Rationality Color Wheel Advanced Notes

In this note, we explain how the Rationality Color Wheel was created, discuss the math behind it, and provide guidance for advanced use.

Bayes' Formula

The Rationality Color Wheel is based on Bayes' formula. Given a hypothesis H and some new evidence E, we can compute the probability that H is true using the facts that were available before E. We express this probability as $P(H)$. We then compute the probability that we would have E given that H is true. We write this probability as $P(E|H)$. Likewise, we compute the probability that we have E given that H is **not** true, which we write as $P(E|\neg H)$. Bayes formula tells us how to combine these probabilities to compute the probability that H is true *given* E, which we write as $P(H|E)$. Bayes' formula can be written as:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|\neg H)(1 - P(H))} \tag{1}$$

Construction

The Color Wheel makes it easy to compute this formula. It divides the range of possible probabilities into five intervals: "very unlikely," "unlikely," "fair," "likely," and "very likely." And assigns a color to each range.

Color	Minimum Value	Maximum Value	Interpretation	"Standard" Value
Red	0%	5%	"Very unlikely"	5%
Orange	> 5%	20%	"Unlikely"	20%
Yellow	> 20%	< 80%	"Fair"	50%
Green	80%	< 95%	"Likely"	80%
Blue	95%	100%	"Very likely"	95%

Bayes' formula takes three input values: $P(H)$, $P(E|H)$, and $P(E|\neg H)$, and returns a single output: the "posterior" value $P(H|E)$. The Color Wheel plots out every possible combination of these values in the form of a radial disk [1]. Each slice of the disk presents one such combination and each slice has four segments. The Outermost segment represents $P(H)$. The second segment represents $P(E|H)$. The third segment represents $P(E|\neg H)$. And, the inner segment represents $P(H|E)$.

Now each color represents a range of probabilities. As a result, each combination of inputs corresponds to a range of posterior probabilities. For example, the inputs "Blue, Green, Green", could represent 100%, 80%, 90%. It could also represent 95%, 94%, and 80%. Each of these combinations have different output values - different "posteriors." To account for this, the innermost segment has three slices. The outermost represents the greatest posterior probability that could result from inputs having the given colors. The innermost represents the smallest

posterior probability. The middle slice represents the posterior probability that would result if each color corresponded exactly to the "standard" values.

Nuances

The Color Wheel is intended to be used when working with probabilities that are not extremely small or extremely large [2]. It's goal is to provide a useful tool for weighing evidence concerning issues that arise in everyday life. The Color Wheel helps people get close to right answer when dealing with probabilities falling roughly between 2% and 97%.

When the input probabilities are close to the standard values, the middle slice of the innermost segment should give a good estimate of the true posterior probability. When dealing with extremely small and extremely large probabilities however, the upper and lower bounds on the Color Wheel become increasingly relevant. For example, the inputs "Red, Red, Red" can imply that the posterior probability is either very high or very low. This is indicated by the upper and lower bounds for this slice.

As a general rule, **increasing the probability of the outermost and second rings moves the posterior probability closer to the upper bound. Increasing the probability of the third ring moves the result closer to the lower bound.**

So, for example, let $P(H) = 79\%$ (high yellow), and the remaining values equal the standard value for Green: $P(E|H) = 80\%$ and $P(E|\neg H) = 80\%$. Then the true result will be closer to the upper bound (Green) indicated in the Color Wheel. Conversely, let $P(H) = 21\%$ (low yellow) and the actual result will be closer to the lower bound (Orange) indicated in the Color Wheel. In both cases, because only one input moved, we know that the true value is between the "middle" value and the mentioned bound.

Conclusion

Bayesian inference is a very powerful method. We hope that the Color Wheel has both empowered you and sparked your curiosity. Have fun and tell others about the Rationality Color Wheel!

Footnotes

---

1. Originally, we intended to create a paper handtool in which users rotated the wheel behind a screen that had a single slit in it that revealed only one slice of the table. This was the original motivation for the radial design. ↩

2. By "extremely small," we mean values much less than 1%. By "extremely large," we mean probabilities much greater than 99%. ↩