The Powerball Problem

Kunal Marwaha

March 2020

1 Why?

This problem came to me in a heated discussion with my roommate about finding people that are similar to you. It's not exactly the same as the Powerball lottery, but in a similar spirit.

2 Problem

Suppose there are $n = 100$ preferences, and you have $k = 5$ of them. Everyone also has k of them, with no preference more common than another. What's the chance your preferences overlap with someone else?

2.1 Chance of full overlap

There are $\binom{n}{k}$ ways to choose k among n. There is only one way to have the same k preferences as you. So, the chance is

$$\frac{1}{\binom{n}{k}} = \frac{(n-k)!k!}{n!}$$

In our example, this happens 1 in $75,287,520$. (Much more rare than one-in-a-million!)

2.2 Chance of all-but-one overlap

What's the chance of having $k-1$ of the k preferences overlap? In this case, there is exactly 1 preference that will not align. There are k ways to choose which preference will not align, and $n-k$ ways to choose which "non-overlapping" preference it will be. So, the chance is

$$\frac{(n-k)k}{\binom{n}{k}} = (n-k)k\frac{(n-k)!k!}{n!}$$

In our example, this happens much more often, about 1 in $158,500$.

2.3 Chance of some overlap

What's the chance of having s of the k preferences overlap? There are $\binom{k}{k-s} = \binom{k}{s}$ ways to choose which preferences will not align, and $\binom{n-k}{k-s}$ ways to choose which "non-overlapping" preferences they will be instead. So, the chance is

$$\frac{\binom{k}{s}\binom{n-k}{k-s}}{\binom{n}{k}} = \frac{k!^2(n-k)!^2}{n!(n-2k+s)!s!(k-s)!^2} = \frac{\binom{n}{2k-s}\binom{2k-s}{k}\binom{k}{s}}{\binom{n}{k}^2}$$

One interpretation of the last expression is that you choose the $2k-s$ preferences of both people, which k of them are yours, and which s of the k are shared (out of $\binom{n}{k}$ ways to choose each person's preferences).

3 Scaling

Can we build any intuition for this solution?

3.1 High overlap ($k >> (k-s)$, $n > 2k$)

With high overlap, $\binom{k}{s} = \binom{k}{k-s} \approx \frac{k^{k-s}}{(k-s)!}$, so the chance is approximately

$$\frac{(k(n-k))^{k-s}}{(k-s)!^2} \frac{1}{\binom{n}{k}}$$

Here, allowing one more non-overlapping preference makes the chance of overlap about nk times as likely!

3.2 Many preferences to choose from ($n >> k$)

In this limit, $\frac{n!}{(n-k)!} \approx n^k$, so the chance of s of the k preferences overlapping is

$$\frac{k!^2}{s!(k-s)!^2} \frac{(n-k)!^2}{n!(n-2k+s)!} \approx \frac{k!^2}{s!(k-s)!^2} \frac{n^{k-s}}{n^k} = \binom{k}{s}^2 \frac{s!}{n^s} \approx \frac{\binom{k}{s}^2}{\binom{n}{s}}$$

Increasing the number n of preferences available makes this chance scale as n^{-s}; requiring another preference to overlap resembles randomly choosing another preference correctly (adding a factor of $1/n$).

3.3 Many preferences to choose from and low overlap ($n >> k$ and $k >> s$)

If we also desire just a few preferences to overlap, then the chance of overlap resembles

$$\frac{\binom{k}{s}^2}{\binom{n}{s}} \approx \frac{(k^2/n)^s}{s!}$$

The expression k^2/n represents the average overlap of two people, since any one preference has a k/n chance of matching with the other person. In this regime, the chance of having s preferences match follows the Taylor series expansion for $e^{(k^2/n)}$.

3.4 Gaussian interpretation

Consider the chance of picking from a "good" region of a finite-valued distribution, which has true chance k/n. This can be approximated by randomly sampling with sample size k, and seeing how much this overlap with the "good" region. As k and n grow large, the distribution of sample means (amount of overlap / sample size) should approximate a Gaussian centered at the true mean, k/n.

```
from scipy.special import comb
import matplotlib.pyplot as plt

def f1(n,k,s):
    return comb(k, s)*comb(n-k, k-s) / comb(n, k)

for k in range(10, 200, 10):
    plt.plot([i/k for i in range(k)], [f1(5*k, k, s) for s in range(k)])
```

The above code simulates the procedure in Python. See Figure 1 for the associated plot.

4 What's next?

Hope this makes you think about how easy (or hard) it is to find like-minded, unusual people. What other ideas strike you? Let me know at marwahaha@berkeley.edu . Have a nice day!

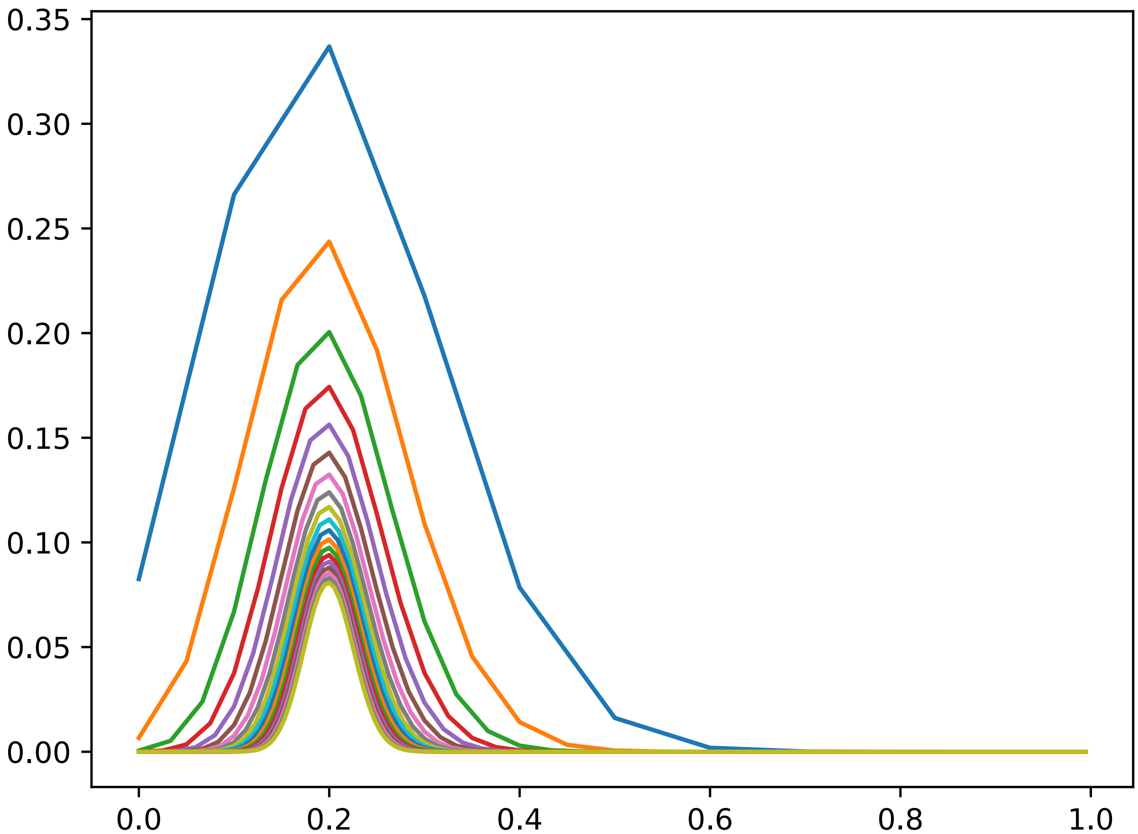

Figure 1: The average overlap per k as k increases, approximating a Gaussian with $\mu = k/n$. In this example, $n = 5k$.