead C.M., Snell J.L. Introduction to Probability. – 1997. – 510 p.
(available free http://www.freebookcentre.net/Mathematics/Probability-Theory-Books.html)

5. Stirzaker D. Elementary Probability. – 2003. – 520 p.

6. Vrbik J. Lecture Notes. Probability //www.brocku.ca/

7. http://www.random.org/
(offers true random numbers to anyone on the Internet.)

8. http://lib.mexmat.ru/catalogue.php
(books (for reading only)).

9. http://www.dartmouth.edu/~chance/index.html
(The goal of Chance is to make students more informed, critical readers of current news stories that use probability and statistics.)

10. http://www.math.uah.edu/stat/
(The Virtual Laboratories in Probability and Statistics, a set of web-based resources for students and teachers of probability and statistics, where you can run simulations.)

11. www-history.mcs.st-andrews.ac.uk/history/index.html
(historical information about math and mathematicians.)

12.http://oli.web.cmu.edu/openlearning/forstudents/freecourses/statistics
(The Open and Free Full Courses. The Probability and Statistics course is comparable to the full semester course on Statistics taught at Carnegie Mellon University.)

13. http://www.icoachmath.com
(…designed specifically to help students strengthen their math skills through self-paced, guided, and online interactive practice.)

14.http://plus.maths.org/content/teacher-package-statistics-and-probability-theory

15. http://home.ubalt.edu/ntsbarsh/