Chapter 2

Dice, Thumbtacks and Data: the Local Uniformity Principle

In everyday language a phrase like “pick at random” carries the connotation of equally likely: saying I pick a student at random from my class is saying that every student should have the same chance to be picked. In mathematics, and in this book, random just means “uncertain: with some probability”, and the technical word uniform is added if we do really mean equally likely. In the first lecture of an introductory course I say

**Non-uniformity is the default.** It is wrong to assume different events have the same probability, without some explicit reason or empirical evidence

and continue with

**Symmetry may justify uniformity.** Artifacts like dice, playing cards and roulette wheels are explicitly constructed to have “physical symmetry” between the associated events. It is this physical symmetry that justifies the usual “equally likely” assumptions of mathematical calculations in such games.

Having equally likely outcomes is so atypical (outside of games of pure chance and controlled statistical schemes) that it makes a misleading starting point for any discussion of probability. A better starting point is the notion of a repeatable experiment, and a down-home example is dropping thumbtacks. After coming to rest, either the point is facing up or facing down, but there’s no symmetry between these two possibilities, and so there’s no reason to think that each has probability 1/2. And indeed they don’t (see next section).

Now consider data such as exam scores (at the end of a year) or baseball team

---

1. Of course coins are not primarily randomization devices; for coins here the symmetry is just an incidental consequence of their design.
standings (at the end of a season). These are not random in the same sense as the examples above, nor are they repeatable in the same sense (students and players are different in different years), yet we still expect some notion of “statistical regularity” from one year to the next.

In many ways the two settings (repeatable experiments; observational data) are just different. What we explore in this chapter is a certain feature that they have in common, the *local uniformity principle*, taking a leisurely route to get there.

### 2.1 Thumbtack data

Though many teachers use thumbtacks as a hypothetical example, a quick online search revealed no very serious actual data, so I did my own little experiment. In brief I did four “trials” of 100 drops each, two onto a hard tile floor and two onto a shallow pile rug.

<table>
<thead>
<tr>
<th>Trial 1</th>
<th>Trial 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tile</td>
<td>59</td>
</tr>
<tr>
<td>Rug</td>
<td>53</td>
</tr>
</tbody>
</table>

**Table 1.** Number of “point up” drops in trials of 100 thumbtack drops.

The data is pretty convincing evidence that the chance of landing point up is greater than 1/2.

OK, dropping thumbtacks is a rather dooey example of a repeatable experiment, so let’s move on to a somewhat more interesting example, throwing darts.

### 2.2 County Fair

There is a fairground game in which playing cards are stuck to a large board in a regular pattern, with space between cards. See Figure 1. You pay your dollar, get three darts, and if you can throw the darts and make them stick into three different cards then you win a small prize.

Let’s think about throwing just one dart. Obviously your chance of hitting a card depends on how good you are at throwing darts! Figure 2 shows data of 100 throws at a dartboard, made by a student Beau La Mont, who was aiming at the center of the board. To show scale, we draw an imaginary playing card centered at the center of the board.

---

2 More details in Notes.
3 Via a test of significance, a topic I don’t want to broach yet.
4 The game is also played with balloons as targets. The balloons are squashed together without empty spaces, making the game appear easier. But you soon learn that to burst a balloon, the dart needs to hit it head-on; a hit on the side just bounces off.
5 One missed the board, so is not counted in the “specific card” data below but is assigned a random position in the “some card in the pattern” data.
Figure 1. Playing cards on a wall. The playing cards on the left are “bridge size” 2.25 by 3.5 inches, with spacing 1 inch between rows. The wall is much larger than shown, with hundreds of cards attached. In the center is the “basic unit” of the repeating pattern. On the right is the pattern shrunk by a factor of 3.

Figure 2. 99 dart throws, centered on a 2.25” × 3.5” playing card.
I could have exploited Beau by getting him to repeat the experiment of throwing darts at different sized targets many times. But it’s much easier to work with this one data set of 100 throws and see what would have happened with differently scaled cards. 36 throws would have hit a normal sized card as target, so we estimate the probability as 0.36. The • in Figure 3 show how this probability increases with the card size. As one expects, this probability is near zero for a postage stamp size and near one for a paperback size. If we repeated the experiment with a different person we would confidently expect a curve of • which was qualitatively similar but shifted left or right according to skill at darts.

Returning to the fairground game, we imagine scaling the pattern (as on the right of Figure 1). For normal size cards, Beau would have 58 hits (that is, 36 on the aimed-at card and 22 fortuitously on a different card) and these probabilities are shown as ◦ in Figure 3. As we explain in a moment, without looking at data we can make a theoretical prediction that, regardless of skill level, when the “pattern repeat distance” becomes small the probability of hitting some card will become about 0.54. And the data shows this is indeed true for Beau on scales smaller than a playing card.

This example is obviously relevant to one of our recurring themes, the skill-luck spectrum, but we defer that discussion to Chapter xxx.

What does theory say about this example? The key point underlying the theory is that there is a regular repeating pattern on the wallboard, consisting of repeats of the basic unit in the center of figure 1; the basic unit is a rectangle of board, partly occupied by a card. Since the space between cards is 1 inch, this rectangle has size 3.25 by 4.5 inches. So the proportion of the area of the basic unit which is occupied by the card equals \((2.25 \times 3.5)/(3.25 \times 4.5) = 54\%\). Because the pattern just repeats the basic unit, this means that a proportion
54% of the wallboard is covered by cards. And this proportion is unchanged by shrinking cards and spaces together. So a dart hitting a region of the board, without propensity to hit or to miss cards, should have a 54% chance to hit a card. So the underlying theory is that, when the cards are small relative to the variability of our throws, we have little chance of hitting the particular aimed-at card, and instead our hit is essentially like hitting a purely random point.

At first sight the notion of “regular repeated pattern” may seem special to circumstances like this example, but we’ll see that it occurs somewhat more broadly.

2.3 The physics of coin-tossing

Why do we think that a tossed coin should land Heads with probability 1/2? Well, the “symmetry may justify uniformity” argument already mentioned goes something like this.

- There is some chance, $p$ say, of landing Heads.
- By symmetry, there is the same chance $p$ of landing Tails.
- Neglecting implausible possibilities (landing on edge, being eaten by passing bird, ...) these are the only possible outcomes.
- Since some outcome must happen, i.e. has probability 1, it must be true that $p + p = 1$.
- So $p = 1/2$.

While most people find this argument (and the corresponding argument for dice, roulette etc) convincing, such an argument by logic doesn’t give much insight into where physically the number 1/2 comes from. But the physics is actually quite simple, if we simplify matters a little. Suppose you toss a coin straight up, that it spins end-over-end relative to a horizontal axis, and that you catch the coin at the same height as you tossed it. Then the coin leaves your hand with some vertical speed $v$ and some spin rate of $r$ rotations per second. And there’s no randomness – it either lands Heads for sure, or lands Tails for sure, depending on the values of $v$ and $r$ via a certain formula. Now we can’t see $v$ or $r$, but we can see the height $h$ that the coin rises before starting to fall. Figure 4 shows the result of the coin toss in terms of $h$ and $r$, for a certain interval of values.

At the instant we toss the coin, we are at some point in the phase space illustrated in the figure, and this point determines whether the coin lands Heads

---

6 The reality is more complicated and more interesting – see Notes.
7 The formula underlying Figure 4 is as follows. The height $h$ and time-in-air $t$ are determined by $h = v^2/(2g)$ and $v = gt/2$ where $g = 32$ feet per sec$^2$. So $t = \sqrt{Sh/g}$. If the coin starts Heads-up, then it lands Heads after $n$ rotations if $n - \frac{1}{4} < rt < n + \frac{1}{4}$. So the curves in the figure are the curves $r \sqrt{Sh/g} = n \pm \frac{1}{4}$. See [?] for more details.
or Tails. Contrary to our policy, we don’t have any honest data for the points in phase space determined by an actual series of tosses. But if you practice tossing a coin 24 inches high, you will find it difficult to be more accurate than 24 ± 3 inches, and so we may envisage a series of tosses as creating a collection of points in phase space scattered in some unstructured fashion in the spirit of Figure 2. The symmetry of the coin is reflected in the fact that the bands determining Heads or Tails have equal width; 50% of the phase space determines Heads. A machine can make tosses in such a consistent way that the spread in phase space was small compared to the width of the bands, but a person cannot. A person tossing a coin is like a person throwing darts at stamp-sized cards – without any bias toward any particular band, we have 50% chance to hit a point in phase space which determines Heads.

Figure 4. Phase space for coin tossing. The shaded bands are where an initially Heads-up coin will land Heads, as determined by the height and rotation rate of the toss. Each band indicates a specific number of rotations, from 7 to 27 over the region shown.

2.4 The fine-grain principle

A mixture of peanuts and cashews is coarse-grained, in that you can pick and choose an individual nut, while a mixture of salt and pepper is fine-grained, in that you can’t avoid picking a mixture. This provides a nice metaphor\(^8\).

\(^8\)Though the principle is well-understood, there is no standard phrase: fine-grain principle is my coinage, though the phrase fine-grain is used literally and metaphorically in several areas of science. Textbook physical explanations of randomness usually involve the long-run behavior of a deterministic physical system not depending on the details on the initial conditions \([?]\), but any “long-run” explanation of the examples in this chapter seems unsatisfactory.
The fine-grain principle. Many instances of physical randomness can be regarded as outcomes of deterministic processes with uncertain initial conditions: the randomness comes only from the initial uncertainty. In a few very special instances, the phase space has enough regularity that one can predict numerical probabilities without any empirical data.

This principle is just restating the ideas of Figures 1 and 4 in more abstract language. We have some activity in which the “initial conditions” determine the final outcome. Any particular initial conditions (speed, direction, spin, in our examples) are mathematically a point in some “phase space”. There is some “pattern” showing which initial conditions lead to a particular outcome. On repeating the activity, the initial conditions are not exactly the same each time; they form some unstructured collection of points in phase space. If the phase space has a certain kind of regularity – that the percentage of phase space leading to the particular outcome is approximately the same percentage \( p \) regardless of which region of phase space we’re in – then we can confidently predict that the probability of the outcome is about \( p \), provided the spread of initial points is at least somewhat large relative to the pattern repeat length. The numerical value \( p \) comes from the pattern, not from any details of the uncertainty in initial conditions.

The fine-grain principle is one of those good news/bad news deals. As a conceptual idea it’s very nice; throwing a die\(^9\) to roll on a table is a much more complicated deterministic process than coin-tossing, but one can imagine a high-dimensional phase space which is divided into six regions in some complicated way analogous to Figures 1 and 4. As a concrete tool it’s terrible, because for die-throwing (or just about any real-world example more complicated than the two we’ve discussed) one can’t actually work out what is the pattern in phase space. The reason we believe that a die lands 5 with probability \( 1/6 \) is the argument by symmetry; we can’t “do the physics” to show this via the fine-grain principle\(^10\).

2.5 The smooth density idealization

We turn to an idea that is much more broadly applicable, though less dramatic in its consequences. We first illustrate the idea with artificial data, then with real data. Consider the simplest kind of data-set, a list of numbers, say 40 numbers between 0 and 20; to have a convenient language think of these as exam scores for 40 students. Figure 5 illustrates several such artificial data sets. The simplest probability model for such data is the “draws from a box” model discussed in Chapter xxx; there is a probability for each possible score between 0 and 20 (shown by the figures on the left), and our artificial data picks 40 scores independently\(^11\) according to these probabilities. The resulting data is

\(^9\)I humorously tell students to write on the blackboard 100 times: the singular of dice is die.

\(^10\)Of course there’s a “proof by economics” – see xxx

\(^11\)The value of one pick doesn’t affect the value of another, as discussed in Chapter xxx
shown by the histograms on the right.

![Probability Histograms vs Data Histograms](image)

**Figure 5.** Each right side shows a data histogram for 40 random picks from the probability histogram on the left side.

In the second case the model is deliberately biasing likely values toward the center, and in the third case toward the left, and this non-uniformity is readily visible in the data. But the aspect I want to emphasize here is a *similarity* between all three cases. Each model is “smooth” in the sense that probabilities do not change much from one possible score to the next, whereas in each case the data is “locally irregular”. That is, in the first model we expect each score to come up “on average” about twice, but the exact number of times is often 1 or 3 (rather than 2) and sometimes 0 or 4; the observed frequencies of successive scores switch unpredictably between these values. The average frequency varies between models and between different parts of the range of scores, but always has this local irregularity.

Moving quickly on to real data, Figure 6 is a histogram of actual scores for a class of 71 students, from data `exam.1`. 
Figure 6. Scores for a class of 71 students: exam 1. The maximum possible score was 92.

With large data sets authors often don’t show the real data as a histogram but instead show a smoothed histogram or a curve – in technical language, an estimated density function – such as Figure 7 below.

Figure 7. A possible theoretical histogram representing predictions for the Figure 6 data.

This is an instance of

The smooth density idealization. Many statisticians implicitly believe that associated with actual data (such as in Figure 6) is some theoretical smooth histogram (such as in Figure 7), by analogy with the data and theoretical histograms in Figure 5.

At this point we have entered philosophically contentious territory – what precisely does the theoretical histogram mean? – but for now let’s just think of it as showing the “best possible prediction” of the proportion of the class who will get each particular score. There is no reason to think that score of 47 will be substantially more or less frequent than a score of 51 – so the theoretical histogram will be smooth – but good reason to believe either of those scores is much more likely than a score of 2 or 89 – so it’s not uniform over the entire range.

As discussed in Chapter xxx, in the context of a repeatable experiment one can think of the theoretical histogram as a “law of averages” limit that one could estimate by repeating the experiment many times. But a particular exam given to a particular class of students clearly isn’t a repeatable experiment. In such non-repeatable contexts, the smooth density idealization is – at first sight, anyway – just a conceptual idea, not amenable to empirical confirmation. For how could one make quantitative predictions, when we have to deal with some
unknown distribution? But after a little thought it turns out that
(a) one can indeed make certain quantitative theoretical predictions; and
(b) these predictions usually check out as well as could be expected.

Here are two examples of such quantitative predictions, described in the
context of our exam data, but then evaluated in other data sets too.

2.5.1 Predictions of the smooth density idealization

Coincidences and near misses. For a given pair of students, they might
get the same score (call this a coincidence) or they might get consecutive scores
(call this a near-miss). So we can count the number of coincidences (that is, the
number of pairs of students with the same score) and we can count the number
of near-misses. The theoretical prediction is

\[
\text{in a typical simple data set, the number of near-misses will be about twice the number of coincidences.}
\]

In the data exam.1 there are 49 pairs representing coincidences and 86 pairs
representing near-misses. So the prediction works pretty well.

xxx Table of predictions/results in other data sets.

The math argument. If student A scores (say) 45 then (by supposition) the
chances of student B scoring 44 or 45 or 46 are approximately equal, so the
chance of a near miss is approximately twice the chance of a coincidence.

The small print. A certain type of reader\(^\text{12}\) reads solely for the purpose of
spotting a perceived mistake and then gleefully informing the author. Such a
reader may be thinking: if we multiply all exam scores by 2 then there will be no
near-misses, so the prediction above is nonsense! But in this case the “smooth
distribution” supposition is obviously unreasonable.

Least significant digit. For a score of 57 the “most significant” first digit is
5 and the “least significant” second digit is 7. Looking at the most significant
digit of the exam.1 data (as one might do for assigning letter grades) we clearly
are going to see substantial non-uniformity, and indeed we do

<table>
<thead>
<tr>
<th>first digit</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequency</td>
<td>1</td>
<td>3</td>
<td>10</td>
<td>15</td>
<td>18</td>
<td>17</td>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>

If (with less motivation) we look at the second digit, there is a theoretical
prediction:

\[
\text{in a typical simple data set, the distribution of least significant digits will look like random picks from the uniform distribution.}
\]

In exam.1 they do:

\(^{12}\text{In my field, Kai Lai Chung was notorious for this.}\)
xxx Table of predictions/results in other data sets.

The math argument. For 3 we add the frequencies of . . . 43, 53, 63, . . . and for 4 we add the frequencies of . . . 44, 54, 64, . . . ; by supposition the probabilities being added are approximately equal.

The small print. As above, the supposition of a smooth theoretical distribution must be plausible. And obviously if all data values are within 7 of each other the prediction can’t be correct, so we need a condition of the form “the spread of data is not small relative to 10”. Measures of spread are a textbook topic; let’s use the interquartile range (the difference between the 25th percentile and the 75th percentile), which in exam.1 is 62 - 43 = 19. Rather arbitrarily, let’s say the prediction should be used only when the interquartile range is more than 15.

2.6 The local uniformity principle

The predictions above are “first order” in that they predict “approximate” values for a statistic, without saying how close an approximation might be expected. Statisticians often implicitly rely on the following principle to make stronger predictions.

\[
\text{The local uniformity principle. For simple data sets, quantitative predictions based on the supposition that the data arises from independent random picks from some unknown “smooth” distribution are usually approximately correct, even when the “independent random picks” part of the supposition is unrealistic.}
\]

At first sight this looks both vague and implausible, so it needs some explanation. For “simple data set” imagine just a list of numbers – the exam scores in Figure 6, or the x-coordinates of the darts throws earlier, or the numbers of wine cases in the next section. The distinction from the previous “smooth density idealization” is that we now suppose the data is like independent random picks from some smooth distribution (the technical term for a “set of probabilities” like those shown on the left of Figure 5). For the darts data this is a reasonable supposition – we are literally repeating the same action on each throw – but data for exam scores or wine case production is clearly not random in the same way, merely “unpredictable” in some vague way.

As one test of the realism of the principle, note that it gives predictions for the local irregularity of observed frequencies. Write \( f_i \) for the frequency of \( i \) in the data. Then “smoothness” says that on average \( f_i \) should be close to \((f_{i+1} - f_{i-1})/2\), and the squared difference \([f_i - (f_{i+1} - f_{i-1})/2]^2\) is a measure of local irregularity. This suggests looking at the statistic

\[
S = \sum_i [f_i - (f_{i+1} - f_{i-1})/2]^2.
\]
As outlined below, the local uniformity principle predicts that $S$ should be around $3n/2$ when $n$ is the number of observations.

xxx get data to see how the prediction works; start with exam.1

The math argument. Smoothness says that the probabilities $p_i$ satisfy $p_i \approx (p_{i+1} - p_{i-1})/2$; “random picks” says that the random variables $f_i$ should be approximately Poisson distributed with mean $np_i$. This gives

$$E[(f_i-(f_{i+1}-f_{i-1})/2)^2] \approx \text{var } [f_i-(f_{i+1}-f_{i-1})/2] \approx n[p_i+(p_{i+1}+p_{i-1})/4] \approx np_i \times 3/2.$$

But isn’t this all just noise? A statistician might react to this section by saying “look, what’s important in exam score data are summary statistics like average, spread, unimodality versus bimodality, and their real-world significance as judged by implications about what students have learned and comparisons with other classes. Your predictions just involve the noise – the unimportant microscopic randomness – in the data.” And that’s perfectly true. However, even noise has its story, and the next two sections indicate more interesting aspects.

Relation to the fine-grain principle. It’s worth mentioning that the current “local uniform randomness principle” is connected to the previous “fine-grain principle” for physical randomness. In the latter context we are assuming that the data of the initial conditions in repetitions of the experiment satisfies the local uniform randomness principle, and then regularity of the pattern (determining outcome in terms of initial conditions) allows us to make predictions in the same way as in this section.

2.7 Asteroid near-misses

99942 Apophis [?] is a 350-meter long asteroid which is confidently predicted to pass Earth just below the altitude (35,000 km) of geosynchronous satellites (which provide your satellite TV) on Friday, April 13, 2029.

This fact prompts discussion of the chances of an asteroid collision with Earth. The actual density of different sized asteroids at different points in the solar system is of course an empirical issue. But it’s perfectly reasonable to make the “local uniformity” assumption that the chance of an asteroid being at one point on the Earth’s orbit is not substantially different from the chance of it being at a different point a quarter million miles away. Then mathematics, and the empirical fact that the ratio (radius of Moon orbit)/(radius of Earth) is approximately 60, shows (see Notes: 3,600 arises as $60^2$)

Amongst asteroids which pass closer to earth than the Moon’s orbit, about one in 3,600 will hit Earth.

Hypothetically, if astronomers could and did detect all asteroids of diameter greater than 50 meters passing within the Moon’s orbit for a period of years,
and found there were on average 3.5 per year, then one could infer that such an asteroid would hit the Earth about once every 1,000 years on average. 

xxx nice combination of model and data

xxx try with auto near-accidents but hard

## 2.8 Benford’s law

*benford.1* is data on total production (number of cases) of each of 337 wines reviewed by Wine Spectator magazine in December 2000 (this data is not claimed to be representative of all wine production). So the data is a list like 517, 5300, 1490, . . . ; the minimum was 30 cases and the maximum was 229,165 cases. Figure 8 shows a histogram of this data. Note the log scale on the horizontal axis, used to fit such widely varying data onto one figure.

![Histogram of wine-case data](image)

**Figure 8.** The wine-case data, *benford.1*.

For obvious economic reasons, few wines have production levels of 1 case or 1 million cases, and one might identify less obvious practical reasons for other features of the data. It may seem surprising that such data could be used to illustrate any general principle, but it can. Look at the first digit of each number in the data – so that 45 or 4,624 or 45,000 are each counted as “4”.

---

13This conclusion is a typical estimate in the literature [?] but was not obtained in this way
14“Medium-bodied and a bit rustic, but a good everyday quaff”
15Here one needs to be careful about the precise definition of a histogram, which represents data by area. See Chapter 3 of [?].
There is a theoretical prediction, called *Benford’s law*[^1], for the frequencies of the 9 possible first digits in data like this. Table 1 shows the predictions, derived from a formula written later.

<table>
<thead>
<tr>
<th>first digit</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequency</td>
<td>.3010</td>
<td>.1761</td>
<td>.1249</td>
<td>.0969</td>
<td>.0792</td>
<td>.0669</td>
<td>.0580</td>
<td>.0512</td>
<td>.0458</td>
</tr>
</tbody>
</table>

Table 1. What Benford’s law predicts.

The surprise is that theory doesn’t predict equal frequencies, but instead predicts that 1 should appear much more often than 9 as a first digit. Figure 6 shows that for our wine-case data the prediction is fairly good – certainly much better than the “equal frequency” prediction which would imply a flat histogram.

![Figure 9. Benford’s law and the wine-case data. The thick lines show the histogram for first digit in the wine-case data; the thin lines are the histogram of frequencies predicted by Benford’s law, Table 1.]()

xxx other data-sets; show only proportions of 1-2 to avoid chi-square test?

I am inclined to use Benford’s law as our first instance in the *pretty shell* category. Though striking and memorable, upon a little reflection we realize it’s just a lightly disguised instance of the ideas discussed earlier in this chapter.

*The math argument (in more detail than usual, but only for the interested reader. Look again at Figure 8. What parts of the horizontal axis correspond to case numbers with first digit 1? That’s easy to picture.*

![10 20 100 200 1000 2000 10000 20000 100000 200000](image)

Compare with the exam score data, where we look at a given *second* digit, say 2:

![20 30 40 50 60 70 80](image)

[^1]: Benford’s law is a common topic in popular science style accounts of probability, and the Wikipedia article [?] is pretty good.
Both pictures show a “repeating pattern” on the one-dimensional line, analogous to the two-dimensional repeating pattern in Figure 1 (playing cards on the wall), whose “basic unit” is shown on the left below.

In each case, a local uniformity assumption implies that the proportion of \( \text{xxx} \) in the marked \( \text{xxx} \) will just be the \( \text{xxx} \). Since we’re working on a log scale, the length of this line from 1 to 10 is \( \log 10 \), which (interpreting “log” as “log to base 10”) equals 1; and the part of the line from 1 to 2 has length \( \log 2 \), which works out to be 0.301. Similarly, for each possible first digit \( i \) there is a repeated unit, as shown on the right in the diagram above, and we get the formula

\[
\text{predicted frequency of } i \text{ as first digit} = \log(i + 1) - \log i
\]

which gave the numbers in Table 2.

The small print. As at \( \text{xxx} \), local smoothness (on the log scale) and sufficient spread, which (copying \( \text{xxx} \) but undoing the log transformation) becomes

\[
(75\text{th percentile})/(25\text{th percentile}) \geq 30
\]

So the key requirement for the plausible applicability of Benford’s law is that the data be widely varying – that it not be too uncommon to find two numbers in the data where one number is more than 30 times the other.

2.9 Notes on Chapter 2

The thumbtack data. Dropping thumbtacks onto your open palm shows they fall point-up, regardless of initial orientation, so the randomness comes from the bounce rather than the flight. On a hard tile floor you can see and hear that they bounce several times; on the rug you can’t hear but visually they appear to bounce only once. I would have expected the probabilities to vary amongst different surfaces, but the data is very close; perhaps a much longer experiment would be needed to detect a difference.

A somewhat vague principle from physics is that the “lower energy” configuration should be more likely, and indeed the “point up” configuration has lower center of mass (i.e. lower potential energy). This principle gives predictions for the direction of change of probability as one varies the shape (pin length, head diameter, etc) of the thumbtack, and one could in principle do experiments to test these predictions.

The darts data. Figure 2 shows the vertical spread is considerably larger than the horizontal spread, as one would predict (there are more variables involved with vertical placement). The data isn’t a very good fit to bivariate Normal, even though textbooks often use dart throws or target shooting as a hypothetical illustration of bivariate Normal.
Real coin tossing. In a fascinating 2007 paper by Diaconis - Holmes - Montgomery [?], high-speed photography of coin tosses shows that in fact the axis of rotation is typically not horizontal and that the axis precesses during the flight. This leads to a theoretical prediction that a tossed coin should land the same way up as it was thrown with probability about 51%. To detect this effect, one would want about 40,000 tosses, and at time of writing no-one has done the experiment.

Asteroid near-misses. xxx not assume directions uniform in 3d

Benford’s law. Our explanation isn’t logically complete, because we didn’t justify starting out by transforming to the log scale. A popular explanation amongst mathematicians and physicists invokes scale-invariance, which supposes the data measures some physical quantity relative to some conventional unit (e.g. the wine case unit equals 9 litres) and then shows that, if there is a limit distribution that does not depend on choice of unit, then the limit must be Benford’s distribution. But this conceals more than it reveals; it conceals the fact that if we don’t have increasing spread then typically Benford’s law will not hold because there is no limit distribution. Also, because the argument applies to (for instance) areas of countries but not populations of countries, it implicitly predicts that Benford’s law should be more accurate for physical data than for purely numerical data, for which I know no evidence. Our argument implicitly gives a different prediction:

For each data set in a collection calculate

\[ y = \text{an estimate of the difference between the true first-digit distribution and the Benford distribution (after subtracting sampling variation);} \]

\[ x = \log(\text{interquartile range}) \text{ for original data.} \]

Prediction: plotting \((x, y)\) for the different data sets, you will see \(y\) tending to decrease as \(x\) increases.