The Law of Small Numbers

Kevin Dorst
kmdorst@mit.edu

23.223 Rationality

I. Small numbers and representativeness

The law of large numbers: with enough trials, the sample mean is likely to be close to the population mean.

And likewise for other statistics. The proportion of samples below x, etc.

The law of *small* numbers: even small samples will be representative of the population, "in all essential characteristics".

The "essential characteristics" rider will do a lot of work for them.

More generally, T&K hypothesize that this is a consequence of a general strategy people use for estimating probabilities:

> **Representativeness Heuristic:** To evaluate the probability of an event, assess the degree to which
> (1) it is similar in essential respects to the parent population, and
> (2) it reflects the salient features of the generating process.

Examples:

- The conjunction fallacy.
- Birth orders: 72 families had GBGBBG. How many had BGBBBB?
- Preserving majority/minority.
  Program A: 65% boys. Program B: 45% boys. You walk into a classroom and see 55% boys. What's best guess for program?

Median = 30; 81% said former more likely

75% said A. Correct answer = B.
$P(X = 55|Bin(100, 0.65)) = 0.0096$,
$P(X = 55|Bin(100, 0.45)) = 0.0108$

But notice: people *don't* think (eg) the sequence HTHTHTHTHT is more likely that HHTHTTTHTH. Why not?

T&K: the outcome also needs to "appear random": "the event should reflect the properties of the uncertain process that generated it"

Which characteristics? That depends on context, but two normal ones are (1) irregularity and (2) local representativeness

1) Irregularity

- Which is more likely: BBBGGG or GBBGBG?
- Subjects not only avoid producing sequences like HHHHT, but also avoid ones like HTHTHTHT or HHTTHHTT.
- 20 marbles distributed to A, B, C, D, E. Which more likely?
  (4, 4, 4, 4, 4) or (4, 4, 5, 4, 3)?

People prefer latter

Most prefer latter.
Former is 0.3%-likely, latter is 0.25%, but if *bin* by (eg) "exactly x away from uniform", then former ($x = 0$) is still 0.3% while latter ($x = 1$) is 10.1%.

2) Local representativeness

- Expect the law of large numbers to apply to small samples as well.
- Eg gambler's fallacy ("negative recency") with coins, or birth orders.

Expect more "switches" than an IID process would produce

Let's fit this into an argument:

P1 People judge probability using representativeness (even in situations XYZ).

P2 If people were rational, they wouldn't judge probabilities using representativeness (in situations XYZ).

C People don't judge probabilities rationally.

Toy Model

You face a sequence S of 100 tosses, in 10 chunks $(C_1, ..., C_{10})$ of 10:
TTTTTHTTHH TTHTHTHTHT HHTTTHTTHH HHHTHHTTHT TTTTHHHTHT
HHHTTTTHHH HTHTHTTHHH THHHTTTHHT TTHTTTTHTH THTHHTHHTH
You (expect to) care about *proportion heads in the overall sequence*, $\overline{H}_S$.

Some chunks are more representative of $\overline{H}_S$ than others. If $\overline{H}_S = 0.48$:
$C_3 = \boldsymbol{HH}TTT\boldsymbol{H}TT\boldsymbol{HH}$ ($\overline{H}_{C_3} = 0.5$) is more representative than
$C_1 = TTTTT\boldsymbol{H}TT\boldsymbol{HH}$ ($\overline{H}_{C_1} = 0.3$).

Naive Encoding | Representative Encoding

$\mathbb{E}(\overline{\boldsymbol{H}}_S) - \overline{\boldsymbol{H}}_S$: | Estimates vs. actual values

$(\mathbb{E}(\overline{\boldsymbol{H}}_S) - \overline{\boldsymbol{H}}_S)^2$: | Mean Squared Error (MSE) (Thick lines = averages)