# Lecture 13: Physical Randomness and the Local Uniformity Principle

David Aldous

March 16, 2016

A wonderful book *The Physics of Chance* by Charles Ruhla treats a cross-section of topics with simple undergrad math. I don't do much physics in this course because it's not easy to get interesting new data. For instance, Maxwell's theory (1860s) of a "perfect gas" predicts that the velocities of molecules have a 3-dimensional Normal distribution. Ruhla describes an experimental device used to confirm this prediction.
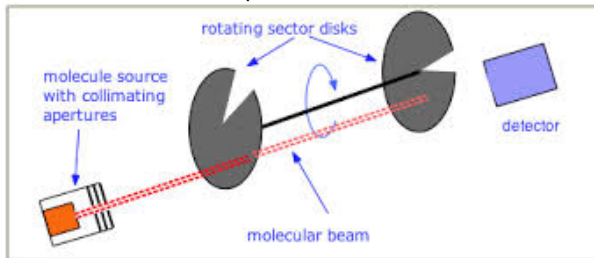
But we can't do such experiments in this course.

This lecture starts with the idea

- Chance as uncertain initial conditions in deterministic processes.

This is usually studied as "chaos" – many non-technical accounts such as *Chaos: Making a New Science* by James Glick. But hard to get data. Instead we will then move on to

- Broader uses of the **smooth density idealization** for data.
- Card shuffling, as the simplest-to-analyze instance of physical mixing.

We start with something very simple for which we can get data – dart throws.
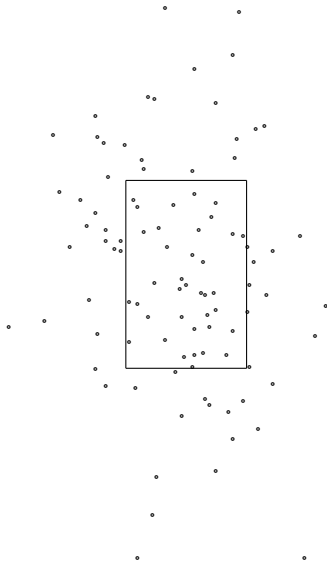
99 dart throws, centered on a 2.25" × 3.5" playing card. By a student Beau La Mont.

(i) Darts provide a vivid and quantifiable instance of the luck-skill combination (recall Lecture 8).

(ii) We can quantify skill by how often you can hit a specified target region (card), or by SD of distance from target point (center of dart board).

(iii) It seems perfectly reasonable to model dart throws by an individual via a probability density function (different for different individuals).

(iv) It seems reasonable to model successive throws as independent (after warm-up, before getting tired/bored).

So we have a convincing model for dart throwing; is there something interesting to do with the model? Want to use this "real data" from Beau instead of some theoretical probability distribution (bivariate Normal or uniform).

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 2. 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.
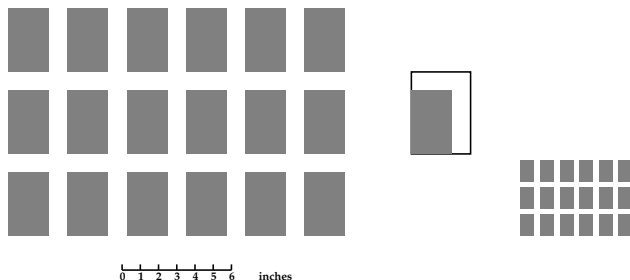

0 1 2 3 4 5 6   inches

Figure 2. 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.

Instead of doing time-consuming experiments with dart-throwers of different skills, I will be lazier and work with the previous data set of 100 throws by Beau, 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 of hitting such a card 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.
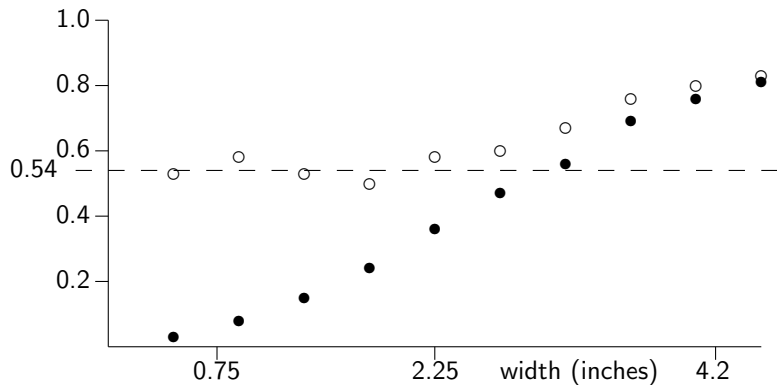
Figure 3. In the setting of Figure 1, Beau's estimated probability of hitting a specific card • and the probability of hitting some card ○, as a function of width. A small postage stamp has width about 0.75, a playing card 2.25, and a cheap small paperback book 4.2.

Returning to the fairground game, we imagine scaling the pattern (as on the right of Figure 2). For normal size cards, Beau would have 58 hits (that is, 36 on the aimed-at card due to skill, and 22 on a different card due to luck) 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.

The key point 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.

- 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 scaling (shrinking cards and spaces together).

- So a dart thrown blindly, without propensity to hit or to miss cards, should have a 54% chance to hit a card.

- Even when we aim, if 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.

Abstractly, this is an example of what I call the *fine-grain principle*.

The physics of coin-tossing.

Why do we think that a tossed coin should land Heads with probability $1/2$? Well, the usual argument by symmetry 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 $p = 1/2$.

Like most people, I find this argument (and the corresponding argument for dice, roulette etc) convincing. But such an argument by logic doesn't give much insight into where physically the number $1/2$ comes from. What is the connection with physics?

Digression; Here are two unusual cases where such an "argument by symmetry" goes wrong.

[in class demo]

After a moment's thought, in this case the physics is actually quite simple, if we over-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.
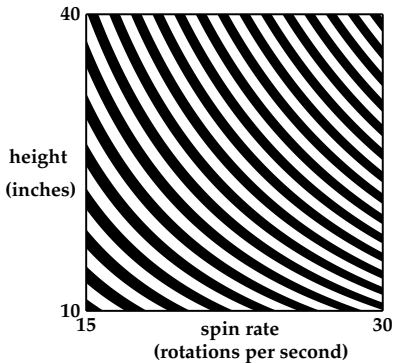
Figure 4. Phase space for coin tossing. The shaded bands are where an initially Heads-up coin will land Heads, as determined by the height and rotation rate of the toss. Each band indicates a specific number of rotations, from 7 to 27 over the region shown. The formula underlying Figure 4 is as follows. 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 $t = \sqrt{8h/g}$. 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 $r\sqrt{8h/g} = n \pm \frac{1}{4}$.

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 or Tails. I 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 very 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 1. 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.

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. A particular outcome corresponds to the initial conditions being in some particular subset of phase space, which we visualize as a collection of clumps such as the rectangles in Figure 2 or the curved bands in Figure 4. If the subset of phase space has a certain kind of regularity – that the local proportion of phase space that lies in the subset is approximately the same proportion p regardless of global position within phase space – then we can confidently predict that the probability of the outcome is about p, provided the spread of the distribution of the initial point is at least somewhat large relative to the distance between clumps. The numerical value p comes from the *pattern* in phase space, 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 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 2 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 actual reason we all believe that a die lands 5 with probability 1/6 is the argument by symmetry (and a supposition that other people have checked it empirically).

- We can't honestly "do the physics" to confirm this via the fine-grain principle, but we have a world-view (i.e. a belief) that it would be confirmed if we could.

Digression: 40,000 coin tosses

[show]

The fine-grain principle answers a philosophical question: a deterministic system may be perceived as random because we cannot see/control initial conditions precisely. Many popular science books on "Chaos" pursue that theme.

As our next topic, we can use essentially the same idea to say something about observational data unrelated to physics.

Figure 6 is a histogram of actual "course scores" for a class of 71 students.
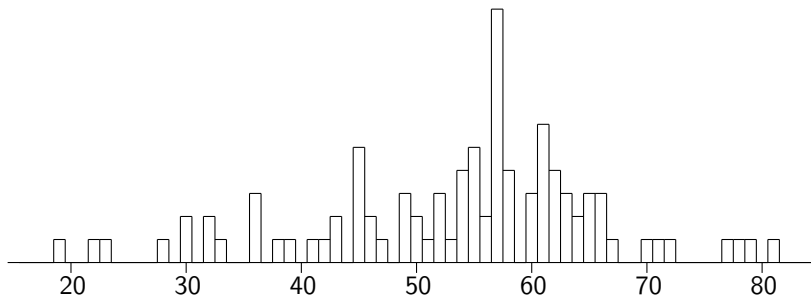

Figure 6. Scores for a class of 71 students. The maximum possible score was 92.

With large data sets authors often accompany the real data histogram with a smoothed curve.
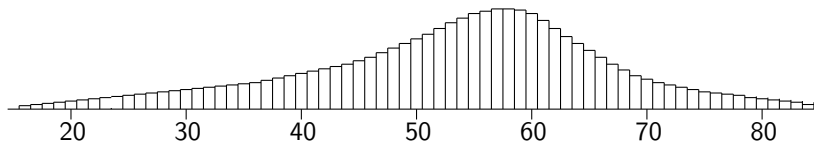

Figure 7. A theoretical histogram one might associate with the Figure 6 data.

And statisticians often regard the data as IID random samples from some unknown smooth density, in which case there is a theory of *density estimation*. Let me call these assumptions **the smooth density idealization.**

But conceptually it is hard to understand why this idealization is reasonable. Data like exam scores, the areas of the 50 States, or the wine-case data later, are "observational", not random samples of anything.

Instead of worrying *a priori* about this conceptual issue, we could ask whether observed data is **consistent** with this assumption. To do this we need to find some quantitative theoretical predictions for such data, which (to be useful) should not depend on the unknown distribution. Let me give two simple examples of such quantitative predictions, described in the context of our exam data.

A **project** is to repeat this kind of analysis for other data sets.

Checking predictions of the smooth density idealization.

Prediction 1: 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

if a data set is associated with a smooth distribution, then the number of near-misses will be about twice the number of coincidences.

In the data of Figure 6 there are 49 pairs representing coincidences and 86 pairs representing near-misses. So the prediction works pretty well. *The math argument.* If student A scores (say) 45 then (by supposition of smooth distribution) 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.

Prediction 2: A test statistic.

As above, imagine data of the form

$$N_i = \text{number of observations of } i$$

with $\sum_i N_i = n$. Suppose we want to test the null hypothesis that the data is consistent with being from IID samples from some unknown "smooth" probability distribution (p_i). One test relies on "local smoothness", interpreted as

$$p_i \approx (p_{i-1} + p_{i+1})/2.$$

[continue on board: leads to test statistic]

$$S = \sum_i (N_i - \tfrac{1}{2}(N_{i-1} + N_{i+1}))^2$$

which under the null hypothesis has $\mathbb{E}S \approx 3n/2$.

Project: try on different data-sets.

Prediction 3: 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 Figure 6 data (as one might do for assigning letter grades) we clearly are going to see substantial non-uniformity, and indeed we do

first digit	1	2	3	4	5	6	7	8
frequency	1	3	10	15	18	17	6	1

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

if a data set is associated with a smooth distribution, then the distribution of least significant digits will be approximately uniform.

In this data set it is:

| second digit	0	1	2	3	4	5	6	7	8	9
frequency	6	8	10	6	4	11	10	7	3	6

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 this data set is 62 - 43 = 19. Rather arbitrarily, let's say the prediction should be used only when

$$\text{the interquartile range is more than 15.} \qquad (1)$$

Benford's law.

Hiistogram for a data-set of the 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. Note the log scale on the horizontal axis, used to fit such widely varying data onto one figure.

Number of cases (log scale). $K = 1000$.
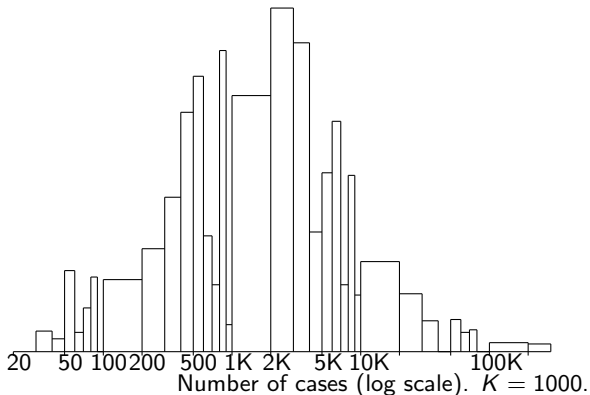
For obvious economic reasons, few wines have production levels of 1 case or 1 million cases, and one might identify less obvious practical reasons for other features of the data. It may seem surprising that such data could be used to illustrate any general principle, but it can. Look at the first digit of each number in the data – so that 45 or 4,624 or 45,000 are each counted as "4". There is a theoretical prediction, called *Benford's law*, for the frequencies of the 9 possible first digits in data like this. This table shows the predictions, derived from a formula written later.

first digit	1	2	3	4	5	6	7	8	9
predicted frequency	.3010	.1761	.1249	.0969	.0792	.0669	.0580	.0512	.0458

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.

For our wine-case data the prediction is fairly good – certainly much better than the "equal frequency" prediction which would imply a flat histogram.
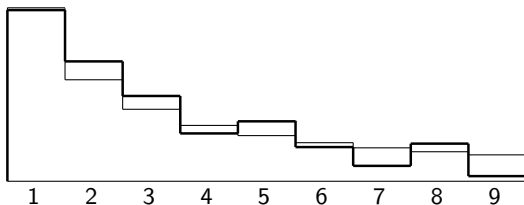

Figure 9. Benford's law and the wine-case data. The thick lines show the histogram for first digit in the wine-case data; the thin lines are the histogram of frequencies predicted by Benford's law, Table 1.

Though striking and memorable, upon a little reflection one realizes that Benford's law is just a lightly disguised instance of an ideas discussed earlier in this lecture.

The math argument. Look again at the histogram. What parts of the horizontal axis correspond to case numbers with first digit 1? That's easy to picture.
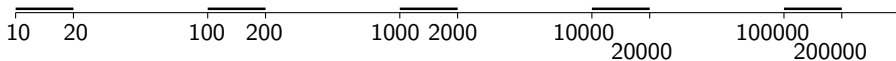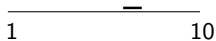

Compare with the exam score data, where we look at a given *second* digit, say 2:
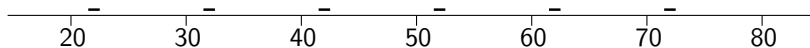
Both pictures show a "repeating pattern" on the one-dimensional line,
analogous to the two-dimensional repeating pattern in Figure 2 (playing
cards on the wall), whose "basic unit" is shown on the left below.
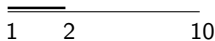
the interval from i to i+1

1   2   10   1   10

In each case, the smooth density idealization implies that the proportion of data in the marked subset will be approximately the density of the marked subset, that is the proportion of the length of the "basic unit" that is in the subset. 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

predicted frequency of i as first digit $= \log_{10}(i+1) - \log_{10} i$

which gave the numbers in Table 2.

The small print. To a mathematician, the Benford prediction is *equivalent* to the second significant digit prediction The implicit assumptions are smoothness (on the log scale) and sufficient spread, which (copying (1) but undoing the log transformation) becomes

(75th percentile)/(25th percentile) $\geq 30 \approx 10^{1.5}$.

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.

Checking Benford's law on data is a natural **project** which has been done – see posted article *When Can One Test an Explanation? Compare and Contrast Benford's Law and the Fuzzy CLT*.

Wikipedia *Benford's Law* gives the same explanation and various applications.