Data Polarization in Movie Reviews

By Thomas Cochran, William Young, and Swaraj Patil

Our Purpose

- Subject:
  - Trying to find out if reviewers often give polarized ratings (extremely low or high) instead of a moderate one
- Motivations:
  - Online ratings can influence if someone watches movies (streaming platforms and in-person)
  - Polarized ratings don’t reflect true viewer sentiment
- Research Question:
  - Do movie reviewers tend to rate films at extreme values ≤1 star or 4.5+ star

Data Set: MovieLens from GroupLens

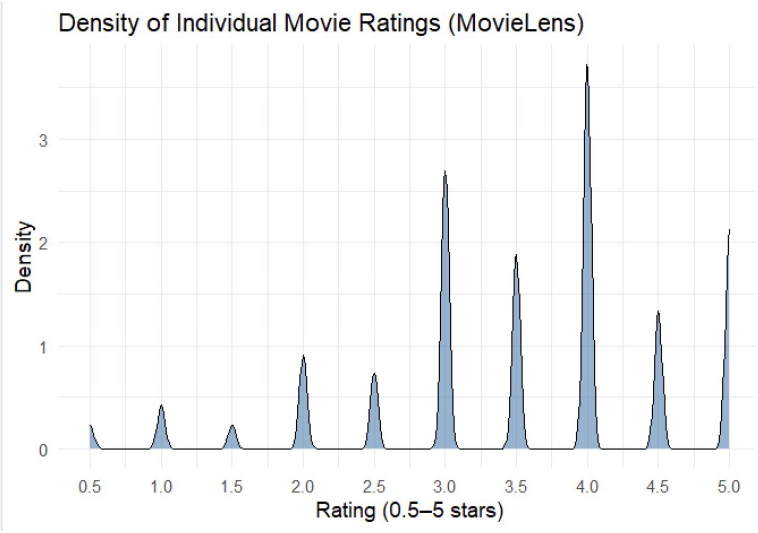
- MovieLens has over 33,000,000 reviews from about 34,000 users.
- Research group in the Department of Computer Science and Engineering at the University of Minnesota
- Components within the data:
  - userId, movieId, rating, timestamp, title, genres

userId	movieId	rating	timestamp
1	1	4.0	1225734739
1	110	4.0	1225865086
1	158	4.0	1225733503
1	260	4.5	1225735204

movieId	title	genres
1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy
2	Jumanji (1995)	Adventure|Children|Fantasy
3	Grumpier Old Men (1995)	Comedy|Romance
4	Waiting to Exhale (1995)	Comedy|Drama|Romance

Descriptive Statistics

- Around 71.4% or 23,892,382 users rated between 1-4.5.
- 4.0 is easily seen as the favorite vote.
- Possible bias when it comes to extreme values
    - Positivity Bias - quicker to praise a movie than hate it
- Not bell-curve shaped
    - Shows peaks




```
# A tibble: 3 x 3
  extreme                n   prop
  <chr>              <int> <dbl>
1 High (= 4.5)     8037731 0.240
2 Low (= 1)        1541499 0.0461
3 Middle (1-4.5) 23892382 0.714
```

Results

Part 1: Overall Extremes

- Approximately 28.6% of all ratings are extreme and the one sample prop. test gave p-value < 2.2e-16
- Reject null hypothesis of at most 20% of ratings being extreme.
- People give extreme ratings more frequently than a non-polarized system normally produces.

```
        1-sample proportions test without continuity correction

data:  extreme_count out of total_ratings, null probability 0.2
X-squared = 1554058, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is greater than 0.2
95 percent confidence interval:
 0.2860612 1.0000000
sample estimates:
        p
0.2861897
```

Part 2: High and Low Extremes

- When splitting extremes into separate tests, we see that data is nearly all on the high side
  - Strong evidence positive scores are overused
- Shows how reviews are actually *not* polarized but inflated with a concentration of very positive reviews for each user
- Large contrasts in P-values showed how high ratings were very statistically significant

```
        1-sample proportions test without continuity correction

data:  high_count out of total_ratings, null probability 0.2
X-squared = 336992, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is greater than 0.2
95 percent confidence interval:
 0.2400143 1.0000000
sample estimates:
        p
0.2401358
```

```
        1-sample proportions test without continuity correction

data:  low_count out of total_ratings, null probability 0.2
X-squared = 4957856, df = 1, p-value = 1
alternative hypothesis: true p is greater than 0.2
95 percent confidence interval:
 0.04599437 1.00000000
sample estimates:
        p
0.04605392
```

Our Next Steps

- We would like to have feedback on us choosing to measure whether the proportion of low ratings is greater than 20% and if it is a viable cutoff for having too many extreme ratings
- If user reviews are very inflated, would we consider them trustworthy to pick out truly great movies
- Investigate in depth the causes of positivity bias
    - More likely to review only when movie is enjoyed
- Decide what movies should truly be considered “must see” for a useful rating system
- Comparing our observations to an ideal rating system:
    - Very few movies are rated/considered masterpieces
    - Most movies fall in the middle range