

Wisdom of the Crowd

An Extension Beyond Independence

Pranav Tikkawar

Wisdom in an individual is a trait developed over years of experience, requiring perseverance through many trials and tribulations year over year. However, the wisdom of a crowd needs about 30-40 people to get 60-80 years of individual experience. Though these two notions of wisdom are somewhat different, it is undeniable that for an estimation problem, more people lead to a more accurate result. This result is derived from the Central Limit Theorem and Law of Large Numbers¹ as it leverages the mathematical concept of averaging to reduce the estimate's variance. The sample mean is an unbiased estimator of the population mean, and as the number of samples increases, the variance of the sample mean decreases on the order of $\frac{1}{n}$. This means that as we get more and more samples, our estimate of the population mean becomes more and more accurate. This is a compelling result, as we can get a very accurate estimate of the population mean with a relatively small number of samples.

Despite the seeming universal utility of this property, there is a key assumption that it requires: independence of each of the observed data points. Without that, the sample mean can drift far away from our desired results. Even social influence, like prior knowledge and mild exposure to other guesses, changes the variance of guesses, which would be integral to estimating the population mean². Due to these limitations, experimental design is severely limited in enforcing the most ideal data collection conditions. However, there are possible scenarios where this assumption can be relaxed while still maintaining reasonable accuracy.

Such a scenario is similar to the canonical example of the crowd's wisdom: guessing a cow's weight. If we had a group of people guess the weight of a cow, each writing their guesses independently, we would expect the average of their guesses to be near the actual weight of the cow. But what if we had a group of people guess the weight of a cow, but each person heard the guesses of the people before them? This scenario is not completely contrived, as a scenario similar to this could occur in a classroom setting where students are presented with the same problem and are instructed to say their guesses out loud. In this scenario, the guesses are not independent, but they are not entirely dependent either. The previous guesses influence each other, but there is still some individual influence.

Consider the following model for each person's guesses. Each person i has an independent guess $Y_i \sim N(\mu, \sigma^2)$, and a random influence $U_i \sim \text{Uniform}(0, 1)$ from the prior guess to influence their complete guess X_i . The formula for the complete guess is given by: $X_i = (U_i)Y_i + (1 - U_i)X_{i-1}$. The first guess is not influenced by any prior guesses, so we set $X_1 = Y_1$ (the first guess is just a normal random variable). Let us denote the first mean and standard deviation as $\mu_1 = \mu$ and $\sigma_1 = \sigma$. The goal is to estimate the population mean μ from the

¹Da Z, Huang X. Harnessing the Wisdom of Crowds. *Management Science*. 2020;66(5):1847-1867. doi:10.1287/mnsc.2019.3294

²Lorenz J, Rauhut H, Schweitzer F, Helbing D. How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences - PNAS*. 2011;108(22):9020-9025. doi:10.1073/pnas.1008636108

guesses X_i . One can see that the guesses X_i are not independent, as each guess is influenced by the prior guess. However, they are not completely dependent either, as each guess is also influenced by an independent random variable. The question is: can we still estimate the population mean μ from the guesses X_i ? Despite the lack of independence, we can still get a reasonable estimate of the population mean μ from the guesses X_i .

We can see that the expected value of each guess is $\mu + (\mu_1 - \mu) \left(\frac{1}{2}\right)^{i-1}$. This means that as i increases, each guess's expected value converges to the population mean μ . This is a good sign, as it means that the guesses are not drifting away from the population mean. Additionally, the sample mean for n individuals is given by $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i = \mu + (\mu_1 - \mu) \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{2}\right)^{i-1}$. This means that as n increases, the sample mean also converges to the population mean μ . This is also beneficial, as it means that the average of the guesses is not drifting away from the population mean. The variance of each guess as we get further and further down the line is given by $\frac{3}{2}\sigma^2 + \frac{12}{5}\mu\mu_1 + \frac{1}{10}\mu^2$. This means that as i increases, the variance of each guess converges to a constant value. This means that the guesses are not becoming more and more spread out. Additionally, the variance of the sample mean for n individuals decreases on the order of $\frac{1}{n}$ at worst (proof omitted). All these properties are similar to the properties of independent random variables³.

Overall, despite the lack of independence, we can still get a reasonable estimate of the population mean μ from the guesses X_i with a similarly reasonable convergence rate. This means that we can relax the independence assumption in some scenarios and still get a reasonable estimate of the population mean. This is useful, because it means we can design experiments in a more flexible manner and still get reasonable results. This is a natural extension of the wisdom of the crowd, as it means we can still get reasonable results even when the independence assumption is not met, and many more real-world scenarios can be modeled in this manner.

This model is not without its limitations. The model assumes that each person is influenced by the prior guess in a linear manner, which may not be the case in reality. Additionally, the original wisdom of the crowd model is prone to failure as biases can be introduced in the guesses, skewing the results⁴. That being said, this model is a good first step in relaxing the independence assumption in the wisdom of the crowd. Further research can be done to explore other models that relax the independence assumption and see if they can still produce reasonable results⁵.

³I will not be providing the proof of all these properties in this file, but if you are interested (im assuming not) I can type set it and email it to you :D

⁴Eger S. (Failure of the) Wisdom of the crowds in an endogenous opinion dynamics model with multiply biased agents. Published online 2013. doi:10.48550/arxiv.1309.3660

⁵wink wink

References

- 1 Da Z, Huang X. Harnessing the Wisdom of Crowds. *Management Science*. 2020;66(5):1847-1867. doi:10.1287/mnsc.2019.3294
- 2 Lorenz J, Rauhut H, Schweitzer F, Helbing D. How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences - PNAS*. 2011;108(22):9020-9025. doi:10.1073/pnas.1008636108
- 4 Eger S. (Failure of the) Wisdom of the crowds in an endogenous opinion dynamics model with multiply biased agents. Published online 2013. doi:10.48550/arxiv.1309.3660