The epidemic lambda-coalescent model

Xavier Didelot[1,2,*], Ian Roberts[2], ...

[1] School of Life Sciences, University of Warwick, United Kingdom

[2] Department of Statistics, University of Warwick, United Kingdom

[*] Corresponding author. Tel: 0044 (0)2476 572827. Email: xavier.didelot@gmail.com

Running title: Epidemic lambda-coalescent

Keywords: lambda-coalescent model; pathogen phylogenetics; multiple mergers; offspring distribution

1 Introduction

Superspreading in infectious disease epidemiology (Lloyd-Smith et al. 2005). For example SARS-CoV-2 superspreading (Wang et al. 2020; Lemieux et al. 2021; Gómez-Carballa et al. 2021). Coalescent model (Kingman 1982a,b). Work by Li and Fraser (Li et al. 2017; Fraser and Li 2017). Lambda-coalescent models (Pitman 1999; Sagitov 1999; Donnelly and Kurtz 1999). Beta-coalescent (Schweinsberg 2003) is a specific type of Lambda-coalescent. Was used in (Hoscheit and Pybus 2019) and (Menardo et al. 2021). David's paper (Helekal et al. 2024).

2 Coalescence probabilities

2.1 General case

Discrete time t. Non-overlapping generations of infected individuals. At time t there are N_t infected individuals. Each of them creates a number $k_{t,i}$ of secondary infections at time $t+1$, following the offspring distribution $\alpha_t(k)$. The mean of this distribution is the basic reproduction number R_t and the variance is v_t. We have:

$$N_{t+1} = \sum_{i=1}^{N_t} k_{t,i} \tag{1}$$

Let $p_{n,t}$ be the probability that n individuals at time $t+1$ have the same infector at time t.

2.2 Poisson case

Here the offspring distribution is Poisson(R_t).

The probability of coalescence for two lines is:

$$p_{2,t} = \frac{1}{N_t} \tag{2}$$

The probability of coalescence for n lines is:

$$p_{n,t} = \frac{1}{N_t^{n-1}} \tag{3}$$

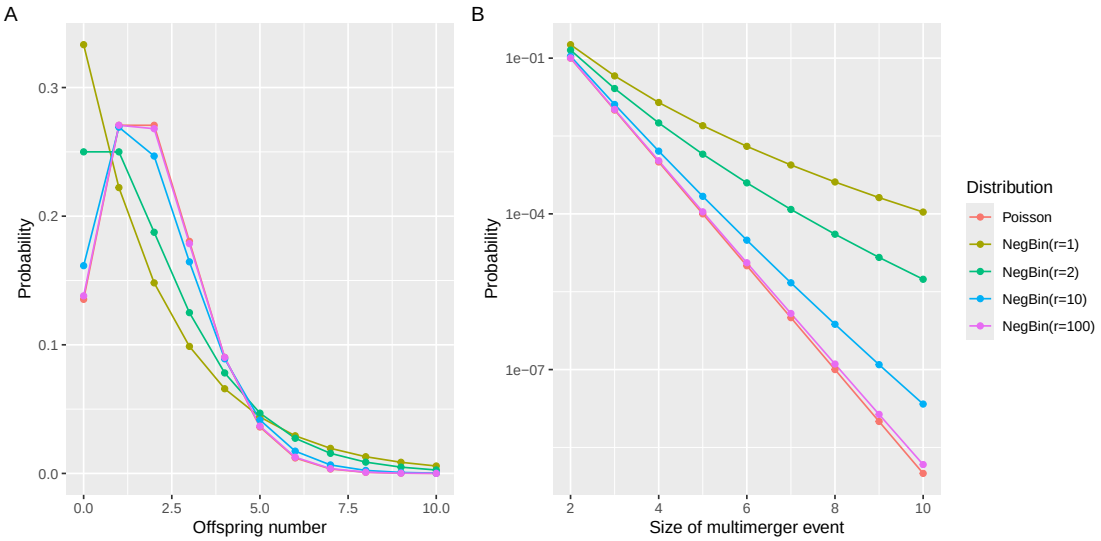

Figure 1: (A) Offspring distribution. (B) Probability of coalescence.

2.3 NegBin case

Here the offspring distribution is Negative-Binomial with mean R_t and variance v_t. The parameters of this distribution are $r = R_t^2/(v_t - R_t)$ and $p = R_t/v_t$.

The probability of coalescence for two lines is:

$$p_{2,t} = \frac{r+1}{N_t r + 1} \tag{4}$$

The probability of coalescence for n lines is:

$$p_{n,t} = \frac{\prod_{i=1}^{n-1} r + i}{\prod_{i=1}^{n-1} N_t r + i} \tag{5}$$

2.4 Example

See Figure 1.

3 Implementation

We implemented the analytical methods described in this paper in a new R package entitled *EpiLambda* which is available at `https://github.com/xavierdidelot/EpiLambda` for R version 3.5 or later. All code and data needed to replicate the results are included in the "run" directory of the *EpiLambda* repository.

4 Discussion

Acknowledgements

We acknowledge funding from the National Institute for Health Research (NIHR) Health Protection Research Unit in Genomics and Enabling Data.

References

Donnelly, P., Kurtz, T.G., 1999. Particle Representations for Measure-Valued Population Models. The Annals of Probability 27. doi:10.1214/aop/1022677258.

Fraser, C., Li, L.M., 2017. Coalescent models for populations with time-varying population sizes and arbitrary offspring distributions. bioRxiv , 10.1101/131730doi:10.1101/131730.

Gómez-Carballa, A., Pardo-Seco, J., Bello, X., Martinón-Torres, F., Salas, A., 2021. Superspreading in the emergence of COVID-19 variants. Trends in Genetics 37, 1069–1080. doi:10.1016/j.tig.2021.09.003.

Helekal, D., Koskela, J., Didelot, X., 2024. Inference of multiple mergers while dating a pathogen phylogeny. bioRxiv , 2023.09.12.557403doi:10.1101/2023.09.12.557403.

Hoscheit, P., Pybus, O.G., 2019. The multifurcating skyline plot. Virus Evolution 5, 1–10. doi:10.1093/ve/vez031.

Kingman, J., 1982a. The coalescent. Stochastic Processes and their Applications 13, 235–248. doi:10.1016/0304-4149(82)90011-4.

Kingman, J.F.C., 1982b. On the genealogy of large populations. Journal of Applied Probability 19, 27–43. doi:10.2307/3213548.

Lemieux, J.E., Siddle, K.J., Shaw, B.M., Loreth, C., Schaffner, S.F., Gladden-Young, A., Adams, G., Fink, T., Tomkins-Tinch, C.H., Krasilnikova, L.A., DeRuff, K.C., Rudy, M., Bauer, M.R., Lagerborg, K.A., Normandin, E., Chapman, S.B., Reilly, S.K., Anahtar, M.N., Lin, A.E., Carter, A., Myhrvold, C., Kemball, M.E., Chaluvadi, S., Cusick, C., Flowers, K., Neumann, A., Cerrato, F., Farhat, M., Slater, D., Harris, J.B., Branda, J.A., Hooper, D., Gaeta, J.M., Baggett, T.P., O'Connell, J., Gnirke, A., Lieberman, T.D., Philippakis, A., Burns, M., Brown, C.M., Luban, J., Ryan, E.T., Turbett, S.E., LaRocque, R.C., Hanage, W.P., Gallagher, G.R., Madoff, L.C., Smole, S., Pierce, V.M., Rosenberg, E., Sabeti, P.C., Park, D.J., MacInnis, B.L., 2021. Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. Science 371, eabe3261. doi:10.1126/science.abe3261.

Li, L.M., Grassly, N.C., Fraser, C., 2017. Quantifying Transmission Heterogeneity Using Both Pathogen Phylogenies and Incidence Time Series. Molecular Biology and Evolution 34, 2982–2995. doi:10.1093/molbev/msx195.

Lloyd-Smith, J., Schreiber, S., Kopp, P., Getz, W., 2005. Superspreading and the effect of individual variation on disease emergence. Nature 438, 355–9. doi:10.1038/nature04153.

Menardo, F., Gagneux, S., Freund, F., 2021. Multiple Merger Genealogies in Outbreaks of Mycobacterium tuberculosis. Molecular Biology and Evolution 38, 290–306. doi:10.1093/molbev/msaa179.

Pitman, J., 1999. Coalescents with multiple collisions. The Annals of Probability 27, 1870–1902.

Sagitov, S., 1999. The general coalescent with asynchronous mergers of ancestral lines. Journal of Applied Probability 36, 1116–1125. doi:10.1239/jap/1032374759.

Schweinsberg, J., 2003. Coalescent processes obtained from supercritical Galton–Watson processes. Stochastic Processes and their Applications 106, 107–139. doi:10.1016/S0304-4149(03)00028-0.

Wang, L., Didelot, X., Yang, J., Wong, G., Shi, Y., Liu, W., Gao, G.F., Bi, Y., 2020. Inference of person-to-person transmission of COVID-19 reveals hidden super-spreading events during the early outbreak phase. Nature Communications 11, 5006. doi:10.1038/s41467-020-18836-4.