



Jun 1st, 9:00 AM - 9:30 AM

# Geometry of the sample frequency spectrum and the perils of demographic inference

Zvi H. Rosen

*University of California, Berkeley, zhrosen@berkeley.edu*

Follow this and additional works at: <https://scholarscompass.vcu.edu/bamm>



Part of the [Life Sciences Commons](#), and the [Medicine and Health Sciences Commons](#)

---

<https://scholarscompass.vcu.edu/bamm/2018/friday/4>

This Event is brought to you for free and open access by the Dept. of Mathematics and Applied Mathematics at VCU Scholars Compass. It has been accepted for inclusion in Biology and Medicine Through Mathematics Conference by an authorized administrator of VCU Scholars Compass. For more information, please contact [libcompass@vcu.edu](mailto:libcompass@vcu.edu).

# Geometry of the sample frequency spectrum and the perils of demographic inference

Zvi Rosen, Anand Bhaskar, Sebastien Roch, Yun S. Song

March 14, 2018

The sample frequency spectrum (SFS), which describes the distribution of mutant alleles in a sample of DNA sequences, is a widely used summary statistic in population genetics. The expected SFS has a strong dependence on the historical population demography and this property is exploited by popular statistical methods to infer complex demographic histories from DNA sequence data. We characterize the geometry of the expected SFS for piecewise-constant demographic histories and use our results to explain some pathological behavior of these statistical inference methods. Using tools from convex and algebraic geometry, we demonstrate that the expected SFS of a sample of size  $n$  under an arbitrary population history can be recapitulated by a piecewise-constant demography with only  $\kappa(n)$  epochs, where  $\kappa(n)$  is between  $n/2$  and  $2n - 1$ . Meanwhile, the set of expected SFS for piecewise-constant demographies with fewer than  $\kappa(n)$  epochs is open and non-convex, leading to pathological behavior.