Chapter 1

Everyday Perception of Chance

Note. *This is the only lecture where I use a laptop projector, to show material from the web site, only a small portion being copied into this chapter. Other lectures are "chalk and talk", except for showing data, diagrams etc from other sources.*

In what contexts do you think of elephants? I suspect you can't answer very confidently, partly because now I have put the idea of elephants into your mind with this question, it's hard to remember the previous times. Our topic today is

> In what everyday contexts do "ordinary people" perceive events in terms of chance?

Just ask them! isn't a helpful way to try to answer – we humans do forget.

There is substantial academic research relating to perception of probabilities, which will be the topic of our "psychology" chapter later. But this typically studies responses when subjects are *prompted* to think about chance by being asked some specific question which plainly involves chance. Our topic today is: when do you think about chance, *unprompted*?

Of course it's easy to imagine many contexts, but a main theme of this course is to look at actual data, not imaginary data:

Don't Make Stuff Up!

I will talk about two sources of actual data (and then, as evil fun, compare with some imaginary data). The first is a data collection exercise done by undergraduates that you could easily repeat yourself.

1.1 References to chance in blogs

Google blogsearch does what you think it does, so one can just search for words or phrases relating to chance. Below are the first 22 instances found when searching on the phrase *one in a million chance*. (Literal text italicized, without correcting spelling or grammar). I have divided them into 4 categories.

1. Past events that happened to writer. (6)

Finding a romantic partner. *There was this weird connection that I felt when I first met him . . . Seeing how its like a one in a million chance to find that one person you connect with.*
(similar quote omitted).
Major life events. *I have . . . syndrome. The fact that I ever became a mother was a "one in a million chance".*
Unusual dramatic events. *. . . and they [adults] all start talking about how im too young to be going out by myself . . . But it's not like im going to listen to them, what happened [witnessing a mall shooting] was a once in a million chance.*
Unusual minor events.
(vacation went unexpectedly well: quote omitted)
(throwing chips in drunken party: quote omitted)

2. Possible future events that might affect writer. (8)

Minor pleasant possibilities. *I'm somewhat hoping to meet friends there . . . it's a one in a million chance.*

i'm waiting for the day they [upcoming movie/TV filming locations] say my city which is one in a million chance

. . . got this contest. It's a one in a million chance to get some people . . . to tell me what they think of my work.

my greatest ambition is to see [a supernova] one day, though there's probably a one in a million chance that i will. smaller than that.

I'd only be satisfied with one particular scenario and there's maybe a 1 in a million chance of that happening . . . no, less. I would get struck by lightning before that happened, twice
Worries. *Of course if I don't go [to the doctor about certain symptoms], there's that one in a million chance that I'll be sorry I didn't.*
(similar quote omitted)

There are things that we were never told that really end up happening to most women [during pregnancy]. Instead we were told the things that we had a one in a million chance of experiencing.

3. Events affecting specific other people. (2)

On the one in a million chance that Christine actually gets hired to do costumes for . . .

There is a one in a million chance that [a particular NHL player] gets picked up on waivers

4. Impersonal speculation. (6)

On the other hand, if you chase after it [a volleyball spike by opponents], who knows? It might just be one-in-a-million chance that you'll get it, but isn't that a chance worth taking?

When you're looking for the one in a million chance of getting a Beethoven you could be overlooking an Einstein.

Becoming a successful actor, singer, or dancer is a one in a million shot in the dark during a snowstorm.

(similar quote omitted)

. . . reflect on how [Valentine's day] has brainwashed a whole lot of people into believing that love could actually happen on that day, which is a one in a million chance by the way (which they would argue is worth the risk anyway, which is also bullshit, by the way).

First [one particular sperm] had to survive and beat out millions of other sperm . . . that's like winning the lotta right there . . . only one in a million, and from that point, you got to survive . . .

The results of other searches can be seen on page `blogs.html`. I will give some analysis after the next section, but let me invite readers to think for themselves the "contexts where we perceive chance" indicated by these examples (the four categories I have used are not the most descriptive).

Once one sees this kind of data, it may seem obvious that "this is the sort of data we expected to see". But actually predicting such things is hard. I challenge readers to stop after the first paragraph of the next section, and try to predict what the data will show!

I suggest as a course project that students gather more data using other search terms and other subsets of the internet. For this I insist on a *repeatable experiment*. You must show (a random sample of) results of a specified search, not human-selected ones. The internet is so big that one could invent examples and then search for similar ones, so *selecting* examples is little different from *inventing* them. Specifying a search protocol that gives the kind of "contexts where we perceive chance" examples is harder than it sounds.

It is important to note that I are not claiming that what we find is a statistically accurate sample of "contexts where we perceive chance", either

in general or specifically within the blogosphere. Our purpose is to illustrate actual usage, as opposed to made-up examples.

1.2 Queries to the search engine Bing

I obtained, from the Bing team, a file of all (around 100,000) queries made to Bing containing the strings "chance of" or "probability of". After excluding those which were not actually looking for the chance of something (e.g. were seeking the movie *Cloudy with a Chance of Meatballs*) I had enough patience to examine 675, sorting them into 66 groups of about 10 similar queries. Picking one from each group gives a sample of 66 "representative" queries, shown on the web site at `bing_chance.html`. Here I prune down to a representative 30.

Before turning the page, I challenge readers to predict what the data will show!

Query: what's the chance of getting pregnant after tubal litigation?
Query: chance of pregnancy after intercourse
Query: how to improve chance of getting pregnant
Query: pecent chance of getting pregnant with clomid
Query: chance of getting pregnant while breastfeeding
Query: if twins run in my family whats my chance of having them?
Query: chance of having multiples using fertility
Query: chance of siblings both having autism
Query: chance of misscarage after 8 weeks
Query: chance of bleeding with placenta previa
Query: any chance of vaginal delivery if first ceaserian
Query: probability of having an adverse reaction to amoxicillin
Query: can aispirin reduce chance of a stroke
Query: does progesterone increas chance of breast cancer
Query: which treatment has the least chance of prostate cancer recurring?
Query: chance of getting a brain tumor
Query: do chargers have a chance of making the playoffs
Query: probability of flopping a set with pocket pair
Query: does a ring of wealth affect the chance of the dragon pickaxe drop in runescape?
Query: persent chance of getting shot if you run from an attacker
Query: chance of surviving severe head injury
Query: chance of having white christmas ontario
Query: chance of rain in september dallas texas
Query: what are the chance of becoming a golf pro
Query: chance of closing airports in mex because of swine flu
Query: chance of getting a short sale
Query: probability of winning a traffic ticket court case
Query: chance of food spoiling if left out over night
Query: probability of life and evolution
Query: wich technology has the least probability of a collision

Discussion. The best descriptive phrase I can devise for these examples of queries is that they are *personal and concrete*: they typically concern near-future events with substantial significance to the person involved. The examples from blogs are loosely similar, though (as "obvious" from the nature of blogs) people also write about past personal events and general thoughts.

It is striking that around half the queries concern medical matters, and more than half of those concern pregnancy and birth control. Of course this data derived from search queries is several steps removed from the conceptual question "how do people think about chance in everyday life" and we certainly do not suggest a quantitative correspondence like "half of our everyday perception of chance involves medical matters".

Comparing these examples to our draft list of 100 *contexts* shows both a reasonable spread over the "everyday life" contexts there, and does not suggest (to me) any significant omitted context.

1.3 Some imaginary data

The previous two sources of "real data" are the only ones that I have actually been able to exploit. Of course I would like 8 other sources which reveal other aspects of "contexts where we perceive chance", so please send me suggestions or do the data collection yourself!

I am confident it is easy to distinguish between real and imaginary examples. Here is a small project. Find two friends who have not seen this material. Ask one to imagine and write down ten instances of how one might use the phrase "one in a million chance" in a blog, or ten instances of "chance of" queries one might type into a search engine. Give this list, and a sample of 10 examples from our earlier lists, to the other friend. I bet the other frind will unhesitantly identify the real examples.

An intriguing source of imaginary examples. is the book *Luck: The Brilliant Randomness Of Everyday Life* by Nicholas Rescher, a former President of the American Philosophical Association. This short book could be viewed as an unusually erudite blog, or as an unusually reader-friendly monograph. And the content of his musings about Luck is perfectly reasonable. But what intereste me here is the examples he cites. On the page `rescher.html` are all the examples from the parts (Introduction; Chapters 1 and 3) that seem closest to the "everyday life" theme. I categorize them as

Specific historical events (11)

Iconic headlines (4)

Conventional examples of luck (20)

Notes for a historical novel? (23)

All except the first category consist of invented examples. Here are the last 12 examples from the fourth category. Italics are exact quotes, others are paraphrases.

potential victim saved because a would-be assassin missed the bus

being wounded by an assassin who mistakes one for someone else
injured as bystander in political demonstration
you were inadvertently delayed and just missed crossing on the Hindenberg
hit by falling icicle
fighter pilot hits ejector button instead of defroster
burglar who breaks into a house just before its owner returns well-armed from a bear hunt
the painter who produces a [long-sought] effect . . . by throwing his brush at the picture in a fit of rage . . .
coming down with a disease for which a cure has just been discovered
author whose biography of a celebrity hits the bookshops just as its protagonist is enmeshed in a highly publicized scandal ...
scam victim accidently profiting
the winner of a lottery who decides to build a dream cottage on Krakatoa

In class, I cannot resist saying "Wow, Everyday Life in a Philosophy department sure seems more exciting than in a Statistics department" and then comparing to the real "everyday life" revealed by the type of data in the previous sections.

Now in one sense I am merely being humorous. You and I both know the author did not intend his examples to be literally "everyday life"; instead, he was interested in the abstract ideas surrounding *luck* and just made up illustrative hypothetical examples as he wrote.

But in another sense I am perfectly serious. The author is adopting a style of intellectual enquiry where he starts with abstract ideas and then invents hypothetical examples to justify the ideas. To what extent is this a useful style of intellectual enquiry?

1.4 Mathematical probability textbook examples

Note: *This section is here for the "popular science" reader, since it's hardly news to my students.*

What is the picture of chance – that is, of the contexts where chance arises – that one obtains from the examples and exercises in an undergraduate mathematical probability textbook? I would put them into 4 style categories, below. Illustrative exercisesare taken from Grinsted-Snell *Introduction to Probability*, available online, which I regard as one of the best textbooks.

1. Purely mathematical.

Let $X_1,\ X_2,\ \ldots,\ X_n$ be n mutually independent random variables, each of which is uniformly distributed on the integers from 1 to k. Let Y denote the minimum of the X_i's. Find the distribution of Y.

2. An (at least somewhat) interesting real-world question and an (at least somewhat) realistic model.

A large number, N, of people are subjected to a blood test. This can be administered in two ways: (1) Each person can be tested separately, in this case N test are required, (2) the blood samples of k persons can be pooled and analyzed together. If this test is *negative,* this one test suffices for the k people. If the test is *positive,* each of the k persons must be tested separately, and in all, $k+1$ tests are required for the k people. Assume that the probability p that a test is positive is the same for all people and that these events are independent.

For small p, show that the value of k which will minimize the expected number of tests under the second plan is approximately $1/\sqrt{p}$.

3. Actions one could do, but with no evident purpose.

A die is rolled 30 times. What is the probability that a 6 turns up exactly 5 times?

4. A real-world story with invented data and/or a very unrealistic model. This is my main concern, so let me give several examples.

a. A student must choose exactly two out of three electives: art, French, and mathematics. He chooses art with probability 5/8, French with probability 5/8, and art and French together with probability 1/4. What is the probability that he chooses mathematics? What is the probability that he chooses either art or French?

b. A restaurant offers apple and blueberry pies and stocks an equal number of each kind of pie. Each day ten customers request pie. They choose, with equal probabilities, one of the two kinds of pie. How many pieces of each kind of pie should the owner provide so that the probability is about .95 that each customer gets the pie of his or her own choice?

c. Take a stick of unit length and break it into two pieces, choosing the break point at random. Now break the longer of the two pieces at a random

point. What is the probability that the three pieces can be used to form a triangle?

d. Suppose you toss a dart at a circular target of radius 10 inches. Given that the dart lands in the upper half of the target, find the probability that

1. it lands in the right half of the target.

2. its distance from the center is less than 5 inches.

3. its distance from the center is greater than 5 inches.

4. it lands within 5 inches of the point $(0, 5)$.

e. You are in a casino and confronted by two slot machines. Each machine pays off either 1 dollar or nothing. The probability that the first machine pays off a dollar is x and that the second machine pays off a dollar is y. We assume that x and y are random numbers chosen independently from the interval $[0, 1]$ and unknown to you. You are permitted to make a series of ten plays, each time choosing one machine or the other. How should you choose to maximize the number of times that you win?

f. A small boy is lost coming down Mount Washington. The leader of the search team estimates that there is a probability p that he came down on the east side and a probability $1 - p$ that he came down on the west side. He has n people in his search team who will search independently and, if the boy is on the side being searched, each member will find the boy with probability u. Determine how he should divide the n people into two groups to search the two sides of the mountain so that he will have the highest probability of finding the boy. How does this depend on u?

Discussion My 4 categories are rather fuzzy – do "drawing balls from urns" problems fit category 1 or 3? So I haven't tried to examine textbooks to find the percentages in each category. But I suspect every introductory probability textbook has an extremely low percentage in category 2. Good textbooks on *statistics* do rather better, e.g. Freedman - Pisani - Purves - Adhikari.

What's unrealistic about the examples in category 4 is (I hope) clear to the reader. In (a) people don't choose at random, and (b) breaks the commandment that should be given in a first class on probability: thou shalt not assume different possibilities are equally likely, without some darn good reason. In (c) I suspect it's physically impossible to break at a uniform random point, and in (d) the implied uniform distribution only applies if

you are *very* bad at throwing a dart. The casino in (e) would rapidly go bankrupt, and in (f) it's hard to justify independence.

And of course the authors know these exercises are unrealistic, just as the philosopher author knew his examples weren't really "everyday life".

Now I must admit that in teaching such courses, I use the same style of examples as do the textbooks. Indeed, part of the reason for teaching this completely separate course is that it's too hard to put realistic material into a conventional mathematics-focussed course. But what *should* one do in a first course, ideally? Certainly one needs some purely mathematical examples to illustrate math techniques, and the justification for phrasing examples in terms of dice and urn is to provide a more concrete visualization than would a purely mathematical formulation.

On the other hand, if you look at a textbook for a course being taught on history or biology, you will see page after page[1] of declarative sentences, and – amazingly enough – there is nothing that the author has "just made up" in the style of our textbook examples.

Refering to our list of 100 contexts, in a typical textbook you will find extensive occurence of "explicit games of chance based on artifacts with physical symmetry" and briefer occurence of "random sampling for representativeness" and a few others. This, and the everyday life examples from blogs and Bing shown earlier, indicate the huge disconnect between introductory textbooks on probability and the big picture of the role of chance.

On a positive note, the best instance I know of academics engaging broad topics of popular interest is the web site *Understanding Uncertainty* from Cambridge U.K. The site is centered around issues of health (recall our Bing data showed that about half the searches were related to health) but covers other popular angles (coincidences, lotteries etc) and gives commentaries on risk and statistics items in the news. The latter is the focus of another valuable site, *Chance News*.

1.5 Which side of the Looking-Glass is the fantasy?

A common view amongst academics involved with probability is that ordinary people are pretty dopey when it comes to understandng randomness. They are superstitious about luck or coincidences. Under irresponsible media influence they have a completely distorted notion of which risks in everyday life are substantial and which are negligible. They gamble when the

[1] In talks I show some typical pages

odds are against them (horses, lotteries, casinos) but not when the odds are in their favor (increase your insurance deductible). They waste billions on useless stock market advice or managed mutual fund fees. They misconstrue positive medical diagnostic test results for rare diseases. And so on.

Now all this is true. But it's one side of a picture. What about the academic side? I should clarify that I am discussing *the impression given by textbooks and papers written by academics*, not what individual academics think. The elements of fiction we have seen in the textbook math exercises and in the examples of luck are mostly harmless in themselves, but set a tone that invites one to approach the world with a *suppose*:

> suppose this event has probability p ....

This starts down a slippery slope, liable to end (see the "toy models" chapter later) with the belief one can learn something about the real world primarily from setting up and studying models, rather than primarily from data and experiment. As Taleb wrote in *The Black Swan*:

> the sterilized randomness of games does not resemble randomness in real life.

Setting up a model presupposes you know the rules, but life does not come with a rule book.