How do "blockbusters" emerge: Understanding the emergence of popularity by analyzing movie revenue data

Raj Kumar Pan[a] and **Sitabhra Sinha**[b]

The Institute of Mathematical Sciences, C.I.T. Campus, Taramani, Chennai - 600 113 India
[a]rajkp@imsc.res.in
[b]sitabhra@imsc.res.in

Understanding how certain products or ideas become popular (i.e., gets adopted by a large section of the population), in comparison to many other qualitatively similar competitors, remains one of the most challenging questions in the social sciences. The availability of large volume of high-quality sales data and the ability to analyze this data using computational algorithms implementing sophisticated statistical techniques has meant that it is now possible to obtain detailed empirical picture of the above process. Commercially released motion pictures (movies) offer one of the most visible products where the dynamics of popularity can be investigated. Revenue collected by a movie during its theatrical run provides a well-defined quantitative measure of its popularity; it has the additional advantage that high-quality sales data is publicly available through multiple sources. Analysis of movie revenue also provides interesting comparison with other popularity distributions, such as, that of scientific papers in terms of the number of citations [1].

We have considered both the opening gross income (i.e., the revenue obtained by a movie in the first week after it is released in theaters) and the total gross income (i.e., the revenue collected over the entire period during which the movie was shown in theaters). Although total gross may be a better measure of movie popularity, the opening gross is often thought to signal the success of a particular movie. The rank-ordered distribution for the opening, as well as the total gross, show an approximate power law in the region where the top grossing movies are located [2]. However, when the data are aggregated together we observe the distribution (Fig.1, left) is better fit by a log-normal (similar to the observation of Redner vis-a-vis citations). The maximum likelihood estimates of the log-normal distribution parameters yield $\mu = 3.49$ and $\sigma = 1.00$. Further, we observe that the total gross distribution is just a scaled version of the opening distribution, which essentially implies that the popularity distribution of movies is decided at the opening itself [4]. An additional feature of interest is that both the opening and the total gross distributions are bimodal (Fig.1, left, inset), implying that most movies either do very well or very badly at the box office.

We next investigate whether the popularity of individual movies can be seen to be a function of their production quality (measured by the production budget). Although movies with higher budget tend to earn more, we see no significant correlation (the correlation coefficient is only 0.62). Further the determination of success of a movie on its opening implies the key role of pre-release advertising. Although the data for advertising budget is often unavailable, the data about the number of theaters where the movie is initially released gives an estimate, since the advertising cost will scale with this quantity. The correlation here is worse, indicating that advertising has often very little role to play in deciding the success or otherwise of a movie.

Finally, we look at how popularity evolves over time by analyzing the gross income per theater (Fig. 1, right). This better represents the dynamics of movie popularity compared

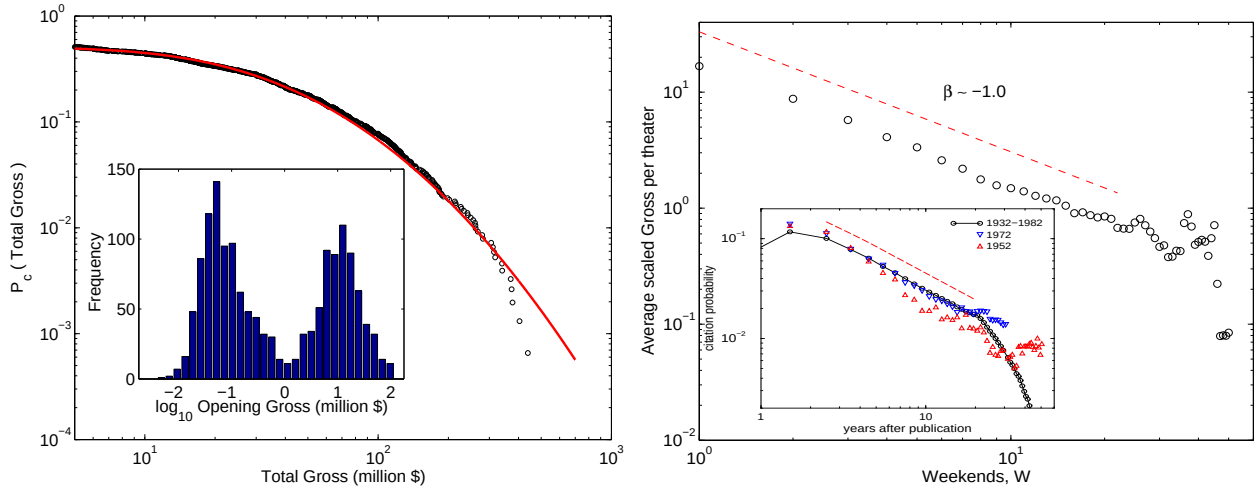

Figure 1: (Left) Cumulative distribution of total gross income for movies released across theaters in USA during 2000-2004, fitted with a log-normal curve. Inset shows the distribution of opening weekend gross. (Right) Average weekend gross per theater for a movie after it has run for W weekends. The initial decline follows a power-law with exponent $\beta \simeq -1$. Inset shows the probability that a paper will be cited t years after publication, which also shows a power law decay with exponent -0.94 [1].

to the time-evolution of the weekly overall gross income, because a movie that is being shown in a large number of theaters has a bigger income simply on account of higher accessibility for the potential audience. Unlike the overall gross that decays exponentially with time, the gross per theater shows a power-law decay [3] in time with exponent $\beta \simeq -1$. This has a striking similarity with the time-evolution of popularity for scientific papers in terms of citations. It has been reported that the citation probability to a paper published t years ago, decays approximately as $1/t$ [1]. In a very different context, namely, the decay in the popularity of a website (as measured by the rate of download of papers from the site) over time t has also been reported to follow an inverse power-law, but with a different exponent. It therefore suggests universal mechanisms at work in the popularity dynamics of very different objects. Indeed, our analysis reveals that the two key properties of popularity distributions that explain most of our observations are fairly general: (i) the bimodal log-normal nature of the distribution, and (ii) the power-law decay with time of sales. This provides benchmarks for future models of the emergence of popularity through self-organization in agent opinions.

Keywords
Sales data analysis, Movie revenue, popularity distribution, power-law, log-normal

References

[1] S. Redner, "Citation Statistics From More Than a Century of Physical Review," *Arxiv preprint physics/0407137*, 2004.
[2] S. Sinha, and S. Raghavendra, "Hollywood blockbusters and long-tailed distributions: An empirical study of the popularity of movies," *Eur. Phys. J. B*, 42: p. 293-296, 2004.
[3] S. Sinha and R.K. Pan, "Blockbusters, Bombs and Sleepers: The income distribution of movies," in *Econophysics of Wealth Distributions* (Eds. A. Chatterjee, B.K. Chakrabarti and S. Yarlagadda), Springer-Verlag Italia, Milan, p. 43-47, 2005.
[4] S. Sinha and R.K. Pan, "How a "Hit" is Born: The Emergence of Popularity from the Dynamics of Collective Choice," in *Econophysics and Sociophysics: Trends and Perspectives* (Eds. B.K. Chakrabarti, A. Chakraborti and A. Chatterjee), Wiley-VCH, Berlin, Chapter-15, p. 417-447, 2006.