Power-law distributions in empirical data
Clauset, Shalizi, Newman

This paper presents a statistical framework for identifying and analyzing power-law distributions in empirical data. The authors address a significant problem in scientific research being that power laws are frequently claimed across disciplines. However, these often rely on inadequate visual inspection of log-log plots rather than sound statistical methods. The authors develop a three-step approach: estimating power-law parameters using maximum likelihood estimation, testing quality of fit with a modified Kolmogorov-Smirnov statistic, and comparing the power-law model against alternative distributions using likelihood ratio tests. These proposed methods are validated through practical tests, albeit on synthetic data. A key insight of the paper is their method for accurately estimating the lower bound parameter xmin, which marks where power-law behavior begins. The authors show that previous approaches, particularly least-squares regression methods on log-transformed data, produce biased results and can falsely suggest power laws where none exist. The authors apply their framework to 24 empirical datasets from diverse fields that have been claimed to follow power laws. Interestingly, they find that only 17 of them are consistent with power-law behavior, many of which are compatible with alternative distributions like log-normal. This reveals that distinguishing between power laws and similar heavy-tailed distributions often requires careful statistical analysis. One limitation of the paper is that for some systems, the statistical tests may have limited power to discriminate between competing models unless sample sizes are very large. Additionally, the framework doesn't directly address truncated power laws or mixed distributions that might better describe certain phenomena. The main strength of the paper is its provision of a rigorous method that can be widely applied across disciplines. The authors have made their methods available as software which allows researchers to implement proper power-law analysis without specialized statistical knowledge. Overall, the work addresses a critical gap between claiming and verifying power-law behavior in empirical research.

A Brief History of Generative Models for Power Law and Lognormal Distributions
Michael Mitzenmacher

This paper provides a comprehensive overview of the historical development of generative models that produce power law and log-normal distributions, revealing deep connections between these distributions. A major insight is that similar generative processes can yield either distribution with only minor modifications. The paper identifies three main mechanisms for generating power laws: preferential attachment, optimization processes, and multiplicative processes with certain constraints. The author demonstrates that many new findings in computer science about power laws have historical predecessors from as early as the early 20th century. The author also shows how minor variations in generative models shift outcomes between distributions. For example, a multiplicative process typically yields a log-normal distribution, but adding a lower bound or randomizing the observation time transforms it into a power law. This shows why researchers across fields debate about whether certain phenomena follow power laws or log-normal distributions. The paper highlights the double Pareto distribution as a particularly useful model that combines features of both distributions, having a log-normal body with power law tails. This hybrid approach could resole many empirical debates about distribution classification. A limitation of this work is that while it

provides historical context and conceptual understanding, it doesn't offer definitive criteria for choosing between models in practical applications. The paper focuses primarily on theory, not statistical testing methods like the first paper, limiting its scope and perhaps practicality. The author makes a valuable contribution by connecting disparate research traditions and demonstrating how computer scientists, economists, linguists, etc. have often independently rediscovered similar principles. This interdisciplinary perspective reveals that power law and log-normal distributions aren't competing explanations, but instead, closely related outcomes from similar underlying processes. The paper serves as a historical review and cautionary tale about reinventing established concepts, encouraging future researchers to look past disciplinary boundaries when studying these patterns in systems.