Diversity Is Beneficial for a Research Group: One More Quantitative Argument

Komsan Suriya[1], Tatcha Sudtasan[1], Tonghui Wang[2],
Octavio Lerma[3], and Vladik Kreinovich[3]
[1]Faculty of Economics, Chiang Mai University, Thailand,
suriyakomsan@gmail.com, miracle.zaza@gmail.com
[2]Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 80003, USA, twang@nmsu.edu
[3]Computational Science Program
University of Texas at El Paso
El Paso, TX 79968, USA
lolerma@episd.org, vladik@utep.edu

Abstract

In this paper, we propose a natural model describing competition between two research groups of the same average research strength. The analysis of this model enables us to conclude that a more diverse group has an advantage: namely, the more diverse the group, the higher the average quality of its publications.

1 Introduction

Diversity is beneficial. Experiments and simulation have shown that, in general, more diverse groups have an advantage over less diverse ones; see, e.g., [2, 3, 5, 6].

What we do in this paper. In this paper, we provide one more quantitative argument in favor of diversity. Namely, we show that if we have two competing research groups with the same average strength, then the more diverse research group has a clear advantage.

Comment. Some of the results from this paper first appeared in [8].

2 Two Competing Research Groups: Description of a Model

Natural assumption: strength is normally distributed. Normal distributions are ubiquitous, they appear in many real-life situations; in particular, they describe the distribution of many characteristics of a human being such as height, weight, blood pressure, or IQ. The ubiquity of normal distribution can be explained by the fact that in many cases, the value of the quantity is caused by many independent factors, and the known Central Limit Theorems states, crudely speaking, that the distribution of the sum of large number small independent factors is close to normal; see, e.g., [7].

It is therefore reasonable to assume that within each of the two competing research groups, strength is normally distributed. In general, a normal distribution is uniquely determined by its mean μ and its standard deviation σ. In terms of strength, the mean is the average strength, while the standard deviation describe diversity: the larger the standard deviation, the more diverse the group.

We assume that both groups have the same average strength μ, but that the first group is more diverse: $\sigma_1 > \sigma_2$.

How the groups compete: a description. We assume that each group coordinate the research efforts of its members, so there is no unnecessary competition within each group; the only competition is between the two groups. Once a member of one of the research groups selects a problem – a problem that people in the field consider to be important – it is highly probably that the same problem will be picked up by some member of another research group.

The groups (being competitors) do not coordinate their research efforts with each other. As a result, the corresponding member of another research group is randomly selected from that group. If two researchers of different research strength $s_1 > s_2$ work on the same problem, it is reasonable to expect that the stronger researcher will get the results first – and this will result in a paper of quality corresponding to this higher strength s_1.

Let us analyze the resulting model.

3 Analysis of the Model

Let us analyze. Under the above assumptions, let us see which of the two groups has an advantage. Intuitively, the answer is not clear:

- on the one hand, the more diverse research group has a larger number of stronger researchers, which gives this group an *advantage* over the less diverse group;
- on the other hand, the more diverse research group also has a larger number of weaker researchers, which gives this group a *disadvantage* over the less diverse group.

At first glance, diversity brings no advantage. In the above competition, which of the two groups will be more successful? Let us first consider the simplest measure of success: the resulting number of publications.

The first group gets a publication is a value s_1 randomly selected from the first group exceeds a value s_2 randomly selected from the second group: $s_1 > s_2$. Thus, the number of publications produced by the first group is proportional to the probability that for randomly selected values s_1 and s_2, we have $s_1 - s_2 > 0$, i.e., equivalently, that $s_1 - s_2 > 0$. The two independent random variables x_1 and x_2 are normally distributed with the same mean μ. It is known that the difference of two independent normally distributed random variables is also normally distributed. The mean of the difference $s_1 - s_2$ is equal to the difference of the means, i.e., to $\mu - \mu = 0$. Thus, $s_1 - s_2$ is a normally distributed random variable with 0 mean. For such random variable, the probability of it being positive is exactly $\dfrac{1}{2}$.

Thus, when the two research groups have the same average strength, in half of the cases, the first group will succeed, in half of the cases, the second group will succeed. So, both groups will generate, on average, the same number of publications.

Towards a deeper analysis. In terms of *number* of publications, diversity does not bring any advantage. However, different publications have different *quality*. What if, instead of simply counting the number of publications, we would instead estimate the average quality of a publication?

According to our model, the first group succeeds if $s_1 > s_2$ and produces a paper of quality s_1. Thus, the average quality q_1 of papers produced by the first research group is equal to the conditional expectation $q_1 = E[s_1 \mid s_1 > s_2]$. Similarly, the the average quality q_2 of papers produced by the second research group is equal to the conditional expectation $q_2 = E[s_2 \mid s_1 > s_2]$.

Let us estimate these two quantities.

Estimating the desired quantities. The first research group produces a paper of quality s_1:

- if there is a person of strength s_1 in this group and
- if this person was stronger than the competitor, i.e., a person with a (randomly selected) strength s_2 from the second research group.

The probability of the first research group having a member of strength s_1 is determined by the normal distribution, i.e., has the form

$$f_1(s_1) = \frac{1}{\sigma_1} \cdot \phi\left(\frac{s_1 - \mu}{\sigma_1}\right),$$

where

$$\phi(x) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2}{2}\right)$$

is the probability density of the standard normal distribution (i.e., a normal distribution with mean 0 and standard deviation 1).

The probability that s_1 will win over the competitor is equal to

$$\text{Prob}(s_2 < s_1).$$

By definition of the cumulative distribution function (cdf) $F_2(x)$ of the random variable s_2, this probability is equal to $\text{Prob}(s_2 < s_1) = F_2(s_1)$. Since the variable s_2 is normally distributed, this probability has the form $F_2(s_1) = \Phi\left(\frac{s_1 - \mu}{\sigma_2}\right)$, where $\Phi(x)$ is the cdf of the standard normal distribution.

Since s_1 and s_2 are independent, the probability distribution function $f(s_1)$ for the publication quality s_1 is proportional to the product of the two probabilities, i.e., has the form

$$f(s_1) = \text{const} \cdot \phi\left(\frac{s_1 - \mu}{\sigma_1}\right) \cdot \Phi\left(\frac{s_1 - \mu}{\sigma_2}\right).$$

Such a distribution is known: it is a *skew-normal* distribution; see, e.g., [1, 4, 9] and references therein. To be more precise, the usual formula for the skew-normal distribution has the form

$$f(s_1) = \text{const} \cdot \phi\left(\frac{s_1 - \mu}{\sigma_1}\right) \cdot \Phi\left(\alpha \cdot \left(\frac{s_1 - \mu}{\sigma_1}\right)\right),$$

which coincides with the above form for $\alpha = \frac{\sigma_1}{\sigma_2}$.

It is known that the mean value of the skew-normal random variable is equal to $q_1 = \mu + \sqrt{\frac{2}{\pi}} \cdot \sigma_1 \cdot \frac{\alpha}{\sqrt{1+\alpha^2}}$. Substituting $\alpha = \frac{\sigma_1}{\sigma_2}$ into this formula and multiplying both the numerator and the denominator of the corresponding fraction by σ_2, we conclude that

$$q_1 = \mu + \frac{\sigma_1^2}{\sqrt{\sigma_1^2 + \sigma_2^2}}.$$

Similarly, the average quality of papers published by the second research group is equal to

$$q_2 = \mu + \frac{\sigma_2^2}{\sqrt{\sigma_1^2 + \sigma_2^2}}.$$

Conclusion. From the above formulas, we can see that the larger the standard deviation σ_i, the larger the average quality q_i of the corresponding publications. Thus, while a diverse group produces, on average, the same number of publications, the average quality of these publications is higher – and the more diverse the group, the higher the quality.

Acknowledgments. This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721, and by a travel grant from Chiang Mai University, Thailand.

The authors are thankful to all the participants of the 7th International Conference of Thailand Econometric Society TES'2014 (Chiang Mai, Thailand, January 8–10, 2014) for valuable discussions.

References

[1] A. Azzalini and A. Capitanio, *The Skew-Normal and Related Families*, Cambridge University Press, Cambridge, Massachusetts, 2013.

[2] L. Hong and S. E. Page, "Problem solving by heterogeneous agents", *Journal of Economic Theory*, 2001, Vol. 97, No. 1, pp. 123–163.

[3] L. Hong and S. E. Page, "Groups of diverse problem solvers can outperform groups of high-ability problem solvers", *Proceedings of the National Academy of Sciences*, 2004, Vol. 101, No. 46. pp. 16385–16389.

[4] B. Li, D. Shi, and T. Wang, "Some applications of one-sided skew distributions", *International Journal Intelligent Technologies and Applied Statistics*, 2009. Vol. 2, No. 1.

[5] J. H. Miller and S. E. Page, *Complex Adaptive Social Systems: The Interest in Between*, Princeton University Press, Princeton, New Jersey, 2006.

[6] S. E. Page, *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*, Princeton University Press, Princeton, New Jersey, 2007.

[7] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall/CRC, Boca Raton, Florida, 2007.

[8] K. Suriya and T. Sudtasan, "Fighter ants algorithm with application to efficient team building", *Abstracts of the 7th International Conference of Thailand Econometric Society TES'2014*, Chiang Mai, Thailand, January 8–10, 2014.

[9] T. Wang, B. Li, and A. K. Gupta, "Distribution of quadratic forms under skew normal settings", *Journal of Multivariate Analysis*, 2009, Vol. 100, pp. 533–545.