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 a layman choose the least biased hypothetical distribution for approximating a real distribution? In this paper, we propose a simple rule of thumb for addressing this challenge. This rule of thumb involves choosing the distribution that maximizes the entropy under the constraint of rank order. For explaining this rule of thumb and its justification, we first show that “guesstimating” a real-world distribution is possible by using the “True Median”.

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

This paper emerged from an interest in the way both human and non-human organisms may apply simple models and heuristics under the constraints of bounded rationality (e.g., Neuman, Cohen, and Tamir, 2021). For informally presenting the context of our paper, we open with imagined examples and then move to formalities.

Imagine you ask a baker from Paris about the scaling size of buildings in his city. The baker 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 the availability heuristic (Tversky & Kahneman, 1973) 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).

Here is another imagined example. If you ask a lady from a poor city about the distribution of wealth in her city, she will probably refrain from answering in terms of power law distribution and its parameters. However, answering in layman's terms, she would probably explain that in her city *few* people are rich, *some* are moderate, and *many* are poor.

In both cases, the individuals could have answered by seeking a reliable scientific source or by applying a scientifically grounded procedure. However, this is not the context of our study. We are interested in the way laymen who use linguistic quantifiers may choose the least biased hypothetical distribution for approximating the real one, without applying formal scientific procedures.

Both of the above-mentioned individuals seem to 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, here is the problem: The *quantifiers* "few", "some", and "many" are not telling about the way the best distribution can be chosen among several hypothetical distributions corresponding with the *same* linguistic labels. Moreover, in the above examples, both individuals rely on tiny hypothetical distributions. As the word "tiny" is in itself a vague term, we use it to describe a distribution 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 make fast and frugal judgments.

The challenge that we address in this paper is as follows. Given the layman's few linguistic quantifiers, is it possible to identify a rule of thumb through which he can choose the best or the least biased hypothetical distribution corresponding with his description? It is important to emphasize that our aim is not:

1. To propose a new scientific way of estimating a distribution, and not
2. To study the way laymen actually approximate distributions (e.g., Goldstein & Rothschild, 2014).

We aim to propose a rule of thumb explaining how to perform a guess which is scientifically grounded. For proposing this rule of thumb, we first present a scientifically grounded procedure for guessing. The aim of this scientifically grounded guessing procedure is to explain the deep scientific rationale underlying our proposed rule of thumb.

The true median and how to use it for guesstimation

To explain the proposed rule of thumb, we first start with combinatorics. Given n different boxes representing our quantifiers (e.g., few, some, many) and r identical balls (i.e., particles), we would like to identify the different possible distributions of the balls assuming *no box is left empty*. For computing the number of distributions, we use the following equation:

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

which is solved using the Binomial coefficient. For example, for 3 boxes (i.e., quantifiers) and 4 balls (i.e., instances of the distribution's particles), we use:

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

and the three possible 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. Three possible distributions for 3 boxes and 4 balls

Through combinatorics, we may generate all possible distributions and first choose those corresponding with the *rank order* proposed by the quantifiers (e.g. many is higher than few). For example, if A and C correspond with "few" and B with "many" then the only relevant distribution is distribution 2, as this distribution preserves the rank order where "many" is higher than "few". In distribution 1, the number of "few" particles is higher than the number of "many" particles (2 vs. 1 respectively) which is incoherent with the rank order imposed by the quantifiers. This is a simple instance, however, there are cases where several hypothetical distributions correspond with the rank order imposed by the quantifiers, and one has to choose between them. To repeat, the quantifiers impose constraints on the ranking order. 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 number of instances, 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.

The layman may therefore work along this line of reasoning. First, he may use a limited number of linguistic quantifiers representing the values of a discrete random variable. Second, he may use a limited number of instances (i.e., 7 plus or minus 2) to form the *hypothetical* distributions corresponding with the rank order imposed by the quantifiers. Finally, if there are several hypothetical distributions from which he must choose, then he uses our rule of thumb for choosing the distribution that best approximates the real one.

First, let us propose a formal procedure for choosing between several *hypothetical* distributions corresponding with the constraints imposed by the quantifiers. Our proposed procedure 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"[1] 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 least biased distribution[2] among the N hypothetical distributions, 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":

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

Given several hypothetical distributions from which we have to choose, 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. In other words, we have several hypothetical distributions composed of a few "boxes" and a few "balls". We would like to choose from these distributions the one which is the least biased and therefore the best guess for the real distribution. According to the above proposal, we should compute the median of each distribution, and through the idea of the true median, simply choose the distribution in which the probability of having n particles (i.e., balls) above the median is the highest.

The scientific justification for the TM criterion

The deep reason why the True Median selection criterion may work so well in choosing the best hypothetical distribution for approximating the real 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 several hypothetical 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 distribution of wealth in the States 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, he may use three boxes and a limited number of seven "balls". Using three boxes and seven balls, two hypothetical distributions from which he has to choose are:

A:

Lower	Middle	Upper
1	5	1

and

B:

Lower	Middle	Upper
2	3	2

Figure 2. Two distributions to choose from

To find the median of this discrete random variable (i.e., the distribution), we need to follow the following steps:

1. Determine the probability mass function (PMF) of the random variable.
2. Sort the values of the random variable in ascending order.
3. Calculate the cumulative probability distribution function (CDF) of the random variable.
4. Identify the smallest value of the random variable such that its cumulative probability is greater than or equal to 0.5. This value is the median.

The PMF of our random variable is:

X	Lower	Middle	Upper
P (X)	1/7	5/7	1/7

The CDF of the random variable is:

X	Lower	Middle	Upper
P (X)	1/7	6/7	1

The smallest value of X such that its cumulative probability is greater than or equal to 0.5 is “Middle”, since P (X<=Middle) = 1/7 + 5/7 = 6/7, which is greater than 0.5. Therefore, the median of this discrete random variable is “Middle”.

Using the idea of the 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. Therefore, choosing the best hypothetical distribution according to the TM criterion shows us that the chosen hypothetical 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 may compute the Multinomial probability of the two distributions using the known distribution of wealth in the States[3] 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 (Weinstein & Adam, 2008).

In the above example, we used a distribution roughly corresponding to normal distribution. However, the same approach may be applied to a Pareto-style distribution. We explain. Let’s assume that asked about the distribution of wealth in the States, another layman explains that a few people hold the majority of the wealth, some are holding less amount of wealth and many hold a minor fraction of the wealth. Using three boxes, and seven plus or minus two balls, he 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 where RICH, MIDDLE and POOR correspond with our layman naïve conception of wealth distribution:

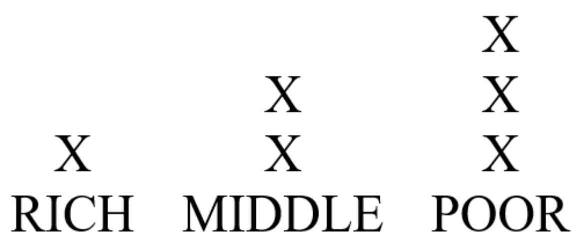


Figure 3. The distribution for the three quantifiers and six balls

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

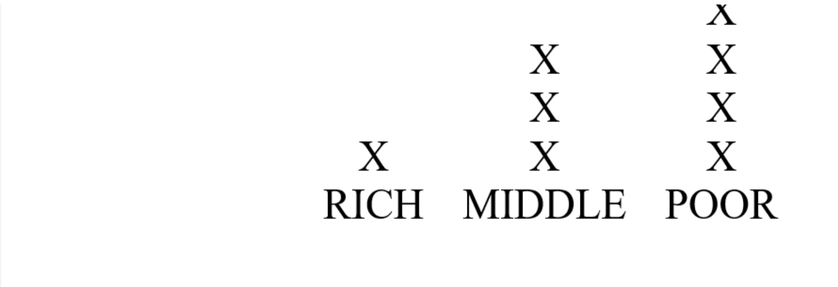
8b:

		X
		X
		X
	X	X
X	X	X
RICH	MIDDLE	POOR

9a:

		X
		X
		X
		X
	X	X
X	X	X
RICH	MIDDLE	POOR

9b:



Figures 4. Four distributions for 8 and 9 balls

For illustration, let's assume that the salary for the lower, middle, and upper income is \$28,700, \$86,000, and \$207,400 respectively. Here, we compute the Median for a Pareto-style distribution[4] using a different procedure from the aforementioned procedure for computing a median. Approaching the distribution as a Pareto distribution, and using X_{min} = 28700, we can estimate the shape parameter α of the distribution using:

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

where x_m is the lowest value of our hypothetical distribution which is used to estimate the real X_{min}. Next, we approximate the Median using:

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

and the Entropy:

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

Applying these equations to our hypothetical distributions, we can estimate the median, the alpha parameter, the probability that n "successes" (i.e., rich) are above the true median, the entropy, and the percentage of wealth held by the upper 20% of the population using this equation:

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

The next table presents the results for our hypothetical distributions:

Table 1. The results for the 4 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%

According to the Federal Reserve, the upper 20% of the population in the States holds 88% of the wealth (Bricker et al., 2020). When generating hypothetical distributions using 8 balls, our layman has to choose between 8a and 8b. Choosing between distributions 8a and 8b, he chooses 8a, and we can see that the approximated percentage of wealth that we derive through this distribution better approximates the real percentage of wealth held by the upper 20% of the population. Choosing between distributions 9a and 9b, he chooses 9b and again we can see that it provides a better approximation of the wealth held by the upper 20% of the population. In both cases, the TM criterion helps us to choose the best distribution among several hypothetical distributions.

The rule of thumb

The only aim of the above procedure was to show that for a layman using a few linguistic quantifiers and a few particles (i.e., balls), there is a scientifically grounded approach for choosing among several hypothetical distributions in order to approximate the real distribution. This procedure which is grounded in the Principle of Max Entropy, leads us to the following rule of thumb:

Given N hypothetical distributions from which you have to choose, simply choose the one in which the “balls” are more homogenously distributed among the “boxes”.

Discussion

This short paper presents a simple albeit powerful rule of thumb for choosing the least biased (i.e., best) hypothetical distribution using a combination of computing with words and maximizing the entropy under the single constraint which is the rank order imposed by the quantifiers of the hypothetical distribution. For the layman, we may simply propose that among the hypothetical distributions, that correspond with the constraints imposed by the quantifiers, s (he) just has to choose the more uniformly distributed configuration, as this distribution maximizes the Shannon information entropy. In sum, and given Simon’s idea of bounded rationality (Simon, 2000), we may understand how individuals, modeled as bounded rational agents, may choose the best distribution for approximating a real-world distribution. Our paper doesn’t involve real data as we aim to propose a rule of thumb only and have no pretension to propose a new way of approximating real distributions or learn about the way real people approximate distributions. Our logic builds on the Principle of Max Entropy, which has been proven to be scientifically solid, and therefore for introducing our rule of thumb there is no need for more. Interested readers may, however, take this proposal further, considering our paper as a trigger only.

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.complexityexplorer.org/news/21-simon-dedeo-talks-about-his-maxent-tutorial

[2] The best or the least biased distribution is the one maximizing the entropy given a few constraints.

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

[4] https://en.wikipedia.org/wiki/Pareto_distribution

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: 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.
- Neuman, Y., Cohen, Y., & Tamir, B. (2021). Short-term prediction through ordinal patterns.*Royal Society Open Science*, *8*(1), 201011.
- 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.