1 Introduction

2 Box: Kappa tutorial

3 Results

Demographic stochasticity can generate "emergent" heterogeneity even in the absence of explicit differences between individual-based rates. In simple models, this heterogeneity can be characterized in including negative binomial funky distributions of activity. We explicate the notion that this is predictable (see Box)
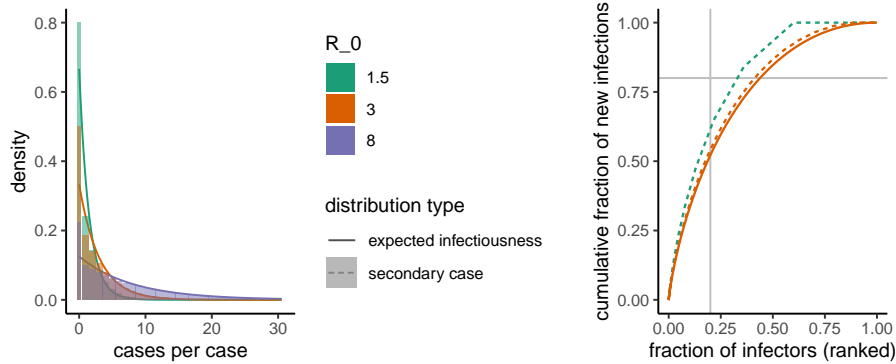

Figure 1: **Heterogeneity emerges even from a simple, linearized compartmental model** due to implicit variation in recovery times among infectors. (left) Activity distributions (density curves) and secondary case distributions (density histograms) for the outset of an SIR epidemic. Because the first bin (at zero) sits at the boundary of support for each distribution, we have plotted this bin as double the density and half the width; this adjustment preserves area-to-area correspondence with the PDF, while facilitating visual comparison of the heights of the density and mass functions. (right) Inequality curves for *activity* distributions from SIR models with differing $\mathcal{R}_0$ are identical (and indestinguishable due to overplotting); inequality in the *case* distribution decreases with R0 towards the theoretical limit of the activity distribution.

But despite differences in a non-dynamic world, we find invariance in case-per-case when looking across the entire epidemic

What is the cause of this surprise? We claim: different relative contributions of variability from between and within-cohorts across R0 (bottom C of legacy/figures/emergentHetPoster.pdf). JD-Azadeh: can you work on putting some code into kappa that can do this?

JD-Azadeh: I remain confused by this picture. If we look at all the time up until 1/4, μ and σ^2_{with} are large, and changes in μ are also pretty large for the
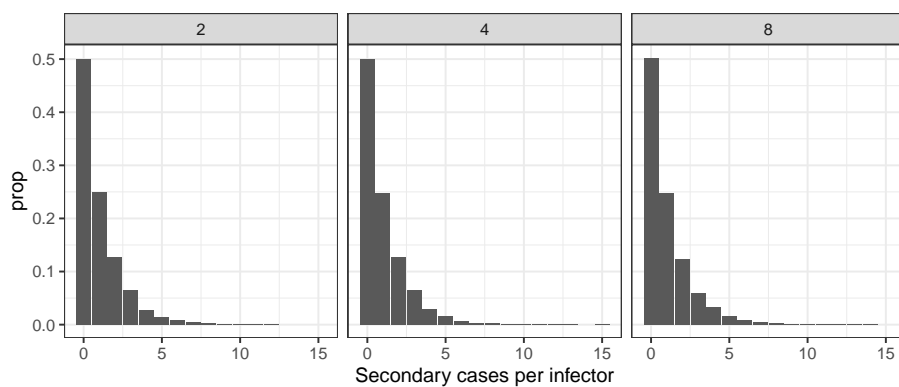
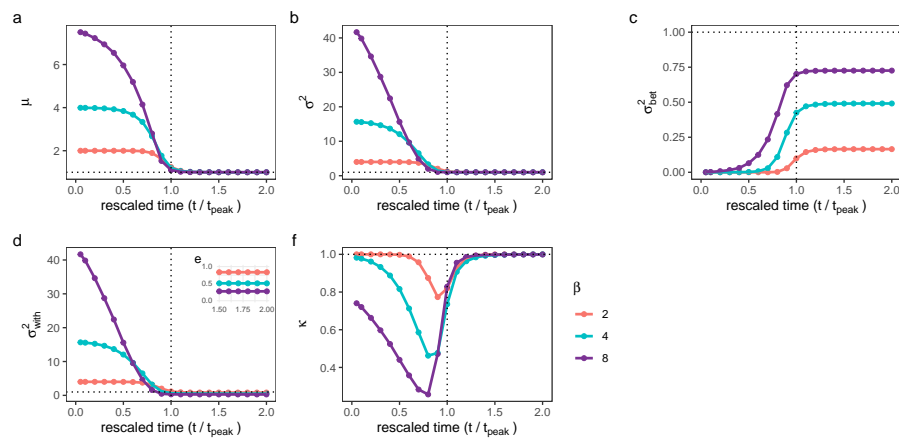Figure 2: Some histograms. Look at poster text and see what we think.

Figure 3: How components of variance are changing through time

large β. How can σ^2_{btw} be that small?

We are also interested in what emergent distributions will look like to people studying outbreaks in real time. We are interested, at least to some extent, both in how cohorts change through time, and in what the outbreak will "look like" if we observe from a particular time.

Observing from a particular time can be done in two ways: either naively, or by trying to correct for the truncation of observations. These can be simulated, respectively, by either simply stopping the simulation at a certain time (or reporting what would be seen if we did), or in an idealized world, by looking at all the cohorts infected up until a given time. It's worth looking at some pictures of both of these views and seeing what we think. It may also be worth looking at statistics for individual cohorts (I guess this is a bit boring, because we only have within-cohort variation in that case, but we should do it and put in the supp).

It's also possible to imagine realistic approaches between these two extremes, but let's put that off for later. There are methods (including by Dushoff and Park) for thinking about this at the cohort level, but not with a focus on individual variation. Maybe this is just for discussion. OR maybe we should also look at plots where we go up until a particular time and only count recovered infectors.

TG: Can we make a note about for epidemics with large R0, if you don't start tracking cases right from the beginning, you'll already underestimate cases/case JD: Yes, this should go into the paper.

4   Box (or appendix?) Tapan's proof?

5   Discussion