Commonsense Explanations of Sparsity, Zipf Law, and Nash's Bargaining Solution

Olga Kosheleva[1], Vladik Kreinovich[1], and Kittawit Autchariyapanikul[2]

[1]University of Texas at El Paso
El Paso, Texas 79968, USA
Olgak@utep.edu, vladik@utep.edu
[2]Maejo University, Chiang Mai, Thailand
kittawit.a@mju.ac.th

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

1. Need for Commonsense Explanations

- In the 19 century,
  - from Adam Smith through Karl Marx,
  - economic phenomena were described by relatively simple intuitively clear formulas.
- However, it was well understood that these formulas provide, at best:
  - a qualitative understanding, and
  - qualitative predictions.
- Economics and finance are complex phenomena.
- So not surprisingly, simple models do not provide a good quantitative description of these phenomena.

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

2. Need for Commonsense Explanations (cont-d)

- Starting with the 20 century, new models appeared that allow us:
  - in many cases,
  - to make reasonable accurate predictions.
- The corresponding models more adequately describe economic and financial phenomena.
- However, they are often very mathematically complicated, far from common sense.
- And here lies a problem:
  - since models are complicated, requiring complex math beyond what most people know,
  - people in general – and politicians in particular – do not fully trust these models.

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

3. Need for Commonsense Explanations (cont-d)

- This lack of complete trust is understandable, since:
  - sometimes, predictions of previous models turn out to be wrong, and
  - decisions based on these predictions made crisis situations worse.
- It is therefore desirable to come up with commonsense explanations:
  - for the current models, or at least
  - for the major assumptions behind these models and for the main features of these models.

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

4. Need for Commonsense Explanations (cont-d)

- This need is similar to the well-publicized need for explainable AI.
- A female CEO who was denied a bank loan by an AI-computer program would like to make sure that:
  - this denial was based on inadequate state of her company
  - and not on the fact that, in the past, most loans were given to men.

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

5. What We Do in This Talk

- Of course, the goal of having explainable econometrics is complex.
- It requires joint efforts of many researchers.
- In this talk, we provide just three examples of the desired commonsense explanations.
- Specifically, we provide commonsense explanation:
  - for sparsity,
  - for Zipf law, and
  - for Nash's bargaining solution.

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

6. What Is Sparsity: A Brief Reminder

- Econometric analysis often starts with trying to use linear regression.
- We try to predict the value $y(t+1)$ of the desirable quantity y at the next moment of time $t+1$ (e.g., month or year).
- We predict $y(t+1)$ as a linear combination of:
  - the values $x_1(t), \ldots, x_n(t)$ of different quantities at the current moment of time,
  - and maybe of their past values $x_i(t-1)$, $x_i(t-2)$, $\ldots$
- Usually:
  - just to be on the safe side,
  - we add as many quantities x_i and as many past moments of time as possible.

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

7. What Is Sparsity: A Brief Reminder (cont-d)

- We understand that some of these values may be irrelevant.
- In the process of regression, it usually turns out that many of these variables are indeed irrelevant.
- We can see it by observing that the coefficients at these variables are close to 0.
- The problem with using the usual linear regression methods for linear regression is that:
  - since the data is noisy, and linear dependence is approximate,
  - we do not get perfect match, and thus, do not get exactly 0 values even where there is no dependence.
- These erroneously non-zero value make the resulting predictions much less accurate.

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

8. What Is Sparsity: A Brief Reminder (cont-d)

- Much more accurate predictions can be obtained if:
  - we explicitly require, from the very beginning,
  - that a large number of the coefficients be zero, i.e., that the dependence is *sparse*.
- This is the main idea behind, e.g., the LASSO method.
- Sparsity is not limited to economics and finance, it is a useful tool in data processing in general.

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

9. Why Do We Need a Commonsense Explanation for Sparsity

- At first glance, what we described makes sense.
- However, one can also raise a serious counter-argument:
  - in economics and in finance, everything is interrelated,
  - often, a small event across the world causes some stocks to go up or down,
  - so why would there be zero coefficients at all?
- The previous text does not provide any commonsense explanation.
- It just states the empirical fact that the resulting best-fit formulas are often sparse.
- It would thus be nice to have some commonsense explanation for this empirical phenomenon of sparsity.

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

10. Continuity: Main Commonsense Feature Underlying Our Explanation

- Our commonsense explanation is based on a reasonable feature of *continuity* – that:
  - small changes in one or several quantities should lead to
  - small changes in other quantities.
- Of course, there are extreme situations when there is no continuity.
- For example:
  - if the amount of money that a company needs to pay right away is slightly smaller than its current amount of liquid assets,
  - it can easily pay its creditors and suppliers.

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

11. Continuity (cont-d)

- On the other hand:
  - if the needed amount is slightly larger than what is available,
  - the company may be in trouble.
- But even this situation is more purely theoretical:
  - if the difference is very small,
  - he company can easily take a loan or negotiate a payment extension with its suppliers.
- And if they refuse because of the bad financial history of this company, this means that:
  - the trouble is in this bad history and
  - *not* in the minor inability to fully pay the debts right away.

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

12. Let Us Describe Continuity in Precise Terms

- Suppose that we have a linear relation between the input quantities $x_1, \ldots, x_n$ and the desired quantity y:

$$y = a_0 + a_1 \cdot x_1 + \ldots + a_n \cdot x_n.$$

- Continuity means, in particular, that:
  - if we make a small change in one or several values x_i,
  - i.e., replace the original value x_i with a new value $x_i + \Delta x_i$, where $|\Delta x_i| \le \varepsilon$ for some $\varepsilon > 0$,
  - then the resulting change Δy in y must also be small,
  - i.e., smaller than $\delta > 0$ for some $\delta \approx \varepsilon$.
- It may be tempting to take $\delta = \varepsilon$, but, as we will see, this is not possible.

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

13. Continuity in Precise Terms (cont-d)

- Similarly:
  - if we analyze the dependence of y on one of the variables x_i,
  - i.e., if we analyze how much we need to change x_i to reach the desired change in y,
  - then we can formulate a similar requirement.
- The requirement is that:
  - for each small change Δy (when $|\Delta y| \le \varepsilon$),
  - the corresponding change Δx_i should also be small: $|\Delta x_i| \le \delta$.
- Of course, this requirement only makes sense only:
  - when a change in x_i can actually change y, i.e.,
  - when the corresponding coefficient a_i is different from 0.

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

14. What We Can Conclude from Continuity

- If we change one of the variables x_i by $\Delta x_i = \varepsilon$, then the value of y changes by $\Delta y = a_i \cdot \Delta x_i = a_i \cdot \varepsilon$.
- So, the requirement that $|\Delta y| \le \delta$ takes the form $|a_i \cdot \varepsilon| \le \delta$, i.e., that $|a_i| \le \dfrac{\delta}{\varepsilon}$.
- Similarly, for the indices i for which $a_i \ne 0$, the change in y caused by a change Δx_i in x_i is equal to

$$\Delta y = a_i \cdot \Delta x_i.$$

- Thus, to get the change $\Delta y = \varepsilon$, we need to take

$$\Delta x_i = \frac{\Delta y}{a_i} = \frac{\varepsilon}{a_i}.$$

- The requirement that $|\Delta x_i| \le \delta$ thus takes the form $\left|\dfrac{\varepsilon}{a_i}\right| \le \delta$, i.e., equivalently, $\dfrac{\varepsilon}{\delta} \le |a_i|$.

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

15. Conclusions from Continuity (cont-d)

- This, by the way, shows why we cannot take $\delta = \varepsilon$.
- In this case, from the above formulas, we would conclude that $|a_i| \le 1$ and $1 \le |a_i|$, so $|a_i| = 1$.
- However, empirically, we can have different values a_i.
- Moreover, we must have $\varepsilon < \delta$.
- Otherwise, if $\varepsilon > \delta$, then we would get two inconsistent inequalities $|a_i| < 1$ and $1 < |a_i|$.
- So far, we considered the case when we change only one input.
- What if we change all of them, e.g., if we take $\Delta x_i = \text{sign}(a_i) \cdot \varepsilon$ for all i?

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

16. Conclusions from Continuity (cont-d)

- In this case, we have

$$\Delta y = \sum_{i=1} a_i \cdot \mathrm{sign}(a_i) \cdot \varepsilon = \sum_{i=1}^{n} |a_i| \cdot \varepsilon = \varepsilon \cdot \sum_{i=1}^{n} |a_i|.$$

- In this formula, we can safely ignore all the indices for which $a_i = 0$, so $\Delta y = \varepsilon \cdot \sum\limits_{i:a_i \neq 0} |a_i|$.
- Our continuity requirement is that $|\Delta y| \leq \delta$.
- In this example, $\Delta y > 0$, so $|\Delta y| = \Delta y$, and the desired inequality takes the form $\varepsilon \cdot \sum\limits_{i:a_i \neq 0} |a_i| \leq \delta$.
- We already know that $\dfrac{\varepsilon}{\delta} \leq |a_i|$, thus, $n_i \cdot \dfrac{\varepsilon}{\delta} \leq \sum\limits_{i:a_i \neq 0} |a_i|$.
- Here by n_i, we denotes the number of non-zero coefficients a_i.

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

17. Conclusions from Continuity (cont-d)

- Multiplying both side of above inequality by ε, we conclude that $n_i \cdot \dfrac{\varepsilon^2}{\delta} \le \varepsilon \cdot \sum\limits_{i:a_i \ne 0} |a_i|$.
- By combining two inequalities, we conclude that

$$n_i \cdot \frac{\varepsilon^2}{\delta} \le \delta, \text{ i.e., } n_i \le \frac{\delta^2}{\varepsilon^2}.$$

- Since $\delta \approx \varepsilon$, this means that the ratio $\dfrac{\delta^2}{\varepsilon^2}$ is reasonably small.
- So, the number n_i of non-zero coefficients is small.
- This is exactly sparsity!

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

18. Zipf's Law: A Brief Reminder

- This law is named after a scientist who discovered that in linguistics:
  - if we order all the words from a natural language from the most frequent to the least frequent ones,
  - then the frequency p_n of the n-th word in this order is approximately equal to c/n.
- This law turned out to be ubiquitous, it is applicable to many phenomena, including economic ones.
- For example, it describes the distribution of companies by size.
- How can we explain this law in commonsense terms?

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

19. Main Idea Behind Our Explanation

- We sorted objects (words, companies, etc.) in the reverse order according to their frequency.
- This means that:
  - if $m < n$, then $p_n < p_m$, and
  - vice versa, if $p_n < p_m$, then $m < n$.
- These formal statements are not very informative:
  - they are satisfied for Zipf's law $p_n = c/n$, but
  - they could be satisfied for all other possible monotonic sequences, e.g., $p_n = c/n^2$ or $p_n = \exp(-c \cdot n)$.
- However, a natural commonsense understanding of these two conditions goes beyond their formal definition.

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

20. Main Idea (cont-d)

- It is natural, e.g., to also conclude that:
  - if m is somewhat smaller than n, then p_n is somewhat smaller than p_m, and that
  - if m is much smaller than n, then p_n is much smaller than p_m.

  This commonsense understanding can be described as follows.

- Intuitively, a statement that A implies B means not only that if A is true then B is true.
- It also means that, more generally:
  - the degree $d(B)$ to which we believe in B is at least as large as
  - the degree $d(A)$ to which we believe in A.

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

21. Main Idea (cont-d)

- In principle, we could formalize the notion of a degree – e.g., by using fuzzy logic.
- However we want to stick to commonsense ideas.
- Let us not go into formalization unless it becomes necessary.

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

22. Resulting Explanation

- From the viewpoint of the above understanding, the two implication statements mean the following:
  - the first statement means that
$$d(m < n) \le d(p_n < p_m);$$
  - similarly, the second statement means that
$$d(p_n < p_m) \le d(m < n).$$
- The above two inequalities imply that these degree are equal: $d(p_n < p_m) = d(m < n)$.
- How do we describe the degree $d(a < b)$ that a is smaller than b?
- In principle, this can be done differently.
- However, in economic applications, this is usually understood in relative – percent – terms.

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

23. Resulting Explanation (cont-d)

- If we say that someone started earning 10K more per year, we are not explaining anything:
  - it may be going to 0 to 10K, or
  - it may be a miserable increase from 300K to 310K.
- A much better commonsense understanding is provided by the ratio of two numbers.
- No matter what we started with:
  - a 3% increase is not much,
  - a 20% increase is always significant, and
  - a 100% increase is always drastic.
- So, in economic applications, the degree $d(a < b)$ to which a is smaller than b is determined by the ratio:

$$d(a < b) = f(b/a) \text{ for some increasing function } f(x).$$

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

24. Resulting Explanation (cont-d)

- From this viewpoint, $d(p_n < p_m) = d(m < n)$ implies

$$f\left(\frac{p_m}{p_n}\right) = f\left(\frac{n}{m}\right).$$

- Since the function $f(x)$ is increasing, this equality is equivalent to $\dfrac{p_m}{p_n} = \dfrac{n}{m}$.
- If we move all the terms containing m to one side and all the terms containing n to the other side, we get

$$m \cdot p_m = n \cdot p_n \text{ for all } m \text{ and } n.$$

- This product is the same for all n, i.e., is a constant:

$$n \cdot p_n = c.$$

- Thus, $p_n = c/n$, exactly the Zipf's law!

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

25. Nash's Bargaining Solution: A Brief Reminder

- In many real-life situations, the outcome depends on the actions of several actors.
- A classical example is several countries producing oil.
- Depending on their actions, they may get different gains $(g_1, \ldots, g_n)$.
- In some possible outcomes, one actor gets more, in other outcomes, another actor gets more.
- For example, if one country cuts off its production and others don't, other countries will benefit.
- If the actors do not coordinate their action – e.g., if we all produce too much oil:
  - the price of oil will dive, and
  - they will all lose.

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

26. Nash's Bargaining Solution (cont-d)

- So, they need to come up with a joint compromise solution.
- A Nobelist John Nash:
  - used some – rather complex – math – to show that
  - the best strategy is to maximize the product $\prod_{i=1}^{n} g_i$ of the participants' gains.

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

27. It Is Desirable to Have A Commonsense Explanation

- From the mathematical viewpoint, it is a good solution.
- However, from the commonsense viewpoint, multiplying gains does not make any sense.
- How can we explain this idea in commonsense terms?

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

28. Main Idea and the Resulting Explanation

- How can we gauge the quality of a solution?
- For athletes, a natural measure of the athlete's quality is the number of athletes whose performance is lower.
- For standardized tests, the person's results is gauged by the percentage of test takers with worse performance.
- This way it is clear that:
  - if this percentage is 3%, the graduate program applicant is not very good, while
  - if it is 99.9%, we should accept this person right away.

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

29. Resulting Explanation (cont-d)

- Similarly:
  - we can gauge the quality of an alternative $g = (g_1, \ldots, g_n)$
  - by counting how many alternatives are worse.
- In other words, by counting how many there are tuples $g' = (g'_1, \ldots, g'_n) \neq g$:
  - which are clearly worse g, i.e.,
  - for which $g'_i \leq g_i$ for all i.
- From the purely mathematical viewpoint, there are infinitely many such tuples.
- In practice, however:
  - there is some "quantum" here,
  - the smallest amount $\varepsilon > 0$ below which the difference is not noticeable.

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

30. Resulting Explanation (cont-d)

- It may be 1 cent if we talk about people.
- It may be 1 million dollars if we talk about oil producing companies.
- From this viewpoint, there are g_1/ε values 0, ε, 2ε, ... which are smaller than or equal to g_1.
- For each of these values, there are g_2/ε values which are smaller than or equal to g_2.
- So, we have $(g_1/\varepsilon)\cdot(g_2/\varepsilon)$ possible combinations.
- Overall, the number of worse-than-g alternatives is $(g_1/\varepsilon)\cdot \ldots \cdot (g_n/\varepsilon) = \dfrac{\prod\limits_{i=1}^{n} g_i}{\varepsilon^n}$.
- Maximizing this number is equivalent to maximizing $\prod\limits_{i=1}^{n} g_i$; this is exactly Nash's bargaining solution!

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .

31. Acknowledgments

This work was supported in part by the National Science Foundation grants:

- 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science),
- HRD-1242122 (Cyber-ShARE Center of Excellence).

Need for . . .
What Is Sparsity: A . . .
Why Do We Need a . . .
Continuity: Main . . .
Zipf's Law: A Brief . . .
Main Idea Behind Our . . .
Resulting Explanation
Nash's Bargaining . . .
Resulting Explanation . . .