STATS 700-002 Class 3.
The Volz Coalescent

Edward Ionides and Aaron King

September 11, 2025

The Volz Coalescent

Volz, E. M., Kosakovsky Pond, S. L., Ward, M. J., Leigh Brown, A. J., & Frost, S. D. (2009). Phylodynamics of infectious disease epidemics. Genetics, 183(4), 1421-1430. https://doi.org/10.1534/genetics.109.106021.

- Overview: Kimgman's coalescent provides a link between the observed rate of coalescence, the *effective population size* and the generation interval, in a fixed-size population.

The Volz Coalescent

Volz, E. M., Kosakovsky Pond, S. L., Ward, M. J., Leigh Brown, A. J., & Frost, S. D. (2009). Phylodynamics of infectious disease epidemics. Genetics, 183(4), 1421-1430. https://doi.org/10.1534/genetics.109.106021.

- Overview: Kimgman's coalescent provides a link between the observed rate of coalescence, the *effective population size* and the generation interval, in a fixed-size population.
- A time-varying population can be approximated as piecewise constant. Volz et al. (volz09) noticed that this can be inappropriate for fast-changing populations and derived an alternative approach.

Cluster size distributions

- Why are (Volz et al., 2009) so interested in cluster size distribution?

The ancestor function

- $A(t, T)$ is defined as the "fraction of the population at t with sampled progeny extant at T.

  Does "population" here mean "host population, N" or "population of infected hosts, $\mathcal{I}$"?

Intuition

- From the introduction, “Coalescent rates are low near peak prevalence, but higher when there is a large ratio of incidence to prevalence.”

  How is this evident?

Definition or derivation

- Consider equation (2):

$$dA(t,T)/dt = f_{SI}p_c.$$

Is this a definition of $A(t,T)$ for some model, or is it derived?

Definition or derivation

- Consider equation (2):

$$dA(t,T)/dt = f_{SI}p_c.$$

  Is this a definition of $A(t,T)$ for some model, or is it derived?

- Either way, once a model gives us $p_c(t,T)$, this defines a coalescent process having coalescent rate $f_{SI}p_c$.

Definition or derivation

- Consider equation (2):

$$dA(t,T)/dt = f_{SI}p_c.$$

  Is this a definition of $A(t,T)$ for some model, or is it derived?

- Either way, once a model gives us $p_c(t,T)$, this defines a coalescent process having coalescent rate $f_{SI}p_c$.

- Note, this depends on the expected number of lineages (normalized as the large sample fraction) not the actual number.

Likelihood

- Equation (12) is written in terms of $A(t, T)$, which is a property of the model not of the data.

  How can this be considered a likelihood?

Comparison with the skyline plot

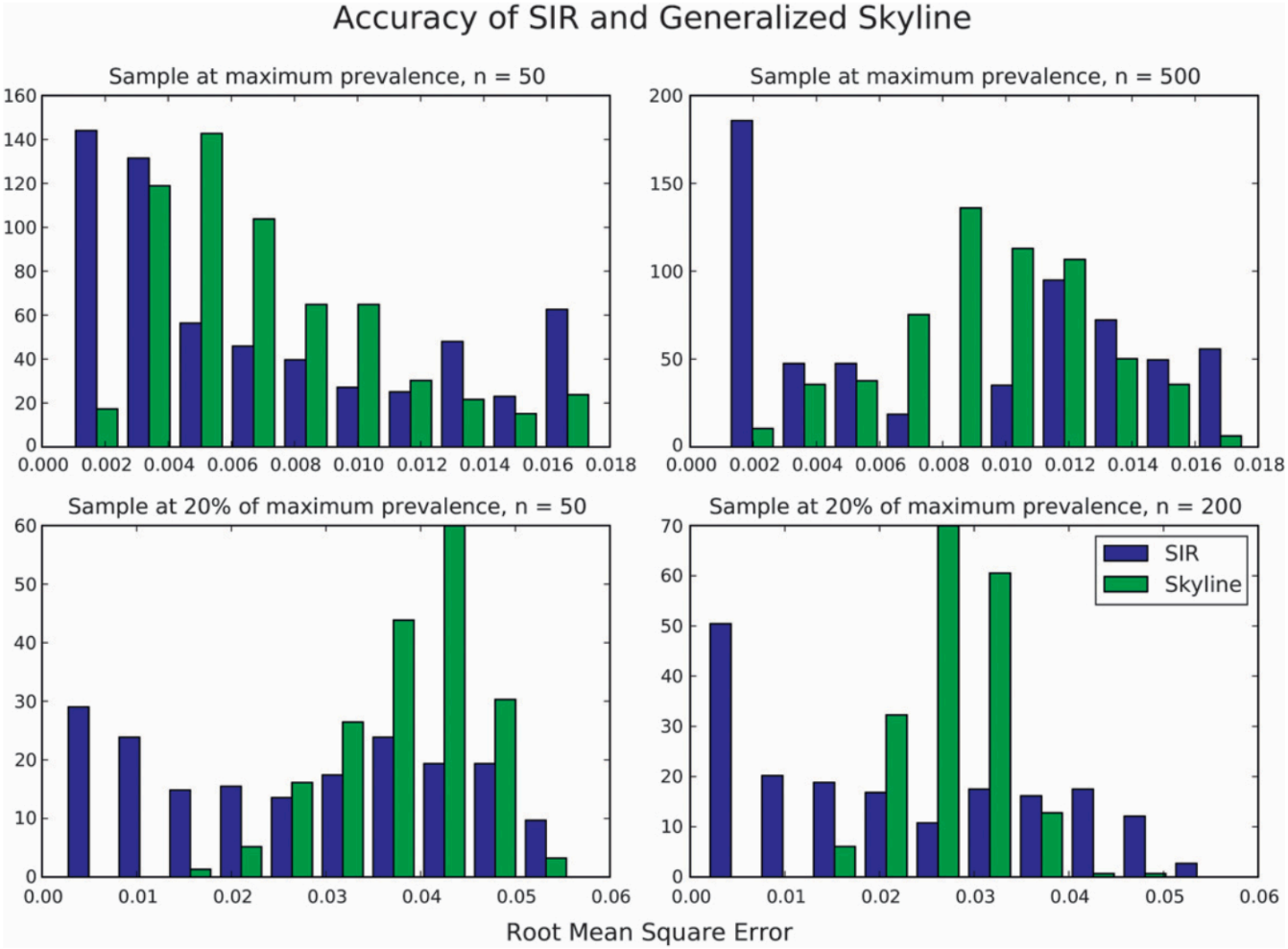

FIGURE 3.—Root mean square error of SIR and generalized skyline estimates of epidemic prevalence. Data are based on 300 simulated epidemics ($R_0 = 2$). RMSE is averaged over 100 time points.

References I

Volz, E. M., Kosakovsky Pond, S. L., Ward, M. J., Leigh Brown, A. J., and Frost, S. D. W. (2009). Phylodynamics of infectious disease epidemics. *Genetics*, 183(4):1421–1430.