Power Laws and the Forecasting of Epidemics

January 29, 2026

Abstract

conteúdo...

1 Introduction

Understanding the dynamics of infectious disease epidemics has lead to a very prolific literature both in terms of mechanistic mathematical models and statistical models(Siettos and Russo, 2013). The former attempting to capture the causal chain of events involved in disease transmission and the latter trying to make sense of growing collection of data about infectious disease incidence.

Among the most salient result of the mathematical modeling of infectious diseases was the definition of the basic reproductive number, R_0, that determines the most important characteristics of epidemics such as its spreading speed and the epidemic threshold, also know as herd immunity threshold. The R_0 appears in equations governing prevalence growth, as an exponent determining the initial exponential growth of an epidemic. It also defines an important bifurcation points $R_0 = 1$ linking the to the existence and stability of equilibria in epidemic models.

On the statistical side of the literature related to infectious disease modeling, an important if not the biggest portion of published literature deals with statistical modeling. for the first two thirds of the 20th century, the literature consisted mainly of parametric frequentist models, with Bayesian statistics gaining more preeminence towards the end of century[REFS].

One of the most commonly modelled kind of epidemic data is incidence data, i.e., the number of cases accumulated over time and space or studied as a stochastic process, such as a Poisson of Negative Binomial processes (Becker and Britton, 1999; Li et al., 2012).

One particular challenge when modeling inference data, is that it follows a power law distribution(Rhodes et al., 1997; Clauset et al., 2009). The very heavy tail of this distribution means most models have difficulty predicting extreme values, which are the most important in epidemiology as they cause the most disease and consequent economic burden.

In this paper we study the statistical distribution of dengue epidemics on a very large dataset of thousands of real epidemics in Brazilian cities of various sizes over more than a decade. Our goal is to determine how well the power law distribution fits incidence data, be it in the form of distribution of weekly case counts or the total size of seasonal epidemics. We also look at the distribution of the power law distribution and how it correlate with other epidemic characteristics. Finally we will discuss how the estimation of these power law distribution can improve the prediction of future incidence, when compared to more traditional forecasting models adopting more tractable distributions, such as log-normal, Poisson and others to approximate the real data distribution.

2   Methods

Based on data made available through the Mosqlimate platform(Araujo et al., 2025) [cite Mosqlimate paper], We have fitted power laws to epidemics to dengue incidence data from 2010 to the present.

2.1   Power law fitting

We follow the methodology of Clauset et al.Clauset et al. (2009) to fit a power law distribution to two datasets: the weekly case counts, and the yearly cumulative incidence, that is the number of cases accumulated over an epidemic year.

3 Results

4 Discussion

5 Conclusion

References

Araujo, E. C., Codeço, C. T., Loch, S., Vacaro, L. B., Freitas, L. P., Lana, R. M., Bastos, L. S., de Almeida, I. F., Valente, F., Carvalho, L. M., et al. (2025). Large-scale epidemiological modelling: scanning for mosquito-borne diseases spatio-temporal patterns in brazil. *Royal Society Open Science*, 12(5):241261.

Becker, N. G. and Britton, T. (1999). Statistical studies of infectious disease incidence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(2):287–307.

Clauset, A., Shalizi, C. R., and Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, 51(4):661–703.

Li, Y., Brown, P., Gesink, D. C., and Rue, H. (2012). Log gaussian cox processes and spatially aggregated disease incidence data. *Statistical methods in medical research*, 21(5):479–507.

Rhodes, C. J., Jensen, H. J., and Anderson, R. M. (1997). On the critical behaviour of simple epidemics. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264(1388):1639–1646.

Siettos, C. I. and Russo, L. (2013). Mathematical modeling of infectious disease dynamics. *Virulence*, 4(4):295–306.