a brief random tour of probability for epidemiologists

Charles DiMaggio, PhD[1,2]

October 22, 2012

[1] Department of Anesthesiology, Columbia University, College of Physicians and Surgeons, New York

[2] Department of Epidemiology, Columbia University, Mailman School of Public Health, New York

1 introduction

> "Chance is a more fundamental conception than causality". Max Born. (Natural Philosophy of Cause and Chance)
>
> "Probability is the very guide of life". Cicero (De Natura Deorum)

Scientific reasoning is often presented as deriving from deduction (premises to fact) or induction (facts to premise) with deduction most often coming out the more sound approach. In the world of epidemiology, deductive reasoning, based on the beauty of pure logic, has limitations. First, there may not be any true or universally acceptable premises upon which to base our reasoning. Second, while a soundly deduced argument can assure us of the truth or falsehood of a thesis, it is less helpful in answering all those questions in between "yes" and "no". Enter probability.

Mathematical probability is a model that applies (with more or less accuracy) to reality in the same way other scientific models do, e.g. Galileo's law on the motion of pendulums. It is in many ways the very basis of epidemiologic analysis.

2 counting

In one of my early public health courses, I was told epidemiology means "counting bodies". (It doesn't) I also remember a bumper sticker that said, "Epidemiologists count!" (They do.) Counting is a big part of everyday life, and while it seems almost intuitive, comes with a set of rules (and a beauty) all its own.

When counting things to see how likely or unlikely they are, we'll soon realize that whether we need to take their order into account becomes an important consideration.

2.1 when order matters

Permutations are a way to count (or aggregate) things so that their order is taken into account.

n things counted (or aggregated) n at a time is denoted nPn and can be done $nxn-1xn-2...n-n+1$ ways, which is denoted $n.$[1] n things can be counted (or aggregated) r at a time (nPr) $nxn-1xn-2...n-r+1 = n!/(n-r)!$ ways.

2.2 when order doesn't matter

More commonly in epidemiology we find that order doesn't really matter. This is referred to as combinations and denoted nCr or for n choose r and has the formula $n!/r!(n-r)!$ [2]

3 a framework

Classical Definition of Probability: $Pr[A] = \#A/(\Omega)$ i.e. the likelihood of event A occurring given Ω possible outcomes

Let's look at a quick example of why probability is important for epidemiologists. Say there are 7 uninfected and 4 infected people in a population. What is the probability of choosing a random pair of people where one is infected and the other is not?

The first step is to determine Ω , or all the possible pairs. There are a total of 11 people. You begin by picking any 1 from the total population of 11. You then pick 1 from the remaining 10. How many times can you do this? $11x10 = 110$

Next, determine how frequently the outcome in which you are interrested can occur. In this example, that is all the ways you can pick an infected and then an uninfected, or an uninfected and then an infected: $= 7x4 + 4x7 = 56$

[1] there is also a neat little calculation for factorials called Stirling's formula that arises from calculus: $n! = e^{-n}/\sqrt{2\pi n}$

[2] This sorry two sentence summary of the expansive study of combinatorics is a grave injustice not only to a noble field of endeavor but also to my good friend Prof. Jim Cox with whom I discussed its finer points over countless hours and beer.

The probability, then, is the number of outcomes in which we are interested, over the total number of possible events: $Pr = 56/110$

3.1 Venn diagrams

It's sometimes more intuitive to think of probability in terms of the two-dimensional space represented by Venn diagrams, where

- $A \cup B$ is the union of space A and space B. All possible outcomes or points are contained in either A *OR* B
- $A \cap B$ is the intersection of A and B which contains just just those outcomes or points present in A *AND* B
- $\bar{A}$ is the complement of A. It is the space with all the points *NOT* in A
- $A \subset B$ in whichA is subset or is *contained* in B

4 some rules

There are sets of rules of probability. They may vary in name and presentation, but this is how I remember them.

4.1 single events

The first set of rules addresses the probability of a single event occurring. From an epidemiological perspective it is akin to descriptive observations of disease occurrence.

- A basic rule is that an event occurs or does not. So, the probability of an event occurring is bounded by zero and one. This is known as the *non-negative rule* ($A \subset \Omega$ then $0 \leq Pr[A] \leq 1$) and while seemingly obvious, has implications for measurement in epidemiology. Rates and rate ratios have no such natural bounds and must be transformed in some way to allow us to make probability statements about them.
- The next rule follows from the first. It's called the *complement rule* and states that Pr$[barA] = 1Pr[A]$
- Since we can subtract probabilities, the *addition rule* states that we can add them, and that the events A_i form a partition of A such that $Pr[A] = \sum_1^i Pr[A_i]$

4.2 more than one event

The next set of rules address the relationship of the probabilities of more than one event occurring. These are the AND/OR rules. We may consider them relevant to how a variable of some kind is related to another factor.

- The first such compound probability rule addresses the situation when event B *only* occurs when event A occurs. This is called the *subset rule* and states if $A \subset B$ then $Pr[A] < Pr[B]$
- The *multiplication rule* addresses the probability of event A *and* B occurring and states $Pr[A \cup B] = Pr[A|B]xPr[B] = Pr[B|A]xPr[A]$ If the two events are independent, the calculation simplifies to $Pr[A]xPr[B]$.[3]
- The *rule of total probability* states that the probability of A *or* B two occurring is $Pr[\Omega] = Pr[A \cup B] = Pr[A] + Pr[B] - Pr[A \cap B]$. It is sometimes called the *inclusion - exclusion rule* because when adding up the individual probabilities you are counting $A \cap B$ twice and that probability must subsequently be excluded. So, for example, if $Pr[A] = .5$, and $Pr[B] = .5$, then $Pr[A \cup B] = .5 + .5 - .25 = .75$ [4]

4.3 some classic problems

The basic rules of probability have been applied to solve some classic and fun (to some) problems. The complement rule provides the critical first step for addressing problems that ask for the probability of something occurring at least once or not at all. For the "once" problems, the complement of doing something at least once, is not doing it at all. For the "not at all" problems, the probability of an event not occurring is 1-Pr of the event occurring.

4.3.1 the birthday problem

Here's one to try the next time you are giving a lecture or presentation. What is the probability that at least 2 people in a room full of any number (n) of people share same day and month of birthday? We will see that beyond a minimal threshold of around 2 dozen people it is better than even bet. With 50 or so folks ina room, it is a virtual certainty.

---

[3] We will return to conditional probabilities and the implications of this rule momentarily

[4] An implication of the law of total probability is that a simpler model is more probable, because, by extension, $Pr[\Omega] = Pr[A] = Pr[A \cap B] + Pr[A\bar{B}] = 1$ So, for example, shortness of breath alone is more likely to be due to a pulmonary embolus than is shortness of breath and chest pain.

We begin with the most basic observation. The probability of any one person having a birthday in any given year is 365/365 or 1. Following that, we observe that the probability of any second person having a different birthday is 364/365. And, the probability of a third person having a birthday different from the first two is 363/365. By the complement rule then, the probability is $1 - (365x364x363/365x365x365) = 1 - (364 - n + 1)/365^n$

So, for example, for 10 people the probability is 12%. For 23 people the probability is 51%. And, for 50 people the probability is 97%.

4.3.2 Montmart's problem

In the game of *trieze* there are 13 sequentially numbered balls. If you draw 13 balls one after the other, draws, what is probability that at least one ball will be drawn in the correct sequence?

Problems like this, where we're interested in *at least* one occurrence of an event, are best approached by starting with the probability of no such events occurring. In this case, the probability of no balls being drawn in their correct sequence is 1/2!+1/3!-1/4!+1/5! etc... After about n=8, this oscillating sequence converges rather neatly to 1/e. [5] And the probability of the event occurring is 1 minus this probability of if not occurring.

So, for example, if 8 men [6] check their coats, and the person behind the counter loses all their tickets, there is still about a 2 to1 chance [7] that at least one of them will get their correct coat back.

subsubsectionthe monty hall problem

You are on a game show. There is a fabulous prize behind one of three doors. You choose a door. The game show host (Monty Hall, from the old "Let's Make a Deal" show) opens one of the doors that you *did not* choose and reveals it is empty. He offers you the option of switching your current choice. Should you switch? Yes.

The probability that the prize is behind the remaining door you did not choose is not, as you might expect, 50:50. It is, rather 2/3. You can think of it in terms of the complement rule. The door you originally chose had 1/3 probability of holding the prize, with the other two doors having a 2/3 probability of not holding the prize (1/3 for each door). Monty has eliminated one of the doors that contributed to that 2/3 probability. The other door of that 2/3 probability combo essentially "inherits" the 1/3 probability from the empty door Monty revealed.

[5] In this case I'm going to ask you to just take my word for it. But it's an easy enough thing to Google

[6] And beyond this number the probability is the same.

[7] $Pr = 1 - 1/e = 1 - .36 = .64$

It might be easier to understand it if you imaging 100 doors rather than just three. Say you choose 1 door, and then Monty reveals 97 empty doors. Should you stay with the 1 in a 100 probability of the door you originally chose? Or accept Monty's generous offer to put 97/100 of the probability into the two remaining doors you did not choose?

Problems like these are relevant for epidemiologists and other researchers, because illustrate that chance and probability can sometimes be far from intuitive. One need look no further than recent controversies over PSA screenings to see applications of this kind of thinking.

5 conditional probability and Bayes rule

5.1 conditional probability

One event depends on another occurring, i.e Pr(A—B) or the probability of A given B. Basically, it changes the Ω possible outcomes that can occur. Consider a roll of two dice. What is the probability that the sum ofnthe two faces is 8? $\Omega = 6x6 = 36$ There are a total of 5 ways you can get a sum of 8: 2+6, 3+5, 4+4, 5+3, 6+2. The probability then is 5/36. What if, though, you first rolled a 2? Ω is now changed, reduced to the 6 possible outcomes of the roll of the second die, of which only one face when combined with 2 results in 8 and the probability becomes 1/6.

The general rule is: Pr(A—B) = Pr($A \cap B$) / Pr(B)

This arises from the multiplication rule for $A \cap B$ = Pr(A—B) x Pr(B) = Pr(B—A) x Pr(A)

5.2 Bayes rule

If we rearrange the multiplication rule, something interesting (and perhaps a little surprising) arises:

$$Pr(A|B)xPr(B) = Pr(B|A)xPr(A) \tag{1}$$

$$Pr(A|B) = \frac{Pr(B|A)xPr(A)}{Pr(B)} \tag{2}$$

Let's consider what this means. If we know something about the individual events A and B occurring, We can determine the probability that a *previous* event (B) occurs, given that the final event (A) has occurred (Pr(B—A))

This relationship was first described by Rev. Thomas Bayes, and in its full form looks like this:

$Pr(B|A) = Pr(A|B)P(Bi)Pr(A|B)Pr(B) + Pr(A|\bar{B})Pr(\bar{B})$

where P(B—A) is called the posterior probability P(B) is called the prior probability P(A—B) is the conditional probability or likelihood

Whether they are aware or it or not, clinicians and clinical epidemiologists use Bayes rule all the time.

Say a disease occurs in 0.01 of a population and that a test for it is positive in 95% of people who have the disease. What is the probability that someone has disease given that they have a positive test or Pr(D—+) ?

Pr(D) = 0.01 $\bar{D}$ = 0.99 Pr(+—D) = 0.95 Pr(+—$\bar{D}$) = 0.05

We simply plug the numbers into the formula:

$Pr(D|+) = \frac{Pr(+|D)Pr(D)}{Pr(+|D)Pr(D)+Pr(+|\bar{D})Pr(\bar{D})} = \frac{(0.95)(0.01)}{(0.95)(0.01)+(0.05)(0.99)} = 0.1$

Again, let's think about what this means. A person with a positive test result has only a 10% probability of actually having the disease. Is this a bad test? Well, we know it is positive in 95% of cases where the disease is present (95% sensitivity) so the test itself is not the problem. It has something to do with the prior probability of disease.

The 1% probability that a person has the disease somehow influences the additional evidence we derive from the test. This is, in essence, the heart of Bayesian thinking. We start with some prior expectation about in outcome. We gather additional evidence. We update or reconsider our original expectation based on that evidence.

In this example, the test was in fact rather helpful. A positive result changed our our thinking by an order of 10. We went from a 1% probability that the person has a disease, to a 10% probability.

Looked at this way, Bayesian thinking is a natural fit to clinical practice. A patient walks in (already a bit of evidence) complaining of epigastric pain. You form a list of possible diagnoses, each with their own (prior) probability. Those probabilities are influenced by things like how common or rare those diagnoses are in your population. Perhaps alcoholism is common among your patients and when you hear abdominal pain one thought is pancreatitis. History, physical exam, and lab tests are ways to update your initial probabilities. In this case an elevated lipase level changes your prior probability from "maybe" to "likely". You have just conducted an intuitive Bayesian analysis.

The challenge of Bayesian approaches in the health sciences is to validly and reliably quantify the probabilities and how we combine or update them.

6 chance, patterns and causality

Random processes almost invariably result in streaks or patterns. In a large series of tosses of an unbiased coin, we will invariably see (relatively) more heads or tails in a row, but the underlying process is still random. We are, though, seemingly 'wired' to confirm our ideas about streaks or patterns rather than falsify them.

6.1 mathematical expectation

Pascal also left us the idea of mathematical expectation. In essence, the expectation for an outcome that is subject to chance is the sum of the value of each possible outcome times its probability. It is a very clever way of summing up the pros and cons of a bet. Say each roll of a die costs you one dollar. If you roll a two, you win two. How good of a bet is this? The expectation is $(\frac{5}{6})(1) + (\frac{1}{6})(2) = .5$ Over the long run, you would expect to walk away with about 50 cents.

Interestingly, Pascal had a more other worldly application for this concept. The story, as reported in his Pensees, is that Pascal had a series of 'visions'. As befits a mathematician, he records the exact date of his mystical experience, 23 Nov 1654. On or about that date, he realized there was a wager one couldn't lose, and it concerned the existence of God.

Say you are completely unsure about God. Your probability is then 0.5. Should you 'bet' on the existence of God? If you are correct and God does exist, the reward is infinite bliss. If you are wrong and God doesn't exist, the "penalty" is having lived a moral life. The expectation then = (0.5 * infinity) + (0.5 * small sacrifice). To Pascal it was as close to a sure thing as possible.

7 probability distributions

A probability distribution is a curve that shows on the x axis all the values a random variable can take with probability of each value on the y axis. There are many kinds of probability distributions, reflecting different underlying likelihoods for different kinds of events. They are often characterized by the parameters which define them, such as their expectation (a measure of central tendency) and their variance or standard deviation (a measure of spread around that central value).

Some probability distributions can be used to describe health-related outcomes, and are of particular interest to epidemiologists.

7.1 the binomial distribution

A string of successes or failures arising from a series of trials gives rise to the binomial distribution. These successes or failures can be any outcome that occurs with a yes or no certainty. In epidemiology we are interested in whether a person does or does not have a disease, or whether someone lives or dies.

The probability that an event occurs k times in n trials is described by the formula $\binom{n}{k}p^k q^{n-k}$ where p is the probability of the event occurring in a single trial, q is the probability of the event not occurring ($1-p$ by the complement rule) and $\binom{n}{k}$ is the binomial expansion "n choose k" the formula for which is $\frac{n!}{k!(n-k)!}$ 1[8]

A binomial probability distribution can be defined by an expectation $= np$ and a $\sigma = \sqrt{npq}$

7.1.1 the binomial expansion and Pascal's triangle

The binomial probability described above is an application of the more general formula for expanding any binomial expression, $(a+b)^n = \binom{n}{k}a^k b^{n-k}$. For example, $(a+b)^4 = \binom{4}{0}a^0b^4 + = \binom{4}{1}a^1b^3 + = \binom{4}{2}a^2b^2 + = \binom{4}{3}a^3b^1 + = \binom{4}{4}a^4b^0 = b^4 + 4ab^3 + 6a^2b^2 + 4a^3b + a^4$

Blaise Pascal (1654) developed his famous[9] triangle as a way to solve for permutations that arise from the $\binom{n}{k}$ permutations . Pascal noticed a pattern and built a series of numbers where each row is made of the sum of the two numbers directly above:

$$abcd \tag{3}$$
$$a + bc + d \tag{4}$$
$$a + b + c + d \tag{5}$$
$$\tag{6}$$

The answer to the binomial expansion $\binom{n}{k}$ is found on the n^{th} row, k steps to the right (or left, since it is symmetric).

[8]! is factorial notation and stands for $nxn - 1xn - 2x...xn - ()$

[9]To some.

8 the normal probability distribution

subsection large numbers and central limits This most famous of probability distributions deserves a healthy preamble. So, before moving to a consideration of the normal curve, let's consider how probability varies and where the curve came from in the first place.

8.0.2 Chebyachev's theorem

This theorem formally links probability to variation, and as importantly, quantifies the relationship. The mathematician Chebyachev noted that at least $1 - \frac{1}{h^2}$ of the total probability of an event occurring is within $h\ \sigma$ of the mean or expectation of the event. So, in general , at least 75% of the probability of an event occurring is within 2 σ since $(1 - 1/2^2 = 3/4)$. As a concrete example, consider a health measure whose mean value in some population is 3.6 and that this value varies by 1.02. Chebyachev states with certainty that 75% of the values will be between 1.56 and 5.64. [10]

8.0.3 central limit theorems

You will probably [11] come across references to the *central limit theorem* (usually when someone is trying to convince you that an assumption of normality is justified). This actually refers to a number of theorems, that became more general and more powerful over time, that attempt to define the *limits* of probability distributions in the sense that the more trials you do, the closer you approach a single estimate for μ or σ.

8.1 normality and error

The historical pedigree of what we now call the normal curve can, in fact, be traced to Blaise Pascal's series of binomial expansions. In 1733, DeMoivre was trying to figure out the bottom-most rows of Pascals Triangle. He calculated a geometric approximation to even the most remote row defined by a mean and deviation from that mean, where he noted that this deviation was about 1/2 the width of the approximating curve at 60% of its maximum height. The German Frederich Gauss later realized this curve could be used to approximate measurement error. In

[10] This is actually quite conservative, since now know that for normally- distributed values, this interval will actually cover 95% of the probability

[11] Pun intended

1810,LaPlace used it to prove his central limit theorem, that the more measurements one makes, the more the errors are normally distributed.

This leads to the important observation that in measuring things such as health outcomes, we are interested in error, because it is most often our errors that are randomly distributed, not the thing itself.

9 building the normal curve

A probability graph represents probability by both height (the y value for a particular valve of x) and by area (for a particular range of x). We also know, in keeping with the laws of probability, the total area in a graph of a probability distribution must be 1.

We can *standardize* a probability graph by (1) transforming the x axis values from observations to units of standard deviation, (2) centering on the mean (or 'break even point'), and (3) multiplying the heights on the y axis by σ to ensure that the total probability equals 1 [12]

After this transformation and rescaling, the Gaussian Distribution can be described by $y = 1/\sqrt{2\pi e^{-x^{2}/x}}$, where x is the number of $.\sigma$ units from μ [13] [14]

We find that for data are distributed in this way, 68% of observations are within 1 σ of μ, 95% within 2 σ and 99% within 3 σ.

Epidemiologists will note that binomial experiments, measurement errors and sums of random observations tend to be normally distributed.

10 the Poisson distribution

A third type of data frequently encountered by epidemiologists (disease counts) is often *Poisson* distributed. It is named after Simeon Denis Poisson, who wrote about probability and criminal behavior. It is apt when we are interested in counting the occurrence of (relatively) rare phenomena, so may be particularly suited to disease outcomes.

It is not a long trip from the Binomial to the Gaussian to the Poisson. In fact, the Poisson may be seen as a special case of the normal distribution where the

[12] NB: after standardization, the y values (heights) no longer represent probability, only the area does

[13] You may see the formula written as $f(x) = 1/\sqrt{2\pi\sigma}e^{-((x-\mu)^{2}/(2\sigma^{2}))}$

[14] $1/\sqrt{2\pi}$ensures the total area is equal to 1

probability of an event is very small and the number of observations is very large. [15]

If we set our expectation to be $\lambda = np$, and substitute λ into the normal probability formula, we will eventually arrive at the formula for Poisson-distributed probability: $Pr[k] = e^{-\lambda} * \lambda^k/k!$ [16]

As an example of how we might apply the Poisson formula to epidemiology, consider the hypothetical (but possible) situation of stocking an expensive medication in a hospital emergency department. If all we know is that on average we use the medicine twice a week, we can apply the Poisson probability formula to figure out the number of doses (k) we should stock: $Pr[k] = e^{-2}2^k/k!$ after plugging in some possible values for k we get the following table:

k	Pr
0	.135
1	.276
2	.276
3	.181
4	.092
5	.036

Preparing at most 4 doses per week is adequate.

10.1 Prussian Cavalry Horse Kicks and the Poisson Distribution

One of the earliest applications of the Poisson distribution to health involved a unique class of injuries: deaths due to horse kicks. In the Prussian cavalry to be precise.

In 1898 Ladislaus von Bortkiewicz reviewed 20 years of data (1875-1895) for each of 14 cavalry troops for a total of 280 data points or "experiments". [17] He was interested in the number of deaths in any given year. There were a total of 196 deaths, so $\lambda = 196/280 = 0.7$ deaths per troop per year. Using this (minimal) information he compared the Poisson prediction to what actually occurred.

[15] We can trace a pedigree from the binomial expansions of Pascal's triangle to their normal approximations courtesy of Gauss to the special case of these normal approximations described by Poisson. And it continues. The exponential distribution (popular in survival analysis) is a special case of the Poisson, specifically the time between Poisson events. And the exponential is actually a special case of the gamma probability distribution with a single shape parameter k for a single transition between events or states, where the gamma can account for sub-periods. This last feature has been applied to the complex modeling of infectious disease transmission among susceptible, immune and recovering populations. A long journey from figuring out how to win at dice, no ?

[16] We don't need to know the underlying probability or even n, just λ

[17] An epidemiologist might refer to these as "troop-year" data

So, for example, the probability of a single death occurring in a single troop over the course of a single year was $Pr[1] = e^{-.7} * .7^1/1! = 0.3476$ This 34% probability translated into $0.3476 * 280 = 97.3$ or approximately 97 instances. So well did the data fit the Poisson probability distribution? Very.

deaths	actual times occurred	predicted times to occur
0	144	139
1	91	97.3
2	32	34.1
3	11	8.0
4	2	1.4
≥ 5	0	0.2

There is an impressive correspondence. We can code up a simple R function to calculate Poisson probabilities.

There is no factorial function in base R so we begin by defining a function for Stirling's formula:

```
stirling<- function(x){
         exp(-x)/(sqrt(2*pi*x ))
                                   }
```

Now we define a function for calculating Poisson probabilities:

```
poisson.prob<-function(k, lambda){
     ( (exp(-lambda))* lambda^k) / sterling(k)
                                                       }
```

Now let's test it out:

```
# lambda^k / k *e^(-lambda)
# e.g. average 2 events per week

2^0/factorial(0)*(2.71828183^(-2))
2^1/factorial(1)*(2.71828183^(-2))

poisson.prob(0,2)
poisson.prob(1,2)
k<-0:7
poisson.prob(k,2)
```

```
# prussion horse  kicks
lambda<-196/280
k<-0:4
predicted<-(poisson.prob(k,lambda))*280
actual<-c(144,91,32,11,2)
cbind(predicted,actual)
```

11 some unfinished business

These are a few stubs of ideas I'd like to pursue and expand on if and when I ever find the time.

11.1 a limit

The idea of a limit stretches back to Zeno in 5th century BC Greece. As you approach an object you can imagine halving the distance again and again ad infinitum: 1/2, 1/4, 1/8, 1/16, 1/32... You can't add up this infinite series, but if you sum it up one item at a time (incrementally), you will see a pattern:

$$1/2 \tag{7}$$
$$1/2 + 1/4 = 3/4 \tag{8}$$
$$1/2 + 1/4 + 1/8 = 7/8 \tag{9}$$
$$1/2 + 1/4 + 1/8 + 1/16 = 15/16 \tag{10}$$
$$\tag{11}$$

The denominator is 2 to the number of elements in series, the numerator one less than denominator. At 20 intervals, the sum is 1,048,575/1,048,576. The sum is, in fact, not approaching infinity as Zeno thought, but unity.

11.2 the quincunx

The quincunx, descrubed by Sir Francis Dalton in 1889, board with evenly spaced pegs which results in a Gausian shaped distribution of balls dropped into it. The name is based on a roman coin called a quincunix, worth 5/12 of a lira, with five spots on it. An impressive example may be found in a Boston children's museum.

11.3 distribution of first significant digits

The numbers 1,2,3, and 4 occur more frequently as 1st digits of any set of random numbers. This is a consequence of our base 10 decimal system. There are 9 possible digits (1-9). The probability of number the number n (or less) is not n/9, but actually $log[base10](n+1)$. So, for example, the probability of the number 4 or less is $log[base10](5) = 0.699$ (not $4/9 = 0.444$).

11.4 statistics vs. probability

It has been said that probability argues from populations to samples, while statistics argues from samples to populations.

11.5 Buffon's Needle

```
Pr[will cross line]=2c(Pi)a
where c=length of the needle, and a=distance parallel lines are apart
so, Pi=2cN/am
where N=#tosses,  and m=# times it crosses a line
```

11.6 rare occurances

The basic question is: "Compared to what?" For example, a particular poker hand has a low probability of occurring, but so do all other poker hands.