

# Probability Review

Stephen R. Addison

January 31, 2003

## Introduction

Over the next several class periods we will be reviewing the basic results of probability and relating probability to the second law of thermodynamics.

## Basic Probability

The theory of probability has its origins in gambling. Chevalier de Mere noted that he could make money by offering even odds on throwing one “6” in four rolls of a die. He reasoned (plausibly) that betting on a double “6” in 24 rolls of a pair of dice would be profitable – it wasn’t.

## The Single Die

We’ll adopt a useful device – what’s the probability of getting no sixes on four rolls of a die? The probability of not getting a 6 on one roll is  $5/6$ . Since the rolls are independent of each other we can multiply the individual probabilities to calculate the net probability:

$$P(\text{no 6}|\text{fair die}) = \left(\frac{5}{6}\right) \left(\frac{5}{6}\right) \left(\frac{5}{6}\right) \left(\frac{5}{6}\right) = \left(\frac{625}{1296}\right).$$

We are, of course interested in the opposite result – what is the probability of getting one six in four throws?

$$P(6|4 \text{ throws of a die}) = 1 - \left(\frac{625}{1296}\right) > \left(\frac{1}{2}\right),$$

–the bet is a winner.

$$P(6|\text{fair die}) = \left(\frac{1}{6}\right) / \text{throw}$$

$$P(\text{not } 6|\text{fair die}) = \left(\frac{5}{6}\right) / \text{throw}$$

Notice that I’ve written my probabilities in a particular way, I have written probabilities that are conditional. The vertical line can be read as “given that”. There are several different schools of probability, until recently the dominant schools was the *frequentist, orthodox, or statistical* school. The use of the conditional symbol indicates that the writer is a member of the *Bayesian, or Laplacian* school. Interestingly, it was shown, in 1946, that if the statistical method produces a different result to the Bayesian method, then the statistical method is wrong. I am an adherent of the Bayesian school, I have always been, when I wanted to learn about this sort of thing I picked up Harold Jeffreys’ book on probability. Since I omitted the first chapter, I wasn’t aware that there were other methods that were accepted. I viewed it as Jeffreys simply showing that other methods that could be used were inefficient. It was a surprise to find that people actually used them. In cases where it is obvious what is meant, I won’t always use the conditioning notation — but I always have the background information in mind.

In order to see why the second bet failed, we will take a brief look at sample space.

## Sample Space

A sample space is a list of all possible outcomes of an experiment. The possible outcomes for a pair of dice are:

111111	222222	333333	444444	555555	666666
123456	123456	123456	123456	123456	123456

. Of the 36 possible outcomes, only one is a double 6.

$$P(\text{double 6}|\text{one roll of fair dice}) = \left(\frac{1}{36}\right)$$

$$P(\text{Not double 6}|\text{one roll of fair dice}) = \left(\frac{35}{36}\right)$$

Compounding probabilities multiplicatively,

$$P(\text{no double 6}|24 \text{ rolls of a pair of fair dice}) = \left(\frac{35}{36}\right)^{24} = 0.5086$$

Thus it is no surprise that Chevalier de Mere lost money on this bet.

## Formal definitions of probability

If there are several equally likely, mutually exclusive, and collectively exhaustive outcomes to an experiment, the probability of an event E is given by:

$$P(E|\text{Conditioning Information}) = \frac{\text{Number of outcomes favorable to E}}{\text{Total number of outcomes}}.$$

If the event cannot be broken down into equally likely events – for instance, what's the probability of snow on June 21<sup>st</sup>.

$$P(E|\text{Conditioning Information}) = \frac{\text{Number of successful occurrences of E}}{\text{Number of trials}}.$$

This second definition can be expected to improve with the number of trials.

## The Terminology of Probability

Event not A  $\Rightarrow$  A does not happen

Event A or B  $\Rightarrow$  In an experiment A or B or both occur

A then B  $\Rightarrow$  If in independent successive experiments A occurs in the 1<sup>st</sup> and B occurs in the 2<sup>nd</sup>.

A,B are disjoint events if it is impossible for both of them to occur simultaneously.

## Compounding Probabilities

If A,B are independent successive events or experiments:

$$P(A \text{ then } B) = P(A)P(B),$$

$$P(\text{not } E) = 1 - p(E),$$

and, if A,B are disjoint,

$$P(A \text{ or } B) = P(A) + P(B).$$

The reason that we calculated  $P(\text{no 6}|4 \text{ rolls})$  was that we couldn't compound the probabilities for one six in four rolls.

## Some Examples

### Playing Cards

In a normal deck, there are four suits and 52 cards.

$$P(\text{spades}) = \frac{13}{52} = \frac{1}{4}$$

$$P(\text{king}) = \frac{4}{52} = \frac{1}{13}$$

If you haven't seen the cards, it doesn't matter if some of them were previously dealt. If the cards are exposed and not replaced differences emerge. In such situations, it is best to use the notation for conditional probability so that the conditions are clear.

$$P(\text{pair}|2 \text{ cards are drawn}) = \frac{3}{51}$$

In this case, it doesn't matter what the first card was.

$$P(5 \text{ spades}) = \left(\frac{13}{52}\right) \left(\frac{12}{51}\right) \left(\frac{11}{50}\right) \left(\frac{10}{49}\right) \left(\frac{9}{48}\right) = 0.000495$$

In this case each of the events was independent.

$$P(\text{black card}) = \frac{26}{52}$$

$$P(\text{red ace}) = \frac{2}{52}$$

$$P(\text{black card or red ace}) = \frac{4}{52} + \frac{2}{52} = \frac{7}{13}$$

In the above case, each of the events is disjoint

## Monkeys and Shakespeare

We'll use our new found methods to answer a question posed by Sir James Jeans. Can Monkeys type Hamlet?

We first need to determine the probability of Hamlet appearing in a random stream. As there were no computers in Jeans' era, we'll assume a typewriter. There are approximately one million characters in Hamlet, and assume that there are 44 keys on a typewriter. (We won't worry about the shift key.) Thus the probability of getting any character is  $1/44$ . So

$$P(\text{Hamlet}|\text{Random typing}) = \left(\frac{1}{44}\right) \left(\frac{1}{44}\right) \cdots \left(\frac{1}{44}\right) = \left(\frac{1}{44}\right)^{10^6}.$$

This is one of those numbers for which your calculator provides little help. We can use logarithms,  $\log_{10} 44 = 1.643453$ . Now if  $\log_a x = y$ , then  $a^y = x$ , so  $44 = 10^{1.643453}$ . Thus we can write

$$P(\text{Hamlet}|\text{Random}) = \left(\frac{1}{44}\right)^{10^6} = \frac{1}{(44)^{10^6}}.$$

So

$$P(\text{Hamlet}|\text{Random}) = \frac{1}{(10^{1.643453})^{10^6}} = \frac{1}{10^{1643453}}.$$

We have now calculated the probability of typing Hamlet if we type one million characters at random. The probability can be expected to increase if more characters are typed. The question is, by how much does the probability go up?

We will now work this out, the question is how many monkeys and how many keystrokes. Let's assume that the number of monkeys that has ever lived is  $10^{10}$ . Further let each of them have lived for the age of the universe, that is  $10^{18}$ . Finally, we'll assume that they can hit ten keys per second. (Some good typists can.) Now clearly the probability will go up. How do we calculate it? By a scheme that we have used before. We'll calculate the probability that they do not type Hamlet.

$$P(\text{Hamlet}|\text{Random}) = 10^{-1643453},$$

$$P(\text{No Hamlet}|\text{Random}) = 1 - 10^{-1643453}.$$

With  $10^{29}$  keystrokes, we can write the probability as

$$P(\text{No Hamlet}|\text{Random}) = (1 - 10^{-1643453})^{29}.$$

Some of you might not like this step, and would want to add  $10^6$  keystrokes to either end of the manuscript. Remember that  $2 \times 10^6$  is negligible compared to  $10^{29}$ . To evaluate this probability we use

$$(1 + x)^p = 1 + px + \frac{p(p-1)x^2}{2!} + \dots, \text{ if } |x| < 1.$$

Therefore,

$$P(\text{No Hamlet}|\text{Random}) = 1 - 10^{29}10^{-1643453},$$

and finally,

$$P(\text{Hamlet}|\text{Random}) = 10^{29}10^{-1643453} = 10^{-1643424}.$$

This could be said to be the meaning of never.

## Methods of Counting

If something can be done  $n_1$  ways, and something else can be done in  $n_2$  ways, then the number of ways of doing these things in succession is  $n_1n_2$ . This is called *Fundamental Principle of counting*.

## Permutations

How many ways can 13 objects be arranged (or rearranged)?

$$13 \times 12 \times 11 \times 10 \times \dots \times 1 = 13! = 6,227,020,800$$

The question that we have really asked here is: given  $n$  objects, how many ways can we permute them? We denote the number of permutations of  $n$  things taken  $n$  at a time as  $nPn$ ,  $P(n, n)$ , or  $P_n^n$ . To clarify this think of  $n$  people sitting in  $n$  chairs. There are

$$n(n-1)(n-2)\dots 1 = n!$$

ways of arranging them. Thus,

$$P(n, n) = n!$$

Now consider  $n$  chairs and  $r$  people, where  $r < n$ . Then there are

$n$  ways of filling the first

$n-1$  ways of filling the second

$n-2$  ways of filling the third

So how many ways are there of filling chair  $r$ ? To determine this consider the following

$$n = n - 1 + 1$$

$$n - 1 = n - 2 + 1$$

$$n - 2 = n - 3 + 1$$

For the  $r^{th}$  this gives us the answer  $n - r + 1$ .

Thus the number of permutations of  $n$  things taken  $r$  at a time is

$$P(n, r) = n(n-1)(n-2)\dots (n-r+1).$$

We can rewrite this as

$$P(n, r) = n(n-1)(n-2)\dots (n-r+1) \times \frac{(n-r)!}{(n-r)!} = \frac{n!}{(n-r)!}$$

resulting in

$$P(n, r) = \frac{n!}{(n-r)!}$$

This calculation assumes that the order of the objects is important. It can be used to answer such questions as: How many different finishes among the first three places can occur in an eight horse race. The answer is

$$P(8, 3) = \frac{8!}{(8-3)!} = 336.$$

It can't answer questions like how many distinct foursomes can be formed from seven golfers. Here the order of the golfers on the scorecards doesn't matter, for this we need *combinations*.

## Combinations

So how many combinations can be formed from seven golfers? The number of permutations is  $P(7, 4) = 7!/3! = 840$ . Since we don't care about order, this contains  $4!$  useless rearrangements of the same names – a given foursome can be ordered in  $4!$  ways. The number of foursomes is  $P(7, 4)/4!$ . We call this the number of combinations of  $n$  things taken  $r$  at a time. This is written as  $nCr$ ,  $C(n, r)$ , or  $\binom{n}{r}$ , we read this as “ $n$ ” choose “ $r$ .”

$$C(n, r) = \frac{P(n, r)}{r!} = \frac{n!}{r!(n-r)!}$$

We can write the relationship as

$$P(n, r) = C(n, r) \cdot P(r, r)$$

Combinations can be used to answer such questions as: What is the coefficient of  $x^8$  in the binomial expansion of  $(1+x)^{15}$ ? I suppose we could work it out, but there is an easier method. Think! We obtain the term in  $x^8$  by multiplying 1's in 7 brackets by  $x$ 's from the other 8. The number of choosing 8 from 15 is

$$C(15, 8) = \frac{15!}{8!7!}$$

This is the desired coefficient of  $x^8$ .

## Detailed Example

Let's check this method for  $(1+x)^3$

$$(1+x)^3 = (1+2x+x^2)(1+x) = 1+2x+x^2+x+2x^2+x^3 = 1+3x+3x^2+x^3$$



The coefficients are

$$x^0: C(3,0) = \frac{3!}{3!0!} = 1$$

$$x^1: C(3,1) = \frac{3!}{1!2!} = 3$$

$$x^2: C(3,2) = \frac{3!}{2!1!} = 3$$

$$x^3: C(3,3) = \frac{3!}{0!3!} = 1$$

Generalizing, if we have  $(a + b)^n$ , then the coefficient of  $a^{n-r}b^r$  is  $C(n, r)$ , usually written as  $\binom{n}{r}$ . The binomial expansion can then be written as

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

## The Basic Problem of Thermal Physics

Many problems that arise in thermal physics reduce to the following: Given  $N$  balls and  $n$  boxes, how many ways can they be arranged so that there are  $N_1$  in the first,  $N_2$  in the second,  $\dots$ ,  $N_n$  in the  $n^{th}$ , and what is the probability that a given distribution will occur? Let's take a particular case.

$N = 15$

Box Number	Number of Balls
1	3
2	1
3	4
4	2
5	3
6	2

Order is obviously unimportant in this case.

How many ways can we choose 3 from 15 for the first box? The answer is  $C(15,3)$ . This leaves

12 box 2  $C(12,1)$

11 box 3  $C(11,4)$

7 box 4  $C(7,2)$

5 box 5  $C(5,3)$

2 box 6  $C(2,2)$

Therefore, the total number of ways is

$$C(15,3)C(12,1)C(11,4)C(7,2)C(5,3)C(2,2)$$

$$= \frac{15!}{12!3!} \frac{12!}{11!1!} \frac{11!}{7!4!} \frac{7!}{5!2!} \frac{5!}{2!3!} \frac{2!}{2!0!} = \frac{15!}{4!(3!)^2(2!)^21!}$$

Now, what's the probability of this distribution occurring? Assume a random distribution, each ball has a one in six chance of being in any box. So we can place the

1<sup>st</sup> ball in 6 ways

2<sup>nd</sup> ball in 6 ways

3<sup>rd</sup> ball in 6 ways

4<sup>th</sup> ball in 6 ways

5<sup>th</sup> ball in 6 ways

...

15<sup>th</sup> ball in 6 ways

The fundamental principle of counting yields  $6^{15}$  total ways. Recall

$$P(E|\text{Conditioning Information}) = \frac{\text{Number of outcomes favorable to E}}{\text{Total number of outcomes}}$$

so

$$P = \frac{\frac{15!}{4!(3!)^2(2!)^21!}}{6^{15}} \approx 8 \times 10^{-4}$$

Now, we'll consider restricting the number of particles allowed in a box. Consider the problem of placing 4 balls in 6 boxes with the constraint that the maximum number of balls in any box is 1. The total number of ways of distributing the balls is not  $6^4$ . There are six ways of choosing the first box, five ways of choosing the second, four ways of choosing the third and three ways of choosing the fourth box. The total number of ways is  $6 \times 5 \times 4 \times 3$ .

Let's rewrite this. Consider the number of permutations of 6 objects taken 4 at a time.

$$P(6, 4) = \frac{6!}{2!} = C(6, 4) \cdot 4! = \frac{6!}{4!1!} 4! = 6 \times 5 \times 4 \times 3$$

Now what's the probability that the first two boxes are vacant when the other four are filled?

The number of ways of arranging 4 balls in the last four boxes is  $4!$ . This is the number of favorable outcomes. The probability of the distribution is

$$P = \frac{4!}{C(6, 4)4!} = \frac{1}{C(6, 4)}.$$

$4!$  is the number of ways of arranging 4 balls in the 4 occupied boxes. This will be the same for any given set of 4 boxes. The quantity  $C(6,4)$  tells us the number of ways of picking the four occupied boxes from the 6 boxes. There is only one way to pick the first two boxes to be vacant so the probability is  $\frac{1}{C(6,4)}$ . We could look at this problem in another way. Consider the set of four identical (or indistinguishable) balls being placed in six distinguishable boxes.

The balls are identical, the  $4!$  arrangements of the 4 balls in the 4 boxes all look alike. So there are  $C(6,4)$  distinguishable rearrangements of 4 identical balls in 6 boxes (one or no balls per box.) All the arrangements are equally probable, so the probability of any one arrangement is  $\frac{1}{C(6,4)}$  as before.

The probability of 2 particular boxes being empty is the same whether the objects are distinguishable or not. This only happened because all the distinguishable arrangements are equally probable. Without the restriction of 1 per box the distinguishable arrangements are not equally probable and we'd get different results.