

Combining Product Ratings

Q6.1 Why is an approximation symbol used in the wisdom of crowds equation in the book?

This result holds as an expectation, *i.e.*, the *expected* error in the average decreases from the *expected* error in each guess. In reality, each individual guess is going to have a different error associated with it. In fact, they *need* to be different for this to work. Think about it: what would happen if all of the guesses were the same? This means they would all be off by the same amount, and would have the same error. If we average over the same number, we will just end up with that number (*e.g.*, $(5 + 5 + 5)/3 = 5$), so averaging really isn't doing anything for us here. The reason is that the guesses in this case are all completely *dependent* on each other, because if we know one of them, we know all of them. Complete *independence* is the exact opposite, where any one guess' error won't give us any information about the others. The independence assumption in the wisdom of crowds is indeed an important one.

Q6.2 Can we figure out how Amazon ranks lists of products?

The exact formula is not known outside of Amazon. As an example, let us consider the list of the top 20 thermoses in January 2013, shown in Illustration 19. It was obtained from Amazon by applying the following sequence of search filters:

- Search query: Thermos
- Then select Home & Kitchen → Kitchen & Dining to enable sorting.

There are two rank-ordered lists:

- The first is the ranking by Amazon, which orders the list in Illustration 19: 1, 2, 3, ...
- The second is the ranking by “average rating,” or the average number of stars. Rather than showing this number explicitly, Amazon rounds it and gives an “average customer review.” This leads to a different order: 17, 15, 1, 2, 4, 3, 18, 7, 6, 13, 5, 9, 10, 19, 8, 12, 11, 20, 16, 14.

Thermos	No. reviews	5 star	4 star	3 star	2 star	1 star	“Ave. review”	Ave. rating
1	210	179	20	7	3	1	4.8	4.776
2	87	74	8	3	1	1	4.8	4.759
3	1064	850	126	42	20	26	4.6	4.648
4	442	349	64	10	6	13	4.7	4.652
5	273	206	37	13	8	9	4.5	4.549
6	411	312	60	16	8	15	4.6	4.572
7	158	122	21	7	6	2	4.6	4.614
8	183	135	27	10	4	7	4.5	4.525
9	110	81	18	4	4	3	4.5	4.545
10	218	166	28	8	6	10	4.5	4.532
11	122	85	24	6	3	4	4.5	4.500
12	181	132	28	9	4	8	4.5	4.503
13	607	477	64	20	18	28	4.6	4.555
14	331	229	54	19	12	17	4.4	4.408
15	24	20	3	1	0	0	4.8	4.791
16	102	75	10	9	4	4	4.5	4.451
17	19	18	1	0	0	0	4.9	4.947
18	69	57	6	1	2	3	4.6	4.623
19	47	29	15	2	1	0	4.5	4.532
20	46	31	10	3	0	2	4.5	4.478

Illustration 19: A list of the top twenty thermoses on Amazon. Most of the items appear strangely out of order compared with the average reviews.

What are the differences between the ranking by Amazon and the one by average customer review?

- Amazon places items 17, 15, 18, and 19 much lower than their average ratings would indicate.
- Amazon places items 5, 8, and 11, and 14 much higher than their average rating would indicate.

From our discussion in this section of the various factors that can affect the quality of a ranking, can we think of what may have caused these differences between these two lists?

1. *Bayesian adjustment*: The population sizes of the ratings matter. The raw rating scores should be weighted with the population size

in some way.

2. *Recency of the reviews*: Perhaps some of the reviewers rated their thermoses as soon as they purchased them, and gave them high ratings because the products were durable initially. Sometimes, items deteriorate quicker than anticipated, which is not apparent at first.
3. *Quality of the reviewers or reviews*
 - (a) *Reputation score of the reviewer*: Reviewers with higher reputations should be given more “say” in the average customer review of a product.
 - (b) *Quality of review*: The quality of a review can be measured in terms of its length or associated keywords in the text.
 - (c) *Review of review*: Higher review scores indicate that customers found the review “helpful” and accurate.
 - (d) *Timing of reviews*: Review spam from those attempting to promote competing products can be partially by looking for unusual patterns in review timing.

Let’s first apply Bayesian adjustment using the equation on page ?? in the book. To do this, we need to know “Overall Num.” and “Overall Avg.” The latter is just the average rating of all the products (here assumed to be the top 20); from Illustration 19, you can apply Bayesian adjustment to obtain an average of 4.59.

But what should we choose for “Overall Num.”? We could very well use the total number of reviews, as in our examples before. But as discussed, some applications may limit this number in practice, such as to the *minimum* number of reviews necessary for a product to be listed (as in the Beer Advocate’s website), or possibly to some number in between the minimum and the total.

Let’s compare the Bayesian ranking result using the following selections of “Overall Num.”: 19, which is the lowest number of reviews for any product (number 17); 1064, which is the highest number of reviews for any product (number 3); 235, which is the average number of reviews per product; and 4704, the total number of reviews entered for the top 20 products. Each choice leads to different Bayesian rankings, ordered from highest product to lowest:

Most Helpful Customer Reviews

60 of 61 people found the following review helpful

★★★★★ **Thermos is the way to go** June 12, 2012

By Leann

Color Name: Teal | Amazon Verified Purchase

Well, we are farmers and have turned to thermos for years for hot beverages to keep time to upgrade our line of thermoses. Not knowing what to get, we bought two of each four Thermos King jars for a total of 8. All performed the same. The Sipps vacuum sometimes hard to open. We have keep everything in these jars from eggs to casses the 7 hours as promised. We found if we fill the thermos with hot water first to warm longer than the promised 7 hours. All in all, we love both lines and would highly recommend meal while away from the house. (sorry for any typos - on smart phone writing this)

Illustration 20: For the 15th thermos, 60 out of 61 people found the “most helpful” review helpful, which might help increase its rank.

- **19:** 17, 1, 2, 15, 4, 3, 18, 7, 6, 13, 5, 9, 19, 10, 8, 11, 12, 20, 16, 14.
- **1064:** 1, 3, 4, 2, 17, 15, 7, 18, 19, 9, 6, 20, 5, 11, 8, 10, 13, 16, 12, 14.
- **235:** 1, 3, 2, 4, 17, 15, 7, 18, 19, 6, 9, 20, 5, 13, 10, 8, 11, 12, 16, 14.
- **4704:** 3, 1, 4, 2, 17, 15, 7, 18, 19, 9, 20, 6, 5, 11, 8, 10, 16, 12, 13, 14.

Clearly, having the number too large or too small results in rankings that are far out of order. Both “1064” and “235” give results closer to Amazon’s ranking, at least in terms of grouping clusters of products together in the rank-ordered list. Still, there are some exceptions in each case. We will focus our attention now on products 8, 15, 17, and 20.

We notice that products 15, 17, and 20 have the lowest number of ratings, specifically 24, 19, and 46, respectively. This helps explain why they are further down than expected. But, even so, why would they be placed *apart* from one another?

Let’s take into account the *review of reviews*: in the case of product 15, for instance, the “most helpful” review had 60 people find it helpful, as shown in Illustration 20. On the other hand, products 17 and 20 had only seven and six, respectively, which could explain why 15 is ranked higher.

★★★★★ **Excellent Travel Mug - Keeps Coffee Hot for Most of the Day**, November 28, 2012
 By **Michael Gallagher** (Houston, TX) - [See all my reviews](#)
VINE™ VOICE HALL OF FAME REVIEWER TOP 10 REVIEWER REAL NAME
This review is from: Thermos Stainless King 16-Ounce Leak-Proof Travel Mug with Handle, Cranberry (Kitchen)
Customer review from the Amazon Vine™ Program ([What's this?](#))
 This travel mug does what it is supposed to do: keep your coffee or other beverage hot / cold for time, and using a vacuum type system vs. whatever else may be out there does a much better job warmer, longer.
 The label on the jar says it keeps things hot for five hours: my experience was it kept coffee hot as long as the slider is in the "close" position to activate the vacuum sealer of this unit. This is the other Thermos travel mug I have, except this one has the handle which I prefer over the one without.
 This container is about ten dollars higher than the one without the handle, which seems a little pricey. It's also higher than other non-Thermos branded travel mugs, but to me that is worth it as you get a product that keeps your beverage hot for an extended period of time. Overall, I consider this a quality travel mug: I wouldn't mind having several more of these!

Illustration 21: One of the reviews for the 8th thermos was from an Amazon “hall of fame” reviewer (7th overall), and he gave it five stars. This may have helped promote the ranking.

We can also look at the *recency of the reviews*: the “most helpful” review for product 17 was made in May of 2011. Product 17’s “saving grace” is that its corresponding review was more recent, made in December of 2012, which would push it closer to product 15 in the ranking. On the other hand, Amazon may have deemed that product 20’s reviews to not have enough overall helpfulness despite their recency.

Finally, product 15 had an extremely high quality of review in its descriptive listing of the pros and cons in each case. This may have also pushed its ranking higher.

Notice that despite being ranked so low, product 17 has the *highest* average rating. In addition to what we just discussed, there are two more factors that can help explain this. First, only half of the reviews are from the current year. Second, average reviewer reputation is fairly low, with all but 2 having less than 10 helpful votes overall.

Now, why would Amazon decide to rank an item such as 8 so high, given that the Bayesian ranking places it around the 16th position? Well, it turns out that item 8’s “most helpful” review had 79 out of 79 people find it helpful, and over 10 of Amazon’s top 5000 reviewers reviewed the product (reviewer rankings extend beyond 3 million on Amazon). Further, the reviewers all have high ratings, and one of them is an Amazon “hall of fame” reviewer (ranked 7th overall), as shown in Illustration 21. Quality of reviews overrides the quantity.

To summarize, the following set of guidelines is inferred from this

(small) sample on how Amazon comes up with its ranking.

1. An initial Bayesian ranking is made, with “Overall Num.” chosen to be somewhere around the maximum number of reviews entered for any product, or the average number of reviews across the products.
2. Products that have a small number of reviews, or low recency of their most helpful review, are ranked separately amongst themselves, and are moved to lower locations (*e.g.*, 15, 17, and 20).
3. Products that have very-high-quality, positive reviews from top reviewers are bumped up in the ranking (*e.g.*, 8).
4. Products for which there are few reputable reviewers (*e.g.*, 17) are severely demoted in the ranking.