How can we Guess the Distribution of Wealth in Greenville, Mississippi? A Proposed Heuristic for Guessing Distributions

Yair Neuman[1], Itzhak Dayan[1]

1 Ben-Gurion University of the Negev

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.

Abstract

How can laymen approximate a real-world distribution while being “rationally bounded”? In this paper, we show that guesstimating a real-world distribution is possible by using the “True Median”. We show that choosing the best distribution among combinatorically potential distributions through the True Median criterion is a straightforward approach for choosing the distribution that *maximizes the entropy* and therefore the one that provides the least biased approximation of the real one. We conclude by proposing a rule of thumb that works as effectively as any of our proposed calculations.

Yair Neuman* and **Itzhak Dayan**

**Head, The Functor Lab, Department of Cognitive and Brain Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel. Email: yneuman@bgu.ac.il*

Keywords: guesstimation, heuristics, median, maximum entropy principle, bounded rationality.

Introduction

If you ask a baker from Paris about the scaling size of buildings in his city, he will probably find it difficult to understand the term scaling. However, rephrasing the question in layman’s terms, he may answer by saying that *few* of the buildings are tall, *some* are moderate, and *many* are small. As people naturally use a limited number of linguistic*quantifiers* (e.g., few) and draw conclusions based on heuristics rather than on scientific inference, we may question the validity of the baker’s description of buildings in Paris. However, although presented in layman’s terms, the baker’s description surprisingly corresponds with the scaling of beautiful cities (e.g., Paris) as described by Salingaros (2010). Using the term “Guesstimation”, which is a form of a wise guess that lacks sufficient information, and which is based on minimal and simplified assumptions, we may wonder whether the baker’s answer may be grounded in a clear guesstimation heuristic.

Here is another example. If you ask a lady from Greenville, Mississippi[1] about the distribution of wealth in her city, she will probably refrain from answering in terms of power law distribution and its parameters. However, rephrasing the question in layman's terms, she would probably explain that in her city *few* people are rich, *some* are moderate, and *many* are poor. Both individuals can nicely describe the real-world distribution while "computing with words" (Zadeh, 2012) meaning that they perform a kind of cognitive computation where "the objects of computation are words, phrases, and propositions drawn from a natural language" (ibid. p. 5). However, the *quantifiers* "few", "some", and "many" are not telling about the way the best distribution can be chosen among several distributions corresponding to the *same* linguistic labels. Moreover, in the above examples, both individuals rely on tiny samples. Their cognitive computation with words probably relies on the *availability heuristic* (Tversky & Kahneman, 1973) and on a tiny and non-representative sample of instances that come to mind. As the word "tiny" is in itself a vague term, we use it to describe a sample composed of seven plus or minus two instances. This number corresponds with Miller's "magic number" (Miller, 1956) and with the number of individuals composing our small social circle as identified by Dunbar (cited in West, 2017) and elaborated by West (2017). These are the individuals that probably come to mind when we perform fast and frugal judgments, like in the case of the lady from Mississippi.

The challenge we address in this paper is as follows. Given the layman's few linguistic quantifiers and a tiny sample, is it possible to identify a heuristic through which one can guess the best distribution corresponding with his description? It is important to emphasize that our study is not a psychological one and we don't study the way laymen approximate distributions (e.g., Goldstein & Rothschild, 2014). Our aim is different and may fall under the title of "Guesstimation" (Weinstein & Adam, 2008) or how approximation may be performed using only calculations that can be written on the "back of a cocktail napkin" (Weinstein & Adam, 2008). We conclude our paper, by showing that our proposed heuristic may be even reduced to a rule of thumb.

The proposed solution

To address the above challenge, we first start with combinatorics. Given n different boxes representing our quantifiers (e.g., few, some, many) and r identical balls (i.e., particles), corresponding with a non-representative and tiny sample of instances, we would like to identify the different distributions assuming *no box is left empty*. For computing the number of distributions, we use:

$$\text{Equation 1. } \binom{r-1}{n-1}$$

which is solved using the Binomial coefficient.

For example, for 3 boxes and 4 balls, we use:

$$\text{Equation 2. } \binom{4-1}{3-1} = \binom{3}{2} = 3$$

and the 3 distributions are:

1.

A	B	C
2	1	1

2.

A	B	C
1	2	1

3.

A	B	C
1	1	2

Figure 1. The first three distributions

Through combinatorics, we generate all possible distributions and first chose those corresponding with the *rank order* proposed by the quantifiers (e.g. many is higher than few). The next step is to select the best among them, the one best representing the real-world distribution.

While the above equations may be “scary” they actually rely on simple counting procedures and a limited number of relevant combinations. For instance, for five instances corresponding with the lower limit of Miller’s magic number, and three quantifiers, we have only six potential distributions, and for nine balls (i.e., the upper limit of the Magic number), we have 28 distributions. However, only some of the combinatorically generated distributions correspond with the quantifiers *few* and *many*, such as in the case *few* are rich, *many* are middle class, and *few* are poor. The quantifiers impose constraints on the ranking order of the distributions. For example, let’s assume that a layman (or laywoman) is asked about the distribution of wealth in the States. She answers by saying that *few* are poor, *many* are in the middle class, and *few* are rich. Using the above distributions, where A represents “poor”, B represents the “middle class”, and C represents “rich”, there is only one distribution corresponding with the quantifiers, which is distribution 2. It means that using a tiny sample, and a limited number of quantifiers, the number of combinatorically potential distributions is limited, and the number is even smaller when the constraint of order is imposed.

Our proposed heuristic is based on (1) the idea of the True Median (TM) (Gott et al., 2001), and (2) Jaynes’s well-known principle of Maximum Entropy (Jaynes, 1996) where "we look for a probability distribution that best describes a set of data, given some constraints related to features of this data"[2] We first present this heuristic and next draws out of it a rule of thumb that may explain a successful guesstimation.

For choosing the best distribution, or in more accurate terms: the less biased distribution among the N combinations, we use the idea of Median statistics and specifically the idea of the True Median (TM) in the population (Gott et al., 2001). As explained by Gott et al., (2001), a large number of measurements with no systematic effects naturally results in half of the measurements above the true value and half below the true value. As N tends to infinity, there is a true median. As each measurement in our tiny sample is statistically independent, it has a 50% chance of being above or below the TM. Therefore, "the probability that exactly n of our N measurements is higher than the TM is given by the Binomial distribution":

Equation 3. $P = 2^{-N}N!/[n!(N-n)!]$

Given the different distributions we may produce through combinatorics only, we can compute the Median for each distribution, and count the number of n observations that score above the Median. Next, we use Equation 3 and compute the probability that exactly these n measurements are higher than the TM. Finally, we select the distribution in which P is the highest.

The deep reason why the True Median selection criterion may work so well in choosing the best distribution is explained in terms of The Principle of Maximum Entropy. To recall, The "Shannon entropy of a distribution is the expected amount of information in an event drawn from that distribution. It gives a lower bound on the number of bits […] needed on average to encode symbols drawn from a distribution P". (Goodfellow, Bengio, & Courville, 2016, p. 74). The TM is the ultimate cut point where the Shannon information entropy (uncertainty) of the distribution is maximized, as by definition the median splits the population into two equal parts. Given the constraints imposed by the quantifiers, choosing among the combinatorically possible distributions by using the TM criterion, is choosing the *simplest* distribution using one constraint only which is the Median of the hypothesized distribution form (e.g., normal, Pareto, etc.).

We illustrate this idea through the following example. Let's assume that a layman is asked about the real distribution and believes that *many* are in the middle, *some* are in the lower and *some* are in the upper. For approximating/choosing the best distribution, we may use three boxes and a limited sample of seven individuals the layman retrieves from his memory to fill the boxes. This choice corresponds with the availability heuristic and involves a non-representative sample of individuals. Using three boxes and seven balls, two possible distributions from which we have to choose are:

A:

Lower	Middle	Upper
1	5	1

and

B:

Lower	Middle	Upper
2	3	2

Figure 2. Two distributions to chose from

Using the idea of a TM, the probability that exactly 1 observation is higher than the TM (Distribution A) is p = 0.054 and the probability that two observations are above the TM (Distribution B) is 0.164. Choosing the best distribution according to the TM criterion shows us that the chosen probability distribution (i.e., Distribution B) is the one *maximizing the entropy* as the Shannon entropy of distribution B is higher (1.56 vs. 1.15 respectively). For proving that it is the better distribution among the two, we compute the Multinomial probability of the two distributions using the known distribution of wealth[3] as described above and can see that the probability of distribution B is higher than the one of distribution A (p 0.0973 vs. 0.0799 respectively), meaning that our criterion chooses the distribution that better represents the real one. So far, and given the limited number of quantifiers and instances, our calculations can be performed on the "back of a cocktail napkin" and therefore can be described as a form of guesstimation. Next, we show how to apply the same guesstimation to a Pareto-style distribution.

Let's assume that asked about the distribution of wealth, a layman explains that few are rich, some are middle, and many are poor. Using three boxes and seven plus or minus two balls, we can generate several distributions that correspond with the quantifiers few, some, many, meaning that there must be more balls in the "many" box than in the "some" box, than in the "few" box. For example, see the following distribution for six balls:

		X
	X	X
X	X	X
RICH	MIDDLE	POOR

Figure 3. The distribution for the three quantifiers and six balls

However, in this case, there is no other six balls distribution from which we have to choose the best. However, for eight and nine balls, we have two options from which we should chose:

8a:

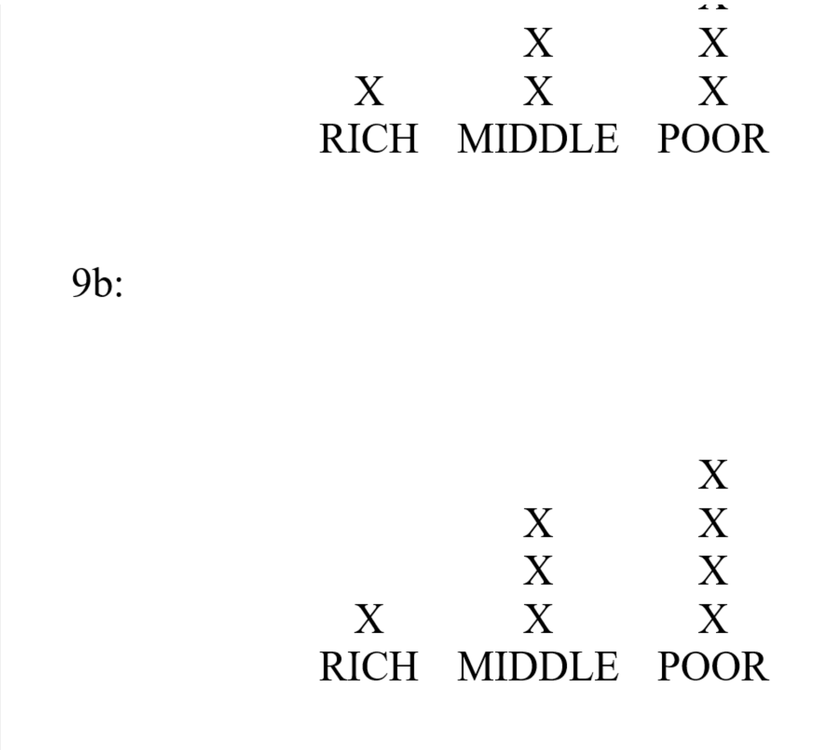
		X
	X	X
	X	X
X	X	X
RICH	MIDDLE	POOR

8b:

		X
		X
		X
	X	X
X	X	X
RICH	MIDDLE	POOR

9a:

		X
		X
		X
		X



Figures 4. Four distributions for 8 and 9 balls

The Median salary for the lower, middle, and upper income is $28,700, $86,000, and $207,400 respectively. Using these data, we can guess that each ball's salary in our rich, middle, and poor categories correspond with the above data. Approaching the distribution as a Pareto distribution, and using X_{min} = 28700, we can estimate the shape parameter α using:

$$\text{Equation 4. } \alpha = \frac{n}{\sum_i \ln(xi/xm)}$$

where x_m is the lowest value of our distribution. We can estimate the median using:

$$\text{Equation 5. } X_m \sqrt[\alpha]{2}$$

and the Entropy:

$$\text{Equation 6. } \log\left(\left(\frac{\mathrm{Xm}}{\alpha}\right)\right) e^{1+\frac{1}{\alpha}}$$

Using these estimators, we can compute the Median for each distribution, the alpha parameter, the probability that n "successes" (i.e., rich) are above the Median, the entropy, and the estimated percent of wealth held by the upper 20% of the population. Moreover, according to the Federal Reserve, the upper 20% of the population in the States holds 88% of the wealth (Bricker et al., 2020). Therefore we can use the following equation:

Equation 7. $$\frac{\log 5}{\log \frac{x}{0.20}} = \alpha$$

to estimate the percent of wealth held by the upper 20% of the population.

The next table presents the results with the chosen distributions marked:

Table 1. The results for the four distributions

Distribution	M	α (estimated)	P	Entropy	Wealth
8a	45282	1.52	0.27	4.99	57%
8b	41178	1.92	0.22	4.83	46%
9a	39559	2.16	0.16	4.76	42%
9b	50193	1.24	0.25	5.14	73%

In this case, and by using the TM criterion, we can see that the chosen distributions among each of the two options are those maximizing the entropy and those that better approximate the percent of wealth held by the upper 20% of the population according to Pareto distribution. It is important to emphasize, that the hypothesized distribution is approximated through the use of linguistic quantifiers only. In the first example, we selected the median using the linguistic description that many are “middle” some are “lower” and some are “upper”. In the second example (i.e., the lady from Mississippi) we approximated a Pareto-style distribution by using the linguistic quantifiers: few, some, and many. So far, we have relied on calculations, but the layman may ask whether we may propose a rule of thumb that is even easier than our napkin calculations. A possible rule of thumb may be as follows:

Given N distributions from which you have to choose, simply choose the one which is more “heterogenous”.

or in different terms: Given N distributions corresponding with your quantifiers, and from which you have to choose, select the distribution where the instances surrounding the median are more uniformly distributed. Using this rule-of-thumb, there is no need for the above calculations.

Discussion

This short paper presents a simple albeit powerful heuristic for guessing the least biased (i.e., best) distribution using a combination of computing with words, basic combinatorics, and the true median as a selection criterion for choosing the distribution that maximizes the entropy under the single constraint which is the chosen Median. Moreover, for the layman, we may propose a rule of thumb where among the combinatorically possible distributions, that correspond with the constraints imposed by the quantifiers, s (he) just has to choose the more uniformly distributed configuration around the median, as this distribution maximizes the Shannon information entropy. Returning to the TM, the deep reason why the

True Median selection criterion works so well is that the median is the ultimate cut point where the Shannon information entropy (uncertainty) of the distribution is maximized, as by definition that median splits the population into two equal parts. The principle of maximum entropy (Jaynes, 1996) suggests that among all possible distributions and given some minimal constraints, we should choose the distribution that maximizes our uncertainty and ignorance, as this is the *least biased distribution*. The best distribution is therefore the least biased. Therefore, selecting the distribution according to the True Median criterion is a straightforward heuristic for choosing the distribution maximizing the Shannon entropy and therefore the uncertainty, given the minimal constraints imposed by the quantifiers and the choice of the relevant median. Choosing among the combinatorically possible distributions by using the TM criterion, is therefore choosing the *simplest* distribution using one constraint only which is the Median of the hypothesized distribution form as inferred from the quantifiers (e.g., normal, Pareto, etc.). This procedure can even be simplified by using the abovementioned rule of thumb, where through eye googling the layman selects the more uniform distribution. In sum, and given Simon's idea of bounded rationality (Simon, 2000), we may understand how individuals modeled as bounded rational agents may gain a reasonable approximation of a real-world distribution even when using a tiny sample and a few categories selected through linguistic quantifiers.

Bionote

Yair Neuman (b. 1968) is a full Prof. at the Department of Cognitive and Brain Sciences, Ben-Gurion University of the Negev, and the head of the Functor Lab. He is the author of numerous papers and ten books and was a Visiting Prof. at MIT, Oxford, U. Toronto, and Weizmann Institute of Science.

Itzhak Dayan (b. 1946) earned his Ph.D. in Operations Research (Hebrew University, Jerusalem) and works as a statistics consultant to various companies in the environmental and transportation fields.

Footnotes

[1] https://www.djournal.com/news/state-news/this-is-the-poorest-city-in-mississippi/article_fd23130e-d0e4-5653-9f10-e316a97640b9.html#:~:text=Among%20the%2014%20places%20in, of%20what%20people%20can%20afford.

[2] https://www.complexityexplorer.org/news/21-simon-dedeo-talks-about-his-maxent-tutorial

[3] https://www.pewresearch.org/fact-tank/2022/04/20/how-the-american-middle-class-has-changed-in-the-past-five-decades/

References

- Bricker, J., Dettling, L. J., Henriques, A., Hsu, J. W., Moore, K. B., Sabelhaus, J.,... & Windle, R. A. (2020). Changes in US family finances from 2016 to 2019: Evidence from the Survey of Consumer Finances. Federal Reserve Bulletin, 106(5), 1-38.

- Goldstein, D. G., & Rothschild, D. (2014). Lay understanding of probability distributions. *Judgment and Decision making*, *9*(1), 1-14.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Gott III, J. R., Vogeley, M. S., Podariu, S., & Ratra, B. (2001). Median statistics, h0, and the accelerating universe. *The Astrophysical Journal*, *549*(1), 1.
- Jaynes, E. T. (2003). *Probability theory: The logic of science*. Cambridge university press.
- Miller, G. A. (1956). The magic number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 91-97.
- Salingaros, N. A. (2010). *Twelve Lectures on Architecture: algorithmic sustainable design*. BoD–Books on Demand.
- Simon, H.A. Bounded rationality in social science: Today and tomorrow. *Mind & Society* **1**, 25–39 (2000). https://doi.org/10.1007/BF02512227
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, *5*(2), 207-232.
- Weinstein, L., & Adam, J. A. (2008). Guesstimation: Solving the world's problems on the back of a cocktail napkin. Princeton, NJ: Princeton UP.
- West, G. B. (2017). *Scale: The Universal Laws of Growth, Innovation, Sustainability, and the Pace of Life, in Organisms*. PENGUIN Books.
- Zadeh, L. A. (2012). *Computing with words: Principal concepts and ideas* (Vol. 277). Springer.