UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Demographic Inference from Large Samples: Theory and Methods

Permalink
https://escholarship.org/uc/item/8nb832p7

Author
Terhorst, Jonathan G.

Publication Date
2017

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Demographic Inference from Large Samples: Theory and Methods

by

Jonathan Terhorst

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Statistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Yun S. Song, Chair
Professor Michael I. Jordan
Professor Rasmus Nielsen
Professor Montgomery Slatkin

Spring 2017

Demographic Inference from Large Samples: Theory and Methods

Copyright 2017
by
Jonathan Terhorst

Abstract

Demographic Inference from Large Samples: Theory and Methods

by

Jonathan Terhorst

Doctor of Philosophy in Statistics

University of California, Berkeley

Professor Yun S. Song, Chair

The emergence of next-generation sequencing has revolutionized our ability to interrogate the genome, leading to fascinating new insights into the nature of humans and many other species. At the same time it has created theoretical and computational challenges associated with the need to perform robust and efficient inference on increasingly massive genetic data sets.

In this thesis I focus on those challenges in the context of demographic inference, which estimates the past history of a population on the basis of genetic data sampled at the present. I derive the basic theoretical models which enable such inferences. I then refine them to formulate a new inference procedure for reconstructing size history using hundreds of whole genomes at a time, a significant increase over existing methods. I complement this algorithmic advance with some theoretical results on the accuracy of demographic inference as sample sizes grow large.

To my parents.

Contents

Contents ii

List of Figures iv

List of Tables v

1 Introduction 1

2 Robust and scalable inference of population history 4
2.1 Motivation . . . 4
2.2 Results . . . 5
2.3 Discussion . . . 11
2.4 Methods . . . 12
2.5 Emission Probabilities . . . 18
2.6 Locus-Skipping EM Algorithm . . . 27
2.7 Extension to multiple populations . . . 30

3 A fundamental limit on accuracy 49
3.1 Motivation . . . 49
3.2 Preliminaries . . . 50
3.3 Related work . . . 52
3.4 Main Results . . . 53
3.5 Discussion . . . 57
3.6 Proofs . . . 58

4 A nonparametric perspective 63
4.1 The site frequency spectrum: diffusion argument . . . 63
4.2 Approximate likelihood of the SFS . . . 65
4.3 Theory of linear inverse problems . . . 67
4.4 Numerical results . . . 71
4.5 Spectral decay of the SFS operator . . . 74
4.6 Appendix . . . 84

Bibliography **87**

List of Figures

1.1 Example size history 2

2.1 Effect of phasing error 33
2.2 Performance of SMC++, MSMC and $\partial a \partial i$ 34
2.3 Split time estimation 35
2.4 Computational performance 36
2.5 Inference of human demography 37
2.6 Inference of human divergence times 38
2.7 Demographic inference for other species 39
2.8 Demographic inference when ρ is unknown 40
2.9 Demographic inference in Africa 41
2.10 Demographic inference in Asia 42
2.11 Demographic inference in Europe 43
2.12 Sensitivity analysis for human demographic inference 44
2.13 Demographic inference for other species 45
2.14 Schematic of differences between PSMC, MSMC and SMC++ 46

3.1 A family $\mathcal{U}_{I,J}$ of piecewise-constant models with $K = I + J$ epochs. 54

4.1 Top five singular functions for $n = 50$ 69
4.2 Piecewise size history functions used to test approximability. 71
4.3 Projection of η_1, η_2, η_3 onto singular components 72
4.4 Performance of spectral cutoff estimator, low-noise case. 74
4.5 Performance of spectral cutoff estimator, high-noise case. 74
4.6 Actual vs. predicted spectral decay for the operator EE^*, $n = 50$ 77
4.7 Bernstein and Jacobi polynomials 78
4.8 Derivatives of $B_{30,10}(x)$ 79
4.9 Actual / bound estimates of singular values of $\mathbf{W}$ 79

List of Tables

2.1 Posterior decoding 47
2.2 Description of modern human data sets analyzed. 48
2.3 Description of other data sets analyzed. 48

Acknowledgments

It is a pleasure to acknowledge the many mentors, friends and loved ones who have made the past six years so rewarding and enjoyable. First and foremost I would like to thank Prof. Yun Song. When people ask me what my adviser is like, I have a one-word answer: "Fantastic." Yun is an exemplary scientist and teacher, and his patience, work ethic, intuition and acuity never ceased to impress me all throughout graduate school. I'm particularly grateful to have had such an inspiring role model as I transition to the next phase of my academic career and begin to advise students of my own. (And did I mention his scotch collection?)

Great advisers attract great students, and I have been fortunate to meet many of them in Yun's group. In particular, I'd like to thank Anand Bhaskar, Jack Kamm, Matthias Steinrücken, Sarah Sheehan, Geno Guerra, Jeff Spence, and Jeff Chan for collaboration, interesting conversations, and fun travel adventures over the years.

One of the things I have most enjoyed about graduate school is the opportunity to form lasting friendships, which otherwise becomes increasingly difficult as we grow older and busier. Adam, Ameet, and Ginger, thanks for all the laughs and good times. I'll cherish those memories forever, and look forward to the adventures to come!

I would like to thank my parents, Alexandra and Peter, for their tireless encouragment and support all throughout my life. It's difficult for me to express the gratitude I feel for all that they have sacrificed in order that their children would grow up healthy and content. I hope that this, in some small way, makes it worthwhile. To my siblings, Telle and Jerry, the two funniest people I know: thanks so much for providing much-needed stress relief and perspective over the past six years. I'm so lucky to have you all in my life.

Last and most important, none of this would have been possible without the affection and support of my wife, best friend, and love of my life, Sarah. I can't believe it's been 20 years since we were sitting in freshman-year Spanish together. Here's to many more.



Chapter 1

Introduction

Encoded in the genome of every living organism is a wealth of information about its antecedents. Unlocking this information promises to provide exciting new insights into the history of humans and many other species. Apart from its intrinsic interest, such knowledge is useful for understanding patterns of historical migration (Tennessen et al., 2012; 1000 Genomes Project Consortium, 2010; Skoglund et al., 2015; Raghavan et al., 2015), natural selection (Tennessen et al., 2012; Huerta-Sanchez et al., 2014; Racimo et al., 2015), the relationship between humans and other hominids (Green et al., 2010; Prüfer et al., 2014; Sankararaman et al., 2014; Vernot and Akey, 2014), and even distantly related phenomena like global climate change (Miller et al., 2012; Stewart and Stringer, 2012).

In order to recover such information, scientists rely on statistical inference procedures which provide a link between

1. Events experienced by a population over the course of its history, and
2. Genetic mutation patterns observed by sampling the genomes of members of that population.

The field of statistics which focuses on developing these procedures is known as *demographic inference*.

The past decade has witnessed a revolution in demographic inference, along with every other area of quantitative genetics. High throughput sequencing has massively increased our ability to acquire genetic data, at a cost which is falling faster than Moore's law. The exciting scientific opportunities ushered in by high-throughput sequencing bring with them some methodological challenges. In particular, the increased availability of genetic data has resulted in a corresponding focus in statistical genetics on scalability and computational performance when analyzing large numbers of genomes. Statisticians face the challenge of developing methods which can make full use of the richness of modern genetic data sets, while remaining computationally stable and efficient.

In this thesis we will focus particularly on the task of inferring the *size history* of a population from a large sample of genomes. The size history is a function $\eta : [0, \infty) \to (0, \infty)$ such

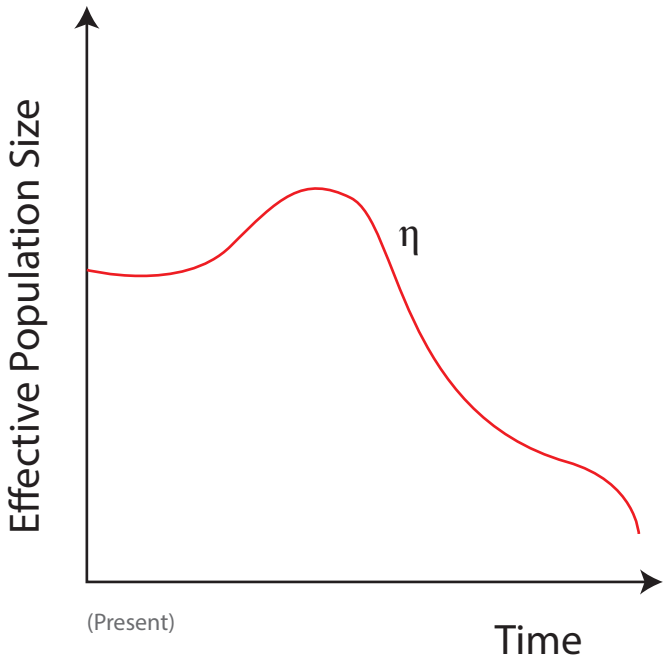


Figure 1.1: An example of a size history function $\eta(t)$.

that $\eta(t)$, under an idealized model of reproductive dynamics, gives the number of individuals alive in the population t generations in the past. (Hence, $\eta(0)$ is the number of individuals alive today.) An example size history is plotted in Figure 1.

Accurate estimation of the size history is an interesting scientific question in its own right, and is also a necessary precursor to identifying areas of the genome that are evolving non-neutrally due to, for example, the action of natural selection. This in turn has many implications for our understanding of diseases, evolution, and other fundamental questions in biology.

Chapter 2 of this thesis describes SMC++ (Terhorst, Kamm, and Song, 2017), a new algorithm designed to infer size history from the whole genome sequences of hundreds of individuals. Compared to existing methods, this is a two order-of-magnitude increase in the analyzable sample size. We demonstrate that the ability to handle larger data sets leads to more accurate and robust estimates, particularly in the recent past. Another important feature of SMC++ is the ability to analyze unphased genotype data, which is especially important when studying species for which a high quality reference panel does not already exist. Theoretically, the main innovation of SMC++ is a novel probabilistic framework for coupling the genealogical process in an individual with auxiliary genotype information from a large panel of related individuals.

As statisticians, our natural inclination when faced with an estimator is to wonder about its properties—is it unbiased? Consistent? What is the rate of convergence? Answering these questions is crucial to quantifying the gain made by algorithms such as SMC++ in analyzing ever larger sample sizes. Chapters 3 and 4 are devoted to providing some answers in this direc-

tion. By analyzing a simpler type of inference procedure which operates only on the so-called *site frequency spectrum*, we can derive some theoretical results about its performance.

Chapter 3 studies the minimax rate of convergence for size history estimators which operate only on the frequency spectrum (Terhorst and Song, 2015). The main result is a minimax lower bound on the convergence rate of *any* such estimator. We show that the lower bound is $O(1/\log s)$, where s is the number of independent segregating sites used to calculate the frequency spectrum. (The number of sampled individuals n does not enter into the bound.) This is exponentially worse than the convergence rate of many statistical estimation problems, which often decay polynomially in the sample size. This result highlights the need to exercise caution when interpreting the results of size history estimation which do not take linkage information into account.

In Chapter 4 we try to complement this negative result with a positive one. Specifically, we analyze the site frequency spectrum from the point of view of nonparametric estimation, where, in the limit as the mutation rate tends to zero, the expected unnormalized frequency spectrum is approximately given by the action of a linear operator on a (time-rescaled version of) the size history function η. This enables us to derive a regularized estimator for η which is asymptotically minimax under some additional assumptions on the function class containing η. We also give numerical evidence that this operator is very poorly-conditioned, in the sense that its singular values decay exponentially quickly towards zero. Finally, we complement these numerical estimates with a theoretical discussion illustrating why this is the case.

Chapter 2

Robust and scalable inference of population history

In this chapter we describe SMC++ (Terhorst, Kamm, and Song, 2017), an algorithm for inferring demographies using hundreds of unphased whole genomes.

2.1 Motivation

Over the past few years, there has been a surge of interest in using whole-genome sequence data from multiple individuals to learn about population demographic histories. There are two popular approaches to this problem: one based on the Poisson Random Field model (PRF; S. A. Sawyer and Hartl, 1992) using the sample frequency spectrum (SFS; R. C. Griffiths and Tavaré, 1994), and the other based on the sequentially Markov coalescent (Wiuf and Hein, 1999; McVean and Cardin, 2005; P. Marjoram and Wall, 2006). SFS/PRF methods (Gutenkunst et al., 2009; Excoffier et al., 2013) employ results from the diffusion or coalescent theory to characterize the sampling distribution of unlinked segregating sites in a random sample of DNA sequences. This distribution can be efficiently computed for a very large sample size (Bhaskar, Wang, and Song, 2015) and for complex demographic models with multiple populations (Kamm, Terhorst, and Song, 2016), but the model is incorrect when applied to modern whole-genome sequence (WGS) data due to correlation among neighboring sites. Also, the number of parameters that can be estimated using SFS-based methods is bounded by the sample size.

With the increasing prevalence of WGS data, attention has shifted to more complex models that incorporate recombination and linkage disequilibrium (LD) information. It has been shown that this approach can reveal past population history in unprecedented detail (Li and Durbin, 2011). Owing to the linear structure of DNA, these models commonly employ some form of hidden Markov model (HMM) (Dutheil et al., 2009; Li and Durbin, 2011; Schiffels and Durbin, 2014; Paul, Steinrücken, and Song, 2011; Steinrücken, Paul, and Song, 2013; Sheehan, Harris, and Song, 2013; Steinrücken, Kamm, and Song, 2015). Exploiting LD gives these meth-

ods more power to reconstruct past demographic events. At the same time, it entails an added level of mathematical and computational sophistication, such that state-of-the-art methods are limited to analyzing whole genome samples from only a few genomes at a time. Also, most of these methods require as input computationally phased haplotypes, entailing a costly and potentially error-prone preprocessing step. As we show later, switch errors in phasing can catastrophically distort the inferred history.

Here, we present a new inference framework called SMC++ that combines the computational efficiency of the SFS and the advantage of utilizing LD information in coalescent HMMs. Our method is designed to take advantage of modern data sets consisting of hundreds of *unphased* whole-genomes. It can also analyze pairs of diverged populations, allowing it to pool information from both populations, as well as directly estimate the time of divergence. We believe that this is the first demographic inference method capable of analyzing unphased whole genome data from a large number of individuals in a computationally efficient and stable manner, while taking linkage information into account.

2.2 Results

We first demonstrate the accuracy and efficiency of our method SMC++ on simulated data. Then we apply the method to analyze real data from several large-scale whole-genome sequencing projects for various species: over a thousand genomes from eight human populations in Africa and Eurasia, hundreds of genomes from a Drosophila population in Africa, and tens of genomes from zebra finch and long-tailed finch populations in Australia. Throughout the rest of the chapter, we denote the haploid sample size of a data set by n.

Accuracy and computational performance on simulated data

The exact details of our simulation procedure are described in Methods. Briefly, each scenario consisted of ten replicates of 3Gb of data simulated under either a "sawtooth" demography (Schiffels and Durbin, 2014) featuring repeated exponential expansions and crashes over the last 1 million years; or "recent expansion", a stylized model of European population growth involving a bottleneck 200 kya and tenfold expansion over the last 10 ky.

Effect of phasing error. Many inference procedures in population genetics require phased sequence data, which leads to an unavoidable and potentially significant source of computational expense and error. Since current phasing methods work best when large reference panels are available (S. R. Browning and B. L. Browning, 2011), these problems are especially acute when studying new populations for which a suitable reference panel is not available.

Phasing errors confound demographic inference by breaking up identity-by-state tracts in closely related haplotypes, biasing downwards the number of inferred coalescences in the recent past and causing haplotype-based inference methods to infer large recent effective population sizes. To quantify this effect, we simulated $n = 4$ genomes under the sawtooth demography to generate three input datasets for MSMC and SMC++. One version of the dataset

contained the exact haplotypes generated by the simulation. The other two contained artificially induced "switch errors" at rates of 1% and 5%, which are approximate upper and lower bounds on the accuracy of existing phasing algorithms (S. R. Browning and B. L. Browning, 2011; Delaneau, Zagury, and Marchini, 2013).

Figure 2.1 shows the result of running SMC++ and MSMC on these datasets. There are three lines corresponding to the performance of MSMC at each level of switch error. (Since it is invariant to phasing, only a single line is plotted for SMC++.) All of the fits exhibit some bias, but the methods show fairly good agreement and low dispersion in the distant past. In the recent past, large differences emerge in the performance of MSMC as error increases. With 1% switch error, MSMC has difficulty accurately inferring the demography more recently than 20 kya, with highly divergent estimates (off by four orders of magnitude) over the past 3 ky. With 5% phasing error performance further deteriorates, with estimates diverging from substantially from the truth beginning 20 kya. With no phasing error, the accuracy of MSMC is similar to that of SMC++, with SMC++ producing higher resolution in the recent past. While we have used MSMC for illustrative purposes, we expect any demographic inference procedure which relies on phased data to exhibit similar inaccuracies in the presence of phasing error.

SMC++ compared to other methods. Next, we compared SMC++ with MSMC and the SFS-based inference method $\partial a \partial i$ (Gutenkunst et al., 2009). Based on the findings in the preceding section, we introduced phasing errors into the data at a switch error rate of 1%. (Note that $\partial a \partial i$ is also invariant to phasing.)

For both the sawtooth and recent-expansion simulations, SMC++ estimates are much more tightly concentrated around the true demography (shown in black) than either of the other two methods. The spline-based regularization scheme (Methods) employed by SMC++ effectively trades a small amount of bias for greatly reduced variance. Estimates obtained from $\partial a \partial i$ appear to be biased only slightly, but vary significantly from run to run and fluctuate substantially across epochs. MSMC has difficulty accurately inferring the true history, with estimates diverging quite severely for the very recent past. We note that these findings are somewhat at odds with the simulation results reported by Schiffels and Durbin (Schiffels and Durbin, 2014), which appear to have been performed on error-free data.

The estimates in Figure 2.2a seem to deviate more from the truth than in Figure 2.2b; in general the sawtooth demography seems to be harder to accurately infer compared to the recent-expansion demography. This may in part be due to the somewhat pathological nature of the sawtooth demography, which experiences several rapid crashes over a relatively short period of time. It is known, for example, that a strong bottleneck complicates efforts to accurately estimate size change events which occurred before the bottleneck (Terhorst and Song, 2015). For the recent-expansion demography (Figure 2.2b), SMC++ tends to smooth over abrupt changes in the historical effective population size, a phenomenon which has been witnessed previously in related methods (Li and Durbin, 2011).

Estimating the rate of recombination. In the preceding simulations, we assumed that the recombination rate (denoted by $\rho/2$) was known ahead of time. When ρ is unknown, SMC++ can estimate this quantity simultaneously with the demography. To test this, we generated ad-

ditional data sets with $n = 50$ under the recent expansion demography, varying ρ within the range $[0.1\theta, 10\theta]$ with $\theta/2$ being the mutation rate. We then used SMC++ to jointly infer both the demography and the recombination rate. Results for the recombination rate estimation are shown in Figure 2.8. SMC++ is able to estimate the rate of recombination fairly accurately over two orders of magnitude. It is most accurate when the ratio of recombination rate to mutation rate is below one. As ρ approaches 10θ, the estimates of the ratio ρ/θ exhibit some downward bias, which causes the inferred demographies to become too large. In this regime recombinations are roughly as common as mutations along the genome, complicating efforts by the HMM to establish the marginal time to the most recent common ancestor (TMRCA). The resulting demographic estimates exhibited additional variation but were qualitatively similar to those obtained when the recombination rate was known (Figure 2.2b and Figure 2.8).

Inference of split times. SMC++ can analyze pairs of populations simultaneously to infer divergence times jointly with population size histories. The current version of our implementation assumes a "clean split" model, in which no gene flow occurs after the populations split (Extension to multiple populations). As in the one-population case, divergence time estimates and the jointly inferred demographies do not depend on phasing.

We verified by additional simulations that SMC++ can accurately determine the divergence time and demography in the clean split case. For these simulations we focused on somewhat smaller sample sizes of $n = 10$ lineages in each population. This illustrates the performance of our method when there are limited data; additionally, the computations needed to jointly analyze two populations have higher computational complexity, so it is expedient to fit the joint model to somewhat smaller data sets.

The two populations were simulated according to the "recent expansion" demography described above. At a time which varied from simulation to simulation, population 2 splits from population 1 (moving forward in time) and maintains a constant effective population size until the present, while population 1 continues to follow the "recent expansion" demography. No gene flow occurs between the two populations after the split.

Simulation results are shown in Figure 2.3. Over the range of approximately 6–120 kya, SMC++ is able to infer divergence times with low error. Additionally, the inferred demographies exhibit an acceptable level of accuracy. In situations where additional data are available, SMC++ can also estimate the demography of each population separately and then combine this information to infer the joint demography and divergence time.

Computational performance. We recorded the run time and peak memory consumption for SMC++, $\partial a \partial i$ and MSMC while analyzing the simulated data sets mentioned above. To allow for fair comparison, we restricted SMC++ and MSMC to six threads, though we note that SMC++ is capable of exploiting all cores when possible. Results are shown in Figure 2.4. At $n = 8$, the largest setting for which we were able to successfully run MSMC, SMC++ requires roughly an order of magnitude less memory and time. $\partial a \partial i$, which operates on the SFS rather than sequence data, required extremely little memory (roughly 100Mb per run) and ran in roughly the same amount of time as SMC++.

For larger sample sizes, memory consumption with SMC++ is approximately constant, and running time also scales quite favorably. It may seem odd that SMC++ running time actually decreases slightly as n grows. This is due to the fact that our method's computational performance improves as the density of segregating sites decreases (EM algorithm in Methods), which occurs when we "thin" the data more aggressively at larger sample sizes in order to break up correlations in the underlying ancestral recombination graph (see Thinning in Methods). All told, even though there are fewer segregating sites in the thinned data for large n, they are more demographically informative.

Improvement in posterior decoding. In addition to demographic inference, SMC++ can also be used to locally infer the time to the most recent common ancestor (TMRCA) of the distinguished lineages. We hypothesized that the CSFS, by establishing a link between the coalescence time of the distinguished pair and the allelic status of the rest of the sample, could help to obtain an improved posterior distribution on the former quantity. To check this, we simulated additional 10 Mb stretches of sequence data and compared true TMRCA at each position in the distinguished pair with the posterior obtained by running our method. To quantify changes in accuracy, we calculated the root mean-squared error between the inferred and true TMRCA. (See Posterior decoding in Methods for a precise definition.)

Table 2.1 shows the results of these simulations. For each scenario, the corresponding table entry gives the average RMSE relative to the PSMC ($n = 2$) baseline decoding. The table has three strata. The outermost indicates the result of varying the ratio of recombination rate to mutation rate. The next stratum shows the effects of introducing genotype error, and the innermost shows the effects of introducing missingness into the simulations. (See Posterior decoding for a description of how we simulated these errors.) A "*" in columns 4-6 indicates that the ratio of RMSEs is different from 1 at a 5% significance level.

In general, the table confirms that incorporating additional information from the CSFS improves the posterior decoding. For large values of the ρ/μ, the effect is on the order of 1-2% and is statistically significant in all scenarios. On the other hand, for very low values of recombination ($\rho/\mu = 0.1$) the results exhibited considerable variance, with certain combinations of simulation parameters leading to substantial improvements, and others resulting in no significant improvement. Manually comparing the posterior decoding and true TMRCA in these cases revealed long spans with very recent coalescence times, which did not accumulate enough mutations to make reliable inference on the TMRCA.

The middle section of the table examines the case when mutation and recombination occur at comparable rates, and is the most relevant for analyzing human data. For high levels of genotype error (0.1% of sequenced bases) we see a very substantial decrease in RMSE as the sample size increases. This could potentially be useful in low coverage sequencing applications. Lower genotype error rates saw improvements of 1-2% when there was 10% missingness in the data, and 5-6% with 20% missingness. Finally, we found that increasing the sample size from 5 to 10 and then to 25 generally decreased the error, but did not observe much of an improvement beyond that. We suspect that this is because the amount of correlation between the CSFS and the TMRCA of the distinguished pair declines marginally as additional samples are

added.

Inference of human demography

We used SMC++ to infer the historical effective size of eight human populations from Africa and Eurasia. We obtained whole-genome sequences from Complete Genomics (Drmanac et al., 2009) and the 1000 Genomes Project (1000 Genomes Project Consortium, 2010). The Complete Genomics data consist of 54 unrelated individuals obtained from various populations and have average coverage in excess of 45×. The 1000 Genomes data has a larger sample size (typically 50-100 individuals per population) but lower coverage data. We combined these datasets by treating the genomes of each individual from the Complete Genomics data as "distinguished" lineages and computing the full "conditioned SFS" using all available individuals (see the Section 2.5 for the details of these concepts). Within each population, we fit a model using composite likelihood, whereby SMC++ was instantiated once for each available pair of distinguished lineages, and the sum of their log-likelihoods maximized using EM. Additionally, we analyzed a single ancient individual ("Ust'-Ishim") who lived ∼45 kya in western Siberia (Q. Fu et al., 2014). The data used in this portion of the analysis are summarized in Tables 2.2 and 2.3.

Results across all populations are shown in Figure 2.5, with estimates broken out by continent in Figures 2.9–2.11. A constant generation time of 29 years (Langergraber et al., 2012) was used to convert each figure from the coalescent time to calendar time, though we caution that the generation time(s) of ancient human populations are not known precisely, potentially distorting x-coordinates of the plots in the distant past. The per-generation mutation rate was fixed at 1.25×10^{-8} per base pair for all extant populations and 1.45×10^{-8} for the Ust'-Ishim individual (Q. Fu et al., 2014).

Each plot shows the inferred size history over the period 1 Mya–1 kya in panel (a), as well as the size history for the past 20 kya on a linear scale in panel (b). Several interesting features emerge. In the period from 1 Mya to 300 kya, all population size histories experience similar dynamics. Around 300 kya, we begin to see the African size histories deviating from all other populations, potentially reflecting the existence of population structure within Africa. In the period from 300 kya to 100 kya, the African and non-African populations start to diverge significantly. During this period the non-African populations experience a steep decline in effective population size, while the African populations experience a more moderate decline. There is strong concordance in the population size histories of the European and Asian populations over this time interval, consistent with the notion that they split from a common ancestral population more recently than 100 kya.

Turning to the period more recent than 100 kya, all of the various populations become more distinct. Part of the observed short-term variability is due to the increased variance of our estimates as we approach the present (Figure 2.12), but differences in the global pattern should reflect real differences between population size histories. The African populations (Figure 2.9) all experience a relatively mild bottleneck followed by growth in the period 100–10 kya. The Luhya (LWK) and Maasai (MKK) reach a nadir around 70–80 kya, and the Yoruba (YRI)

somewhat later, around 50 kya. All three populations seem to experience growth from at least 50 kya until present, with a noticeable uptick starting approximately 15 kya. Note that additional variability and lower recent effective population size seen for Maasai is likely due in part to the much smaller sample size of this population (Table 2.2).

Among the Asian populations (Figure 2.10), a somewhat different pattern emerges. All three populations experience a sharp bottleneck reaching a nadir around 55 kya, followed by growth until the present. Sudden rapid growth is experienced by the Gujarati (GIH) population around 5–10 kya, whereas in the other two populations (CHB, JPT) growth begins earlier, around 15 kya. It is interesting to note that the Gujarati population appears distinct from the east Asian populations starting around 100 kya.

The European populations (Figure 2.11) experience a similar bottleneck as Asians ending around 50 kya, followed by a period of rapid growth starting 10–15 kya. The size histories of the Tuscan (TSI) and northern European (CEU) populations are nearly identical until around 5 kya.

The plots show that several of the populations in the sample experienced similar size histories before diverging at some point in the recent past. To study this further, we jointly estimated their size histories and split times using the two-population model described above. Results are shown in Figure 2.6. Our model estimates a divergence time of 13 kya for the European and Tuscan populations; 18 kya for the Han Chinese and Japanese; 47 kya for the European and Han; and 110 kya for the northern European and Yoruban. The joint estimates of population sizes shown in Figure 2.6 differ slightly from the marginal estimates shown in Figure 2.5, because the assumption of continuity between the two size histories at the time of divergence places additional constraint on the estimation problem. Overall, however, the estimates are quite consistent between the two figures.

Finally, we note the close agreement between the Ust'-Ishim individual (black line) and modern populations in Figure 2.5 over the period 45 kya–1 Mya. The highly disparate origins of these samples—the former consisting of ancient DNA, the latter obtained from present-day individuals—gives us confidence that the features emerging in this plot correspond to actual historical events, and not data or modeling artifacts.

Inference of demography in other species

To verify that SMC++ has applications beyond human demographic inference, we also inferred population size histories for two species of finch (the zebra finch, *Taeniopygia guttata*, and the long-tailed finch, *Poephila acuticauda*) from Australia (Singhal et al., 2015), as well as a wild African population of *D. melanogaster* from Zambia (Lack et al., 2015). These datasets are described in Table 2.3. For the finch, we assumed a generation time of 3 months and a per-generation mutation rate of 7×10^{-10} per base pair (Singhal et al., 2015). For *D. melanogaster*, we assumed a generation time of 1 month and a per-generation mutation rate of 3×10^{-9} per base pair (Keightley et al., 2014). Due to uncertainty surrounding the generation times of these species, we have also plotted these results in generations (Figure 2.13). For all three species we estimated the per-generation recombination rate from data. The genome-wide average ratios

of recombination to mutation were estimated to be 1.6 : 1 (*D. melanogaster*), 1.15 : 1 (*P. acuticauda*) and 1.1 : 1 (*T. guttata*).

Results of demographic inference are shown in Figure 2.7. The two finch species have similar, stable size histories between 500 kya and 1 Mya. Starting around 500kya, the populations experienced a decline which persisted until about 60–80kya, depending on species. This was followed by a period of expansion, which accelerated around 15 kya and led to more than tenfold increase in effective population size by 1 kya.

Estimates for *D. melanogaster* are overall lower, though we caution that the short generation time of the Drosophila, combined with smaller genome size, complicate efforts to peer this far back into its past using genetic data. *D. melanogaster* appears to decline from 600–100 kya, at which point it experiences steady growth leading to present. Unlike the other species analyzed in this chapter, we do not see as much evidence of a sudden increase in effective population size in the recent past.

2.3 Discussion

We have presented SMC++, a new demographic inference method capable of analyzing hundreds of unphased whole-genome sequences at a time, while being fast, robust and easy to use. The ability to analyze much larger sample sizes makes our method's estimates substantially more accurate than previous methods, especially in the recent past. On simulated data, we obtain accurate estimates of the true effective population size history across a time span of three orders of magnitude. Furthermore, our method is able to estimate population split times with low error over a wide range of timescale. In real data we obtain convincing estimates of the effective population size history of a number of different populations. In most cases our estimates agree with previous findings concerning divergence times and historical migration patterns, while also bringing to light some intriguing new features which could merit further study.

Two aspects of our method in particular are worth re-emphasizing. We have shown using simulations that introducing a modest amount of phasing error can severely corrupt the estimates of an existing demographic inference method. We conjecture that any method which is not invariant to phasing suffers from similar issues. Hence, the phase invariance of our method makes it more robust. Additionally, it eliminates a burdensome preprocessing step when analyzing real data.

We have also demonstrated that SMC++ requires an order of magnitude less memory and processing time than existing methods. We have found in practice that this is extremely useful for exploring and testing hypotheses in real data. The results of any demographic analysis depend on a number of *a priori* modeling assumptions, such as the functional form of the demography and various tuning parameters. At present we lack a theoretical understanding of how to optimally choose these parameters. This is an important area for future research, and in the meantime the ability to explore the model space and receive rapid feedback from the algorithm is essential.

Our general theoretical framework, which couples the genealogical process for a given diploid individual with the allele frequency information in a large collection of other individuals, can be extended to more complex demographic models. In particular, we plan to extend our method to incorporate gene flow between populations. We believe that this approach opens up a new window of opportunity to utilize the information contained in a large collection of whole-genomes to infer population demographic history in finer detail and higher accuracy than previously possible.

2.4 Methods

SMC++ stands for "**S**equential **M**arkov **C**oalescent **+ Pl**enty of **U**nlabeled **S**amples." SMC++ unites the PRF and coalescent HMM approaches, combining the strengths of each while overcoming several of their limitations. The inclusion of "unlabeled samples" into the standard coalescent HMM is achieved via novel theoretical results on what we term "conditioned SFS" (CSFS), the sample frequency spectrum conditioned on the coalescence time and allelic state of a distinguished diploid individual. Compared to existing methods, the main advantages of SMC++ are:

1. Scalability: SMC++ can analyze hundreds of individuals at a time while requiring a modest amount of memory and processing time. Analyzing a hundred human genomes takes roughly an hour on a laptop.

2. Accuracy: By accommodating larger sample sizes, SMC++ has significantly improved power to infer demographic events, particularly in the recent past.

3. Phase invariance: SMC++ only requires unphased sequence data as input (i.e., results do not depend on phasing).

4. Regularity: SMC++ uses cubic splines to enforce a smoothness constraint on the inferred demographies. Compared with existing methods, the resulting estimates exhibit far less variance, with only a minimal increase in bias.

SMC++ extends a line of work which approximates the intractable sampling distribution of a contiguous segment of DNA (that of the so-called *ancestral recombination graph*, (R. C. Griffiths and P. Marjoram, 1997)) by a tractable Markov process. We first give a brief overview of these related methods, followed by a description of the novel aspects of SMC++.

Related work

The starting point for our model is the well known pairwise sequentially Markov coalescent (PSMC) Li and Durbin, 2011. That paper applied a simple likelihood model for the pattern of genetic mutations observed in a single diploid individual, and showed that it had surprising power to infer the history of the individual's population. PSMC works as follows: each pair

of chromosomes is divided into blocks of 100 base pairs, and then mapped to a binary string according to whether each block contained one or more heterozygous sites. These "emitted" symbols are modeled as an HMM whose hidden state is the marginal TMRCA, assumed constant within each block. Conditional on the TMRCA being equal to T, the probability of observing a heterozygous block is $1 - e^{-100\theta T}$ where $\theta/2$ is the per-base mutation rate. The distribution of the TMRCA is discretized into a number of disjoint intervals, and its stationary distribution and (discrete) transition function are then computed with respect to a piecewise-constant population size history. The transition function, which describes the conditional distribution of TMRCA between neighboring sites, is intractable in its exact form (Wiuf and Hein, 1999). Instead, a simple Markov approximation is used (McVean and Cardin, 2005). The stationary, emission, and transition probabilities together define a map from population size history to sampling distribution via the HMM. The likelihood function is then maximized iteratively using the EM algorithm.

Despite its apparent simplicity, PSMC has proved very effective in practice and has been utilized in hundreds of subsequent analyses of many different species. Nevertheless, there are a few clear areas where PSMC could be improved:

- The so-called SMC approximation (McVean and Cardin, 2005) utilized to calculate transition probabilities in PSMC is inaccurate. A slight modification known as SMC' (P. Marjoram and Wall, 2006) results in a much more accurate approximation of the transition matrix, with negligible added computational burden (Hobolth and Jensen, 2014; Wilton, Carmi, and Hobolth, 2015).

- The computational performance of PSMC, though adequate, can be substantially improved via parallelism on modern, many-core workstations. Doing so renders the blocking approximation unnecessary.

- More fundamentally, PSMC is limited to analyzing a single diploid individual. This strongly limits the power of PSMC, particularly in the recent past where there will be few coalescent events in a sample of size two.

A follow-on method called MSMC (Schiffels and Durbin, 2014) extends PSMC to larger sample sizes, and also allows for the analysis of multiple populations. MSMC does so by redefining the HMM hidden variable to be the first coalescent event among the n haplotypes in the sample, with transition probabilities based on SMC'. For the emission probabilities, MSMC mainly uses the singleton mutations, only using the higher-frequency mutations to disallow certain coalescence events (if the 2 coalesced haplotypes have different alleles the emission probability of the non-singleton mutation is 0, otherwise the mutation is ignored). For the case of $n = 2$, MSMC essentially reduces to PSMC. Since the time to first coalescence depends more strongly on recent demography for larger sample sizes, MSMC has greater power to infer size changes in the recent past. On the other hand, the computational burden of MSMC also grows substantially with sample size, such that it is limited, in practice, to analyzing no more than four diploid whole genome samples.

A different approach to performing demographic inference via a coalescent HMM may be found in diCal (Sheehan, Harris, and Song, 2013) and related methods (Tataru, Nirody, and Song, 2014; Steinrücken, Kamm, and Song, 2015). These rely on the so-called *conditional sampling distribution* (CSD), which characterizes the probability of observing an k-th sampled chromosome conditional on the previously observed $k-1$ samples. Computing the CSD becomes substantially more difficult for larger sample sizes, so that even the most recent version of diCal is limited to analyzing no more than 10 samples at a time at whole-genome scale. On the other hand, diCal is capable of analyzing complex demographic scenarios involving size changes, admixture, migration and population splits.

Theoretical model for SMC++

Above, we described how MSMC generalizes PSMC in two different directions, by altering both the hidden state space as well as the space of emitted symbols used in the HMM. We also demonstrated in simulation that MSMC has difficulty scaling to large n. This is partly due to the fact that there are $O(n^2)$ pairs of haplotypes that may be involved in the first coalescence; thus, the HMM must integrate over $O(n^2)$ hidden states per time interval per locus.

To scale to larger sample sizes, SMC++ generalizes PSMC in a different direction. The hidden state of SMC++ is exactly the same as in PSMC; specifically, the hidden state is the TMRCA between the 2 haplotypes of a single individual. Where SMC++ and PSMC differ are in the observed state. Whereas PSMC emits a binary symbol indicating whether the individual is heterozygous, SMC++ additionally emits the allele frequency of an extra $n-2$ haplotypes. Specifically, the emission probability of SMC++ is based on the site frequency spectrum (R. C. Griffiths and Tavaré, 1994; Polanski and Kimmel, 2003), *conditioned* on the TMRCA of a single "distinguished" individual.

The differences between these three approaches are depicted in Figure 2.14. PSMC and SMC++ track the same hidden information, but their emissions are different. Whereas the PSMC emissions are binary symbols, SMC++ emissions are 2-tuples $(a, b) \in \{0, 1, 2\} \times \{0, 1, \dots, n-2\}$. Here a and b denotes the number of derived alleles present in the "distinguished" individual and the additional "undistinguished" $n-2$ haplotypes, respectively. We refer to the distribution of these tuples as the *conditioned SFS* (CSFS): the site frequency spectrum conditional on the event that the TMRCA of the distinguished member falls in a given interval. The CSFS is a two-dimensional generalization of the site frequency spectrum; indeed, summing the CSFS across rows exactly recovers the standard site frequency spectrum on n haploids. We explicitly define the SMC++ emission probability, and rigorously derive the CSFS using coalescent theory, in the Section 2.5.

In order to calculate the CSFS we assume that the mutation rate is known. Uncertainty in the mutation rate can potentially alter the estimates of SMC++, though in practice this effect will be attenuated by also using linkage information to infer the demography.

Thinning

An important caveat to our approach concerns correlation between observations of the conditioned SFS. The hidden Markov model formally assumes that neighboring observations are independent conditional on knowing the underlying hidden state. If we just use the distinguished individual, without the additional undistinguished haplotypes, this assumption is correct (up to the SMC' approximation and time discretization). However, the conditional independence assumption is violated when we add the additional undistinguished lineages: correlations in branch lengths of the sample's underlying genealogy will remain even after conditioning on the TMRCA within the distinguished individual. Emitting the "full" conditioned SFS at every site therefore leads to model misspecification. To prevent this we adopt a "thinning" strategy, in which the full conditioned SFS is emitted only at every k-th site for some pre-specified constant k.

Here, we adopted the heuristic $k = Cn$ for a large constant C, so that the density of full SFS emissions decreases linearly in sample size. We found that $C = 400$ worked well in practice and all analyses reported are the result using that choice of parameter.

Transition function

As discussed earlier, the SMC' transition function more accurately approximates the sequential coalescent. In SMC++ we employ a continuous time, conditioned Markov chain approximation to the sequential coalescent (Hobolth and Jensen, 2014). This approximation, which derives from the exact continuous time description of the two-locus coalescent with recombination (Simonsen and Churchill, 1997), is actually slightly more accurate than SMC' as originally formulated (P. Marjoram and Wall, 2006) since it integrates over any number of recombinations and back-coalescences occurring between neighboring sites.

EM algorithm

For a hidden Markov model with M hidden states and L observations, the complexity of the forward-backward algorithm is $O(M^2L)$. Genetic data usually contains long stretches of non-polymorphic sites, a fact which can be used to decrease the running time of this algorithm to $O(M^2L_p)$, where p is the number of polymorphic loci in the data set (Paul and Song, 2012).

However, extending this speedup to the EM algorithm is nontrivial: in particular, it was previously unknown how to compute the posterior expected number of transitions and emissions from each hidden state, which are needed to execute the "M" step of the EM algorithm during model fitting. In Section 2.5, we present new results which compute these posterior expectations while taking advantage of the sparsity of polymorphic sites. The complexity of our modified algorithm is $O(M^3L_p)$ which improves the running time $O(M^2L)$ of the non-sparse algorithm when $ML_p < L$. For large sample sizes $L = O(10^2L_p)$ so this requires that $M = O(10)$, a setting which we found produces acceptable results in practice.

Regularization

As mentioned above, we found that regularization dramatically improved convergence of our algorithm and the quality of the resulting estimates. SMC++ enforces a mild smoothness constraint by fitting a cubic spline to the data. Hence, the space of models considered by our algorithm consists of continuously twice-differentiable (C^2) curves. (A weaker constraint requiring only C^1 smoothness is also available to the user.)

In addition to this "implicit" regularization scheme, we added an explicit "roughness" penalty to the optimization. Here roughness is defined as

$$\mathcal{P}(f) \stackrel{\text{def}}{=} \int_R (f'')^2,$$

where R is the (compact) support of the spline f. The penalized likelihood used for fitting is then $\mathcal{Q}(f) - \lambda\mathcal{P}(f)$, where $\mathcal{Q}(f)$ is the so-called complete log-likelihood used in the EM algorithm (Bishop, 2006), and $\lambda > 0$ is the regularization parameter.

By penalizing curvature, this type of regularizer shrinks to a linear fit, and encodes our prior belief that real populations are unlikely to experience repeated crashes and expansions over short time intervals. For our simulated results, we found that $\lambda = 0.01$ worked well. On real data, which contains noise, missingness, and other model violations, we found that slightly larger values, ranging from $\lambda = 1.0$ to $\lambda = 10.0$ depending on the amount and quality of available data, were necessary to obtain smooth fits.

Simulation procedure

We used the coalescent simulation program `scrm` (Staab et al., 2015) to generate simulated data sets. Each simulated data set consisted of 30 independently generated chromosomes of length 100Mb. Ten replications were performed for each combination of demography, inference procedure, and sample size. When running MSMC we fixed the value of the recombination rate to its true value, while for SMC++ the recombination rate was estimated from the simulated data.

The command line used to simulate from the sawtooth demography was:

```
scrm <n> 1 -p 10 -t 71560.0
-r 17889.9998211 100000000
-oSFS -seeds <s1> <s2> <s3> -eN 0.0 5.0
-eG 0.0 0.0 -eN 0.000582262 5.0
-eG 0.000582262 1318.18 -eN 0.00232905 0.500002043581
-eG 0.00232905 -329.546 -eN 0.00931919 5.00496081234
-eG 0.00931919 82.3865 -eN 0.0372648 0.500618264601
-eG 0.0372648 -20.5966 -eN 0.149059 5.0061592508
-eG 0.149059 5.14916 -eN 0.596236 0.500615510407
-eG 0.596236 0.0
```

The command line used to simulate from the human demography was:

```
scrm <n> 1 -p 10 -t 50000.0
-r 12499.999875 100000000
-oSFS -seeds <s1> <s2> <s3> -eN 0.0 10.0
-eG 0.0 230.258509299 -eN 0.01 0.5 -eG 0.01 0.0
-eN 0.07 1.0 -eG 0.07 0.0 -eN 0.2 4.0 -eG 0.2 0.0
```

In these commands `<n>` is the (haploid) sample size, which varied from experiment to experiment, and `<s{1,2,3}>` are random number generator seeds, which varied from experiment to experiment but were retained to enable reproducibility.

To model the effect of computational phasing we introduced switch error into the data sets using the following procedure. Let $\mathbf{X} \in \{0,1\}^{2\times S}$ be a binary matrix representing a pair of dimorphic haplotypes at S segregating sites. The ith column of $\mathbf{X}$ is denoted $\mathbf{X}_i = (X_{i,1}, X_{i,2})$. Also let $U_i, i = 1, \ldots, S$ denote a sequence of i.i.d. $Uniform(0,1)$ random variables and let α denote the switch error rate. Execute the following random algorithm:

1. set `switch_error` = False
2. For $i = 1, \ldots, S$
   a) If $U_i < \alpha$ then `switch_error` $= \neg$ `switch_error`
   b) If `switch_error` = True, then $Y_{i,1} = X_{i,2}$ and $Y_{i,2} = X_{i,1}$; else $\mathbf{Y}_i = \mathbf{X}_i$.
3. Return $\mathbf{Y} \in \{0,1\}^{2\times S}$.

Posterior decoding

The root mean squared error between the true TMRCA and inferred posterior is defined as

$$\text{RMSE} = \left[\sum_{\ell=1}^{L} \sum_{m=1}^{M} \gamma_{\ell m} (h_m - t_\ell)^2 \right]^{1/2}.$$

Here, $\gamma_{\ell m}$ is the posterior probability of coalescence in the m-th hidden state for position ℓ, t_ℓ is the true TMRCA at position ℓ, and h_m is the expected coalescence time conditional on coalescence occurring in hidden state m.

To simulate genotype error, we "flipped" a Poisson-distributed number of randomly chosen bases in each simulation. To simulate missingness, we replaced a Poisson-distributed number of segregating sites in each sample with missing values. The means of the Poisson distributions were chosen such that in expectation, the fraction of bases affected by each type of error equaled the corresponding value in columns 2–3 of Table 2.1.

Code Availability

Our software, including source code and a Python-based command line interface, is publicly available at the PopGenMethods repository.

Data Availability

All data analyzed in this study are publicly available. Whole genome sequencing data was obtained from the Complete Genomics diversity panel. Frequency spectrum data was obtained from the 1000 Genomes Phase 3 release. Sequence data for the two species of finch were obtained from the European Nucleotide Archive, accession number PRJEB10586. Drosophila data was obtained from the Drosophila Genome Nexus. Scripts used to generate the simulated data are available in the code repository given above.

2.5 Emission Probabilities

In this section we present results and proofs needed to compute the emission probabilities described in the main text. To simplify the equations, we depart slightly from the notation in the main text and now assume that n "undistinguished" + 2 "distinguished" haploids are sampled from a single panmictic population for an overall sample size of $n + 2$ haploids. The two distinguished lineages are labeled 1 and 2. The effective size of this population at t coalescent units in the past was $1/\alpha(t)$, and we define the cumulative coalescence intensity function $R(t) := \int_0^t \alpha(s)\,ds$. The random tree describing the ancestry of these samples will be referred to as simply "the coalescent" throughout this section, despite the inhomogeneous rate function. Following standard terminology, we say that a branch of this tree has *size* k if it is ancestral to k sampled individuals at the present. The samples are genotyped at a single biallelic site, and the allelic state (number of derived alleles) at this site is denoted $(a, b) \in \{0, 1, 2\} \times \{0, 1, \dots, n\}$. Finally, we set $a_m \overset{\text{def}}{=} \binom{m}{2}$ throughout.

Let C_{12} denote the (random) time at which 1 and 2 coalesce. Given $C_{12} = \tau$, the *conditioned site frequency spectrum* $\text{CSFS}(\tau)$ is a $3 \times n$ matrix whose (a, b)-th entry is the expectation, conditional on $C_{12} = \tau$, for the total branch length subtending a distinguished and b undistinguished lineages. Under neutrality, mutations occur as a rate-$\frac{\theta}{2}$ point process with respect to branch length, and so $\text{CSFS}(\tau)$ is also the leading-order coefficient in the Taylor expansion for the conditional probability of observing (a, b):

$$\mathbb{P}((a, b) \mid C_{12} = \tau) = \frac{\theta}{2} \cdot [\text{CSFS}(\tau)]_{ab} + O(\theta^2),\ 0 < a + b < n + 2.$$

Since the per-site mutation rate is small, i.e. $\frac{\theta}{2} \ll 1$, SMC++ approximates the emission probability by

$$\mathbb{P}((a, b) \mid C_{12} = \tau) \approx \frac{\theta}{2} \cdot [\text{CSFS}(\tau)]_{ab}$$

which is essentially equivalent to assuming at most 1 mutation per site.

To simplify the derivation, we decompose $\text{CSFS}(\tau)$ into the branch lengths above and below τ and derive each quantity separately:

$$\text{CSFS}(\tau) = \underbrace{\text{CSFS}(\tau\downarrow)}_{\text{below } \tau} + \underbrace{\text{CSFS}(\tau\uparrow)}_{\text{above } \tau}.$$

The strategy in both cases is similar: we first compute the total expected branch length in the coalescent on $n+2$ lineages, and then rely on combinatorial arguments to calculate the probability that a branch of size $2 \leq k < n+2$ subtends a sample of (a, b) distinguished and undistinguished lineages.

The conditioned coalescent

We start with an explicit generative model for the conditioned coalescent given C_{12}, which will be useful for computing both CSFS($\tau\downarrow$) and CSFS($\tau\uparrow$). The conditioned coalescent given $\{C_{12} = \tau\}$, denoted as $\mathcal{T}_{n+2}$, can be constructed via a sequence of subtrees $\mathcal{T}_2, \mathcal{T}_3, \ldots, \mathcal{T}_{n+2}$, where $\mathcal{T}_k$ is the genealogy relating the first k leaves $\{1, 2, \ldots, k\}$. $\mathcal{T}_k$ is obtained by recursively adding the kth leaf to the subtree $\mathcal{T}_{k-1}$. Specifically:

1. Initialize $\mathcal{T}_2$ as a cherry (tree with two leaves) with fixed height $A_2 = \tau$.
2. For $k = 3, \ldots, n+2$:
   a) Let $\sigma : \{2, \ldots, k-1\} \to \{2, \ldots, k-1\}$ be the permutation which sorts $A_2, \ldots, A_{k-1}$, in descending order, $A_{\sigma(2)} \geq A_{\sigma(3)} \geq \cdots \geq A_{\sigma(k-1)}$, and set $U_j = A_{\sigma(j)}$. Additionally define $U_1 = \infty$ and $U_k = 0$.
   b) Sample A_k from an inhomogeneous point process with piecewise intensity function
   $$\mu(t) := j \times \alpha(t), \quad U_{j+1} \leq t < U_j.$$
   c) Select a branch uniformly at random in $\mathcal{T}_{k-1}$ from among those surviving to time A_k, and form $\mathcal{T}_k$ by adding a new external branch (labeled k) to $\mathcal{T}_{k-1}$ which coalesces with the chosen branch at A_k.

To see why $\mathcal{T}_{n+2}$ is distributed as the coalescent conditional on $\{C_{12} = \tau\}$, suppose we replace step 1, and instead sample $\mathcal{T}_2$ as a random cherry whose height is given by an exponential waiting time with inhomogeneous rate $\alpha(t)$. Then it is straightforward to verify that at time t, every pair of lineages is coalescing at rate $\alpha(t)$, i.e. the resulting tree is distributed as the (unconditioned) coalescent.

We note that $\mathcal{T}_{n+2}$ can also be generated backwards in time, as follows. For $t < \tau$, the ancestors of the distinguished pair cannot coalesce with each other, but every other pair coalesces at rate $\alpha(t)$. So the total coalescence rate if there are k lineages is $(a_k - 1)\alpha(t)$. For $t > \tau$, every pair of lineages coalesces at rate $\alpha(t)$, and the total coalescence rate for k lineages is $a_k\alpha(t)$.

Computing CSFS($\tau\downarrow$)

In this section we compute the distribution of allelic configurations for a mutation occurring below τ, the time at which the lineages 1 and 2 are conditioned to coalesce. We first fix some notation. We denote by $\xi_{m,k}$ be the partition structure of $\mathcal{T}_m$ when it has $2 \leq k \leq m$ lineages

remaining: $|\xi_{m,k}| = k$. Similarly, $T_{m,k}$ denotes the amount of time in $\mathcal{T}_m$ during which k lineages remain. For $\tau \geq 0$ let $\mathcal{A}_m(\tau)$ denote the "ancestral process", i.e. the pure-death process which gives the number of lineages remaining in $\mathcal{T}_m$ at time τ.

Throughout this section we make repeated use of the event $E_{m,k} \stackrel{\text{def}}{=} \{\mathcal{A}_m(\tau-) \leq k\}$. The truncated times

$$T^{\tau\downarrow}_{m,k} \stackrel{\text{def}}{=} T_{m,k}\mathbf{1}_{E_{m,k}} \tag{2.1}$$

record the duration, until 1 and 2 coalesce, that $\mathcal{T}_m$ has k lineages remaining.

Finally, we will need the following important fact concerning the $T_{m,k}$ and $\xi_{m,k,}$, which follows directly from the generative algorithm given above.

Lemma 2.1. *For $k \leq j \leq m$, the random variables $T_{m,j}$ and $\xi_{m,j}$ are independent conditional on $E_{m,k}$.*

Branch lengths

$\mathbb{E}[T^{\tau\downarrow}_{m,k}]$ can be recursively computed by the following theorem:

Theorem 2.2. $\mathbb{E}[T^{\tau\downarrow}_{m,k}]$ *satisfies the recursion*

$$\left[1 - \tfrac{(k+1)(k-2)}{(m+1)(m-2)}\right]\mathbb{E}T^{\tau\downarrow}_{m,k} = \mathbb{E}T^{\tau\downarrow}_{m-1,k} - \tfrac{(k+2)(k-1)}{(m+1)(m-2)}\mathbb{E}T^{\tau\downarrow}_{m,k+1}$$

with base case $\mathbb{E}T^{\tau\downarrow}_{k,k} = \int_0^\tau e^{-(a_k-1)R(t)}\,dt$.

Proof. Generate $\mathcal{T}_2, \ldots, \mathcal{T}_m$ as above. Then

$$T^{\tau\downarrow}_{m-1,k} = T^{\tau\downarrow}_{m,k}\mathbf{1}\{\{m\} \notin \xi_{m,k}\} + T^{\tau\downarrow}_{m,k+1}\mathbf{1}\{\{m\} \in \xi_{m,k+1}\} \tag{2.2}$$

because if $\mathcal{T}_m$ has j lineages at time t, then $\mathcal{T}_{m-1}$ has $j - \mathbf{1}\{\{m\} \in \xi_{m,j}\}$ lineages at that time. Using (2.1), we can write the first term in (2.2) as

$$T^{\tau\downarrow}_{m,k}\mathbf{1}\{\{m\} \notin \xi_{m,k}\} = T_{m,k}\mathbf{1}_{E_{m,k}}\mathbf{1}\{\{m\} \notin \xi_{m,k}\} \tag{2.3}$$

Taking expectation in (2.3) and using Lemma 2.1, we obtain

$$\begin{aligned}\mathbb{E}(T^{\tau\downarrow}_{m,k}\mathbf{1}\{\{m\} \notin \xi_{m,k}\}) &= \mathbb{P}(E_{m,k})\mathbb{E}(T_{m,k} \mid E_{m,k})\mathbb{P}(\{m\} \notin \xi_{m,k} \mid E_{m,k}) \\ &= \mathbb{E}(T^{\tau\downarrow}_{m,k})\mathbb{P}(\{m\} \notin \xi_{m,k} \mid E_{m,k})\end{aligned} \tag{2.4}$$

Analogous manipulations of the second term in (2.2) yield

$$\mathbb{E}T^{\tau\downarrow}_{m-1,k} = \mathbb{E}T^{\tau\downarrow}_{m,k}\mathbb{P}(\{m\} \notin \xi_{m,k} \mid E_{m,k}) + \mathbb{E}T^{\tau\downarrow}_{m,k+1}\mathbb{P}(\{m\} \notin \xi_{m,k+1} \mid E_{m,k+1}). \tag{2.5}$$

Now, we have

$$\begin{aligned}\mathbb{P}(\{m\} \in \xi_{m,k} \mid E_{m,k}) &= \Big(1 - \frac{m-1}{a_m - 1}\Big) \cdots \Big(1 - \frac{k+1-1}{a_{k+1} - 1}\Big) \\ &= \prod_{j=k+1}^{m} \frac{(a_j - 1) - (j-1)}{a_j - 1} \\ &= \frac{(k+1)(k-2)}{(m+1)(m-2)}. \end{aligned} \tag{2.6}$$

Combining (2.5) and (2.6) yields the first part of the claim. The base case follows from direct computation:

$$\begin{aligned}\mathbb{E}T_{k,k}^{\tau\downarrow} &= \tau\mathbb{P}(T_{k,k} > \tau) + \mathbb{E}T_{k,k}\mathbf{1}\{T_{k,k} < \tau\} \\ &= \tau\mathbb{P}(T_{k,k} > \tau) + \int_0^\tau t(a_k - 1)\alpha(t)e^{-(a_k-1)R(t)}\,dt \\ &= \int_0^\tau e^{-(a_k-1)R(t)}\,dt,\end{aligned}$$

where we integrated by parts in the final equality. □

Block Structure

Next we compute the distribution of partition block structures conditional on 1 and 2 not having coalesced. Following the notation of the preceding subsection, let $\xi_{m,k} = \{L_1, \dots, L_k\}$ be the coalescent partition of $[m] := \{1, \dots, m\}$ when $\mathcal{T}_m$ has k lineages. As partitions are unordered, we are free to adopt a labeling, so let L_i contain the smallest numbered leaf which is not in $L_1 \cup \cdots \cup L_{i-1}$. Conditioned on $\{C_{12} = \tau\}$, this is equivalent to $1 \in L_1, 2 \in L_2$ beneath τ. The following results characterizes the block structure of $\mathcal{T}_m$ before 1 and 2 coalesce. In what follows, we use the conditioning notation $\mathbb{P}(\cdot \mid k)$ to signify that they pertain to the distribution on partitions in the coalescent while there are k blocks remaining.

Theorem 2.3. *The distribution of the block sizes $|L_1|, \dots, |L_k|$ in $\xi_{m,k}$, conditional on $\{1 \in L_1, 2 \in L_2, E_{m,k}\}$, is given by*

$$\mathbb{P}(|L_1|, \dots, |L_k| \mid 1 \in L_1, 2 \in L_2, k, E_{m,k}) = \frac{|L_1||L_2|}{\binom{m+1}{k+1}}.$$

Proof. Write

$$\mathbb{P}(|L_1|, \dots, |L_k| \mid 1 \in L_1, 2 \in L_2, k, E_{m,k}) = \frac{\mathbb{P}(|L_1|, \dots, |L_k| \mid k, E_{m,k})}{\mathbb{P}(1 \in L_1, 2 \in L_2 \mid k, E_{m,k})}\mathbf{1}\{1 \in L_1, 2 \in L_2\},$$

where in the denominator we are marginalizing over all partitions of $[m]$ into k blocks. A computation resembling equation (2.6) in the proof of Theorem 2.2 shows that

$$\mathbb{P}(1 \in L_1, 2 \in L_2 \mid k, E_{m,k}) = \prod_{i=k+1}^{m} \frac{a_i - 2}{a_i - 1} = \frac{(m+1)(k-1)}{(m-1)(k+1)}.$$

Now, the random variables $\xi_{m,k}$ are *not* in general distributed according to the coalescent due to the conditioning of $\mathcal{T}_k$ on $\{C_{12} = \tau\}$. Nevertheless, by the remarks preceding Lemma 2.1, beneath τ they are equal in distribution to the classical coalescent. Therefore, we may apply a result of Kingman (1982c) to conclude that

$$\mathbb{P}(\xi_{m,k} = \{L_1, \ldots, L_k\} \mid k, E_{m,k}) = \prod_{i=1}^{k} |L_i|! \times \frac{k!\,(k-1)!\,(m-k)!}{m!\,(m-1)!} \tag{2.7}$$

Let

$$\mathcal{S}_\xi = \{\zeta : \zeta_1 \uplus \cdots \uplus \zeta_k = [m], |\zeta_i| = |L_i| \ \forall i, 1 \in \zeta_1, 2 \in \zeta_2\} \tag{2.8}$$

be the set of partitions which have the same block sizes as $\xi_{m,k}$ and 1 and 2 in different blocks. There are

$$|\mathcal{S}_\xi| = \frac{1}{(k-2)!}\binom{m-2}{|L_1|-1, |L_2|-1, |L_3|, \ldots, |L_k|} \tag{2.9}$$

such partitions, all of which have the probability given in equation (2.7). Hence,

$$\begin{aligned}
\mathbb{P}(|L_1|, \ldots, |L_k| \mid 1 \in L_1, 2 \in L_2, k, E_{m,k}) &= \frac{1}{\mathbb{P}(1 \in L_1, 2 \in L_2 \mid k, E_{m,k})} \sum_{\zeta \in \mathcal{S}_\xi} \mathbb{P}(\zeta \mid k, E_{m,k}) \\
&= \frac{\mathbb{P}(\xi_{m,k} \mid k, E_{m,k})}{\mathbb{P}(1 \in L_1, 2 \in L_2 \mid k, E_{m,k})} |\mathcal{S}_\xi| \\
&= |L_1| \times |L_2| \times \frac{k-1}{m-1} \times \frac{1}{\binom{m}{k}} \times \frac{(n-1)(k+1)}{(m+1)(k-1)} \\
&= \frac{|L_1||L_2|}{\binom{m+1}{k+1}}.
\end{aligned}$$

□

Theorem 2.4. *Under the conditions of Theorem 2.3,*

$$\mathbb{P}(|L_i| = \ell_i \mid 1 \in L_1, 2 \in L_2, k+2, E_{m,k+2}) = \frac{\ell_i}{\binom{n+3}{k+3}}\binom{n+2-\ell_i}{k+1},$$

for $i = 1, 2$, while for $3 \le i \le k+2$,

$$\mathbb{P}(|L_i| = \ell_i \mid 1 \in L_1, 2 \in L_2, k+2, E_{m,k+2}) = \frac{1}{\binom{n+3}{k+3}}\binom{n+3-\ell_i}{k+2}.$$

Proof. For the first result, if $k = 0$ then only L_1 and L_2 remain, and the claim follows immediately. Otherwise, using Theorem 2.3,

$$\mathbb{P}(|L_1| = \ell_1 \mid 1 \in L_1, 2 \in L_2, k+2, E_{m,k+2}) = \frac{\ell_1}{\binom{n+3}{k+3}} \sum_{\ell_2=1}^{n+2-\ell_1-k} \ell_2 \binom{n-\ell_1-\ell_2+1}{k-1},$$

where the binomial term counts the number of k-way compositions of the integer $n+2-\ell_1-\ell_2$. The sum simplifies to

$$\sum_{\ell_2=1}^{n+2-\ell_1-k} \ell_2 \binom{n-\ell_1-\ell_2+1}{k-1} = \sum_{\ell_2=0}^{n-\ell_1} \binom{\ell_2+1}{1}\binom{n-\ell_1-\ell_2}{k-1} = \binom{n+2-\ell_1}{k+1},$$

where the second equality is from Graham, Knuth, and Patashnik (1994, eq. 5.26). By symmetry,

$$\mathbb{P}(|L_2| = \ell \mid 1 \in L_1, 2 \in L_2, k+2, E_{m,k+2}) = \mathbb{P}(|L_1| = \ell \mid 1 \in L_1, 2 \in L_2, k+2, E_{m,k+2}). \quad (2.10)$$

For the second result, using the same combinatorial identity twice yields

$$\begin{aligned}
\mathbb{P}(|L_i| = \ell_i \mid 1 \in L_1, 2 \in L_2, k+2, E_{m,k+2}) &= \frac{1}{\binom{n+3}{k+3}} \sum_{\ell_1=1}^{n+2-\ell_i-k} \ell_1 \sum_{\ell_2=1}^{n+3-\ell_i-\ell_1-k} \ell_2 \binom{n-\ell_1-\ell_2-\ell_i+1}{k-2} \\
&= \frac{1}{\binom{n+3}{k+3}} \sum_{\ell_1=1}^{n+2-\ell_i-k} \ell_1 \binom{n-\ell_1-\ell_i+2}{k} \\
&= \frac{1}{\binom{n+3}{k+3}} \binom{n+3-\ell_i}{k+2},
\end{aligned}$$

for all $3 \le i \le k+2$. □

Computing CSFS($\tau\downarrow$)

Using the preceding results we can compute CSFS($\tau\downarrow$). Let

$$\mathbf{T}^{\tau\downarrow} = \mathbb{E}(T^{\tau\downarrow}_{n+2,2}, \ldots, T^{\tau\downarrow}_{n+2,n+2})$$

be the row-vector of expected inter-coalescence times below τ and

$$\gamma(\tau) = \mathbb{E}(T^{\tau\downarrow}_{2,2}, \ldots, T^{\tau\downarrow}_{n+2,n+2})$$

be the row-vector of first-coalescence times, which can be computed using the base case in Theorem 2.2. By unwinding the recursion proved in that theorem, we obtain a matrix $\mathbf{B} \in \mathbb{Q}^{(n+1)\times(n+1)}$ such that $\mathbf{T}^{\tau\downarrow} = \gamma(\tau)\mathbf{B}$. Additionally, if $\mathbf{D} = \mathrm{diag}(2, 3, \ldots, n+2)$ then $\gamma(\tau)\mathbf{BD}$ is the total expected branch length in each inter-coalescence interval.

Next, let $\mathbf{P}_0 \in \mathbb{Q}^{(n+1)\times n}$ and $\mathbf{P}_1 \in \mathbb{Q}^{(n+1)\times(n+1)}$ be matrices whose entries are given by Theorem 2.4:

$$
\begin{aligned}
(\mathbf{P}_0)_{i,j} &= \mathbb{P}(|L_3| = j \mid 1 \in L_1, 2 \in L_2, i+1, E_{n+2,i+1}), \quad 2 \le i \le n+1 \text{ and } 1 \le j \le n-i+1 \\
(\mathbf{P}_1)_{i,j} &= \mathbb{P}(|L_1| = j \mid 1 \in L_1, 2 \in L_2, i+1, E_{n+2,i+1}), \quad 1 \le i \le n+1 \text{ and } 1 \le j \le n-i+1 \\
(\mathbf{P}_0)_{i,j} &= (\mathbf{P}_1)_{i,j} = 0, \quad \text{elsewhere.}
\end{aligned}
$$

The (i,j)th entry of $\mathbf{P}_0$ is the probability that a lineage in level $i+1$ (that is, when $i-1$ undistinguished lineages remain) subtends j undistinguished lineages at present, given that it does not contain 1 or 2. Similarly, the (i,j)th entry of $\mathbf{P}_1$ is the probability that a lineage in level $i+1$ subtends $j-1$ undistinguished lineages, in addition to 1 or 2 (but not both).

Finally, let $\mathbf{E} = 2 \cdot \mathrm{diag}(1/2, 1/3, \dots, 1/(n+2))$ be the diagonal matrix whose i-th entry is the probability that a randomly chosen lineage chosen at level $i+1$ contains 1 or 2. We then have that $\mathbf{E}\mathbf{P}_1 \in \mathbb{Q}^{(n+1)\times(n+1)}$ is the matrix whose (i,j)th entry is the probability that a lineage at level $i+1$ subtends a sample $(1, j-1)$. Similarly, $(\mathbf{I}-\mathbf{E})\mathbf{P}_0$ gives the probability of subtending $(0,j)$.

Now consider the $(0,i)$ entry of CSFS$(\tau\downarrow)$, which is the (conditional) probability of observing a derived allele in i undistinguished lineages, but in neither distinguished lineage. Write $\xi_{m,k} = \{\zeta^{(1)}_{m,k}, \dots, \zeta^{(k)}_{m,k}\}$, where the $\zeta^{(i)}_{m,k}$ are the blocks of $\xi_{m,k}$ randomly ordered. Under the SFS approximation,

$$
\begin{aligned}
\frac{2}{\theta}\mathbb{P}((0,j) \mid C_{12} = \tau) &= \sum_{k=2}^{n+2} \mathbb{E}\big(T^{\tau\downarrow}_{n+2,k} \sum_{\zeta \in \xi_{n+2,k}} \mathbf{1}\{|\zeta| = j, \{1,2\} \cap \zeta = \varnothing\}\big) \\
&= \sum_{k=2}^{n+2} k\mathbb{E}\big(T^{\tau\downarrow}_{n+2,k} \mathbf{1}\{|\zeta^{(1)}_{n+2,k}| = j, \{1,2\} \cap \zeta^{(1)}_{n+2,k} = \varnothing\}\big) \\
&= \sum_{k=2}^{n+2} k\mathbb{E}T^{\tau\downarrow}_{n+2,k} \mathbb{P}(|\zeta^{(1)}_{n+2,k}| = j, \{1,2\} \cap \zeta^{(1)}_{n+2,k} = \varnothing \mid E_{n+2,k}) \\
&= \sum_{k=0}^{n} [\gamma(\tau)\mathbf{B}\mathbf{D}]_k [(\mathbf{I}-\mathbf{E})\mathbf{P}_0]_{kj}.
\end{aligned}
$$

Here, the third equality holds by the same argument as in equation (2.4) in the proof of Theorem 2.2.

We see that the last n columns of the first row of CSFS$(\tau\downarrow)$ are given by the following matrix-vector product:

$$
[\mathrm{CSFS}(\tau\downarrow)]_{0,1:n} = \gamma(\tau)\mathbf{B}\mathbf{D}(\mathbf{I}-\mathbf{E})\mathbf{P}_0. \tag{2.11}
$$

Using similar arguments, the second row can be shown to equal

$$
[\mathrm{CSFS}(\tau\downarrow)]_{1,0:n} = \gamma(\tau)\mathbf{B}\mathbf{D}\mathbf{E}\mathbf{P}_1. \tag{2.12}
$$

All other entries of CSFS$(\tau\downarrow)$ are zero.

Interval Calculation

The above formulas can be used to compute CSFS($\tau\downarrow$) for a fixed coalescence time τ. Since the hidden states of our model consist of intervals for τ, we must integrate these expressions with respect to the coalescence density of 1 and 2. This is easily accomplished since the integral commutes with the linear transforms expressed above. Specifically, for an interval $[t_1, t_2)$ we integrate the vector of first coalescence times $\gamma(\tau)$ against the conditional coalescence density

$$f_{[t_1,t_2)}(t) = \frac{\alpha(t)e^{-R(t)}}{e^{-R(t_1)} - e^{-R(t_2)}} \tag{2.13}$$

to obtain $\gamma' = \int_{t_1}^{t_2} f_{[t_1,t_2)}(\tau)\gamma(\tau)\,d\tau$, which can then be used in (2.11) and (2.12).

Computing CSFS($\tau\uparrow$)

To calculate the CSFS above τ, we adapt the approach of Kamm, Terhorst, and Song (2016) for computing the multi-population coalescent. The reader is referred to that paper for motivating details on the method; briefly, the strategy is to compute the unconditioned SFS on $n+1$ lineages beginning at time $\tau+$, and multiply it by the probability that a lineage which has size $k = 1, \ldots, n$ at time $\tau+$ has size (a, b) at time 0. The latter quantity is calculated as the transition probability of a certain forward-time Moran model on $n+1$ lineages which is dual to the conditioned coalescent.

Formally, we have

$$[\text{CSFS}(\tau\uparrow)]_{ab} = \sum_{i=1}^{n} [\nu(\tau)]_i \times \mathbb{P}(i \to (a,b); \tau) \tag{2.14}$$

where $[\nu(\tau)]_i$ denotes the total branch length of lineages above time τ that have size $i \in \{1, \ldots, n\}$ at time $\tau+$ (i.e. the unnormalized SFS on $n+1$ lineages), and $\mathbb{P}(\cdot;\tau)$ is the aforementioned transition function.

To compute $\nu(\tau)$, let $\varepsilon(\tau) \in \mathbb{R}^n$ be a row-vector whose ith entry is the expected time to first coalescence in a sample of size $i+1$ beginning at time τ:

$$[\varepsilon(\tau)]_i = \int_{\tau}^{\infty} \alpha(t)e^{-a_{i+1}[R(t)-R(\tau)]}\,dt.$$

Polanski and Kimmel (2003) have derived a matrix $\mathbf{W} \in \mathbb{Q}^{n\times n}$ such that the unnormalized SFS on a sample of size $n+1$ (i.e., the total expected branch length having sizes $1, \ldots, n$) equals $\nu(\tau) = \mathbf{W}\varepsilon(\tau)$.

Transition probability

Next we calculate the probability that k lineages at time τ have (a, b) descendants at present (time 0). Note that a lineage above τ must subtend either $a = 0$ or $a = 2$ distinguished lineages

at present. For $a \in \{0, 2\}$, let $M_a \in \mathbb{Q}^{(n+1)\times(n+1)}$ be the tridiagonal matrix

$$(\mathbf{M}_a)_{i+1,j+1} = \begin{cases} (2-a)i + \frac{1}{2}i(n-i), & j = i-1, \\ a(n-i) + \frac{1}{2}i(n-i), & j = i+1, \\ 0, & \text{otherwise.} \end{cases}$$

and let $\mathbf{Q}_a$ be the rate matrix of a continuous time Markov chain whose embedded jump chain has transition matrix $\mathbf{M}_a$, i.e., its diagonals are the negative row sums of $\mathbf{M}_a$.

The processes defined by $\mathbf{Q}_a$ are Moran models whose state is the number of undistinguished lineages bearing a derived allele as time runs towards the present. The integer a is the number of derived alleles in the two distinguished lineages. The two distinguished lineages copy onto each undistinguished lineage in the opposite allelic class at rate 1; additionally the undistinguished lineages copy between classes at rate 1/2 as in the usual Moran model. Copying onto either distinguished lineage is disallowed.

One can verify that the coalescent trees embedded in this Moran model are equal in distribution to those of the conditioned coalescent below τ. Accordingly, the transition density function in equation (2.14) is precisely the matrix exponential

$$\mathbf{M}_a(\tau) := e^{R(\tau)\mathbf{Q}_a}. \tag{2.15}$$

Computing CSFS(τ ↑)

Let $\mathbf{S}_2 := \text{diag}(1, \dots, n)/(n+1) \in \mathbb{Q}^{n\times n}$ be the matrix whose ith diagonal entry is the probability that i lineages randomly chosen from $n+1$ lineages at $\tau+$ subtend both of the distinguished lineages, and set $\mathbf{S}_0 = \mathbf{I} - \mathbf{S}_2$. If $a = 0$, the probability that a lineage which has size i at time τ has size (a, b) at time 0 is $(\mathbf{S}_0)_{ii} \times [\mathbf{M}_0(\tau)]_{i+1,b+1}$. If $a = 2$, then this probability is $(\mathbf{S}_2)_{ii} \times [\mathbf{M}_2(\tau)]_{i,b+1}$, because $i-1$ mutated undistinguished lineages remain after one of the i blocks "splits" into the two distinguished lineages at time τ. Accordingly, define the submatrices $\mathbf{M}'_0(\tau) = [\mathbf{M}_0(\tau)]_{[2:(n+1),2:(n+1)]}$ and $\mathbf{M}'_2(\tau) = [\mathbf{M}_2(\tau)]_{[1:n,1:n]}$. By (2.14) and (2.15) we have

$$\sum_{i=1}^{n} [\nu(\tau)]_i \mathbb{P}(i \to (a,b); \tau) = \sum_{i=1}^{n} (\varepsilon(\tau)^T \mathbf{W}^T)_i (\mathbf{S}_a)_{ii} [\mathbf{M}'_a(\tau)]_{i,b} = [\varepsilon(\tau)^T \mathbf{W}^T \mathbf{S}_a \mathbf{M}'_a(\tau)]_b.$$

Integrating CSFS(τ ↑)

To integrate CSFS(τ ↑) against the conditional coalescence density (2.13), let $\mathbf{U}_a \cdot e^{R(\tau)\mathbf{D}} \cdot (\mathbf{U}_a^{-1}) = \mathbf{M}_a(\tau)$ be the eigendecomposition of $\mathbf{M}_a$, where matrix $\mathbf{D} = \text{diag}(D_1, \dots, D_{n+1})$ and $\mathbf{U}_a \in \mathbb{Q}^{(n+1)\times(n+1)}$ can be computed exactly using results from the next section. Let $\mathbf{U}'_a$ and $(\mathbf{U}_a^{-1})'$ denote appropriate submatrices of $\mathbf{U}_a$ and $(\mathbf{U}_a^{-1})$, respectively, that appear in the decomposition of $\mathbf{M}'_a(\tau)$. Defining $\mathbf{X}_a := \mathbf{W}^T \mathbf{S}_a \mathbf{U}'_a \in \mathbb{Q}^{n\times(n+1)}$, we have

$$\int_{t_1}^{t_2} f_{[t_1,t_2)}(\tau) \varepsilon(\tau)^T \mathbf{W}^T \mathbf{S}_a \mathbf{M}'_a(\tau)\, d\tau = \int_{t_1}^{t_2} f_{[t_1,t_2)}(\tau) \varepsilon(\tau)^T \mathbf{X}_a e^{R(\tau)\mathbf{D}}\, d\tau \cdot (\mathbf{U}_a^{-1})'.$$

The integral simplifies to

$$\int_{t_1}^{t_2} f_{[t_1,t_2)}(\tau)\varepsilon(\tau)^T \mathbf{X}_a e^{R(\tau)\mathbf{D}}\, d\tau$$
$$= \int_{t_1}^{t_2} f_{[t_1,t_2)}(\tau)\left[\varepsilon(\tau)^T(\mathbf{X}_a)_1 e^{-D_1 R(\tau)}, \dots, \varepsilon(\tau)^T(\mathbf{X}_a)_{n+1} e^{-D_{n+1}R(\tau)}\right]^T d\tau = \mathbf{1}^T(\mathbf{X}_a \circ \mathbf{C}^T),$$

where $(\mathbf{X}_a)_j$ denotes the jth column of $\mathbf{X}_a$, $\mathbf{1}$ is a vector of 1s, $\circ$ denotes the Hadamard product, and matrix $\mathbf{C} \in \mathbb{R}^{(n+1)\times n}$ is defined by

$$(\mathbf{C})_{ij} = \frac{1}{e^{-R(t_1)} - e^{-R(t_2)}} \int_{t_1}^{t_2} \alpha(\tau) e^{-(D_i+1)R(\tau)} [\varepsilon(\tau)]_j \, d\tau$$
$$= \frac{1}{e^{-R(t_1)} - e^{-R(t_2)}} \int_{t_1}^{t_2} \alpha(\tau) e^{-[D_i+1-a_{j+1}]R(\tau)} \int_{\tau}^{\infty} e^{-a_{j+1}R(t)}\, dt\, d\tau.$$

Spectral Decomposition of $\mathbf{Q}_a$

We utilize an exact spectral decomposition of $\mathbf{Q}_a$ to perform computations.

Lemma 2.5. *The eigenvalues of* $\mathbf{Q}_a$ *are* $\{1 - a_m : m = 2, 3, \dots, n+2\}$.

Proof. Let $\mathbf{T}_a = \mathbf{I} + (a_{n+2} - 1)^{-1}\mathbf{Q}_a$ be the discrete time version of $\mathbf{Q}_a$. The eigenvalues of $\mathbf{T}_a$ are known to be $\{G_{ii} : i = 0, \dots, n\}$ where G_{ii} is the probability that i randomly sampled individuals at generation t have distinct parents in the generation $t-1$ (Gladstien, 1978). (Here, the probability measure is with respect to the Markov chain corresponding to $\mathbf{T}_a$.) Let $p_{ni} = \binom{n}{i-2}\Big/\binom{n+2}{i}$ be the probability that the random sample contains both distinguished lineages. We have

$$G_{ii} = p_{ni}\left(1 - \frac{a_i - 1}{a_{n+2} - 1}\right) + (1 - p_{ni})\left(1 - \frac{a_i}{a_{n+2} - 1}\right)$$
$$= \frac{(n + i + 1)(n + 2 - i)}{n(n + 3)}$$

after much simplification. The result now follows from basic spectral theory. □

Using the lemma we obtain the left and right eigenvectors $\mathbf{U}_a, \mathbf{U}_a^{-1} \in \mathbb{Q}^{(n+1)\times(n+1)}$ of $\mathbf{Q}_a$ by solving tridiagonal systems for each eigenvalue.

2.6 Locus-Skipping EM Algorithm

In this section we outline the procedure, described in Section 2.4, for efficiently computing the posterior expected number of emissions and transitions in long stretches of nonpolymorphic

sites. We presume some familiarity with the forward-backward and Baum-Welch algorithms; a good introduction as well as the source of most of the notation in this section may be found in Bishop (2006, §13.2). In particular, the notation in this section is separate from the notation of the rest of the chapter (e.g. α, ξ are overloaded to have a different meaning here).

To begin, let $\mathbf{X} = (x_1, \dots, x_L)^T$ be the vector of observed data and let $\mathbf{Z} = (z_1, \dots, z_L)^T$ be the (unobserved) vector of hidden states at each location in the HMM. We suppose that between observations x_k and x_ℓ there is a long stretch of identical monomorphic observations $x_{k+1} = \cdots = x_{\ell-1}$. Let T be the transition matrix and B be a diagonal matrix with entries $B_{ii} = \mathbb{P}(x_{k+1} \mid z_{k+1} = i)$. Define W to be their product $W = TB$.

The forward probabilities at $\ell - 1$, denoted by the vector $\alpha(z_{\ell-1})$, are then

$$\begin{aligned}\alpha(z_{\ell-1}) &= \mathbb{P}(z_{\ell-1}, x_1, x_2, \dots, x_{\ell-1}) \\ &= (W^T)^{\ell-k-1}\alpha(z_k)\end{aligned}$$

which can be computed in a single step from $\alpha(z_k)$, skipping over the positions $\{k+1, k+2, \dots, \ell-2\}$. The locus-skipping transition matrix $(W^T)^{\ell-k-1}$ is efficiently obtained by performing an eigendecomposition of W; see below.

Similarly, the backward probabilities at k, denoted by $\beta(z_k)$, can be computed in a single step from $\beta(z_{\ell-1})$:

$$\begin{aligned}\beta(z_k) &= \mathbb{P}(x_{k+1}, x_{k+2}, \dots, x_L \mid z_k) \\ &= W^{\ell-k-1}\beta(z_{\ell-1}).\end{aligned}$$

To prevent underflow one must rescale the forward and backward probabilities (Bishop, 2006, §13.2.4). To do so, we define $\hat{\alpha}_k \overset{\text{def}}{=} c_k^{-1}\alpha_k$ and $\hat{\beta}_k \overset{\text{def}}{=} \left(\prod_{m=k+1}^{L} c_m\right)^{-1}\beta_k$, where $c_k \overset{\text{def}}{=} \mathbb{P}(x_k \mid x_1, \dots, x_{k-1})$. The $\hat{\alpha}_k$ and $\hat{\beta}_k$ recursions are identical to the unrescaled case, and c_k can be easily computed as the ℓ_1-norm of $W^T\hat{\alpha}_{k-1}$. Note also that $\mathbb{P}(\mathbf{X}) = \prod_{k=1}^{L} c_k$.

The EM algorithm also requires the Baum-Welch quantities $\xi(z_{j-1}, z_j) = \mathbb{P}(z_{j-1}, z_j \mid \mathbf{X})$ and $\gamma(z_j) = \mathbb{P}(z_j \mid \mathbf{X})$ to be summed over all j, including the "skipped" loci j with $k < j < \ell$. To sum these quantities at the skipped loci, we first note

$$\begin{aligned}\hat{\alpha}(z_j) &= \Big(\prod_{i=1}^{j} c_i\Big)^{-1}\alpha(z_j) = \Big(\prod_{i=1}^{j} c_i\Big)^{-1}(W^T)^{j-k}\alpha(z_k) = \Big(\prod_{i=k+1}^{j} c_i\Big)^{-1}(W^T)^{j-k}\hat{\alpha}(z_k) \\ \hat{\beta}(z_j) &= \Big(\prod_{i=j+1}^{L} c_i\Big)^{-1}\beta(z_j) = \Big(\prod_{i=j+1}^{L} c_i\Big)^{-1}W^{\ell-j-1}\beta(z_{\ell-1}) = \Big(\prod_{i=j+1}^{\ell-1} c_i\Big)^{-1}W^{k}\hat{\beta}(z_{\ell-1})\end{aligned}$$

so that

$$\gamma(z_j) = \operatorname{diag}\big(\hat{\alpha}(z_j)\hat{\beta}(z_j)\big)^T\big) = \Big(\prod_{i=k+1}^{\ell-1} c_i\Big)^{-1}\operatorname{diag}\Big((W^T)^{j-k}\hat{\alpha}(z_k)\hat{\beta}(z_{\ell-1})^T(W^T)^{\ell-j-1}\Big) \quad (2.16)$$

Now let

$$PDP^{-1} = TB \tag{2.17}$$

be the eigendecomposition of W.[1] The expected number of positions in the interval (k, l) spent at each hidden state is obtained by summing (2.16):

$$\begin{aligned}\sum_{j=k+1}^{\ell-1} \gamma(z_j) &= \operatorname{diag}\Big[P\Big(\sum_{j=k+1}^{\ell-1} D^{j-k}P^{-1}\hat{\alpha}(z_k)\hat{\beta}(z_{\ell-1})^T PD^{\ell-j-1}\Big)P^{-1}\Big] \\ &= \operatorname{diag}\big(PAP^{-1}\big)\end{aligned} \tag{2.18}$$

for $A \stackrel{\text{def}}{=} \sum_{j=k+1}^{\ell-1} D^{j-k}P^{-1}\hat{\alpha}(z_k)\hat{\beta}(z_{\ell-1})^T PD^{\ell-j-1}$.

It is not difficult to show that

$$\Big(\sum_{i=0}^{m} D^i U D^{m-i}\Big)_{ab} = U_{ab}\frac{D_{aa}^{m+1} - D_{bb}^{m+1}}{D_{aa} - D_{bb}} \stackrel{\text{def}}{=} U_{ab}Q_{ab} \tag{2.19}$$

for diagonal matrix D and arbitrary matrix U. Using (2.19) with $m = \ell - k - 2$ and setting

$$U \stackrel{\text{def}}{=} P^{-1}\hat{\alpha}(z_k)\hat{\beta}(z_{\ell-1})^T P$$

we obtain $A = D(U \circ Q)$, where "$\circ$" denotes Hadamard (entry-wise) product.

To obtain $\sum_{j=k+1}^{\ell-1} \xi(z_{j-1}, z_j)$ we simply note that

$$\xi(z_{j-1}, z_j) = \mathbb{P}(z_{j-1}, z_j, \mathbf{X}) = (\hat{\alpha}(z_{j-1})\hat{\beta}(z_j)^T B) \circ T$$

so that

$$\begin{aligned}\sum_{j=k+1}^{\ell-1} \xi(s_{j-1}, s_j) &= \Big[P\Big(\sum_{j=k+1}^{\ell-1} D^{j-k-1}P^{-1}\hat{\alpha}(z_k)\hat{\beta}(z_{\ell-1})^T PD^{\ell-j-1}\Big)P^{-1}B\Big] \circ T \\ &= [P(U \circ Q)P^{-1}B] \circ T.\end{aligned} \tag{2.20}$$

Complexity

Assuming a total of M hidden states, we see that the computational cost of evaluating equations (2.18) and (2.20) is that of an $M \times M$ matrix-matrix multiply which is at most $O(M^3)$. (At each segregating site we employ the standard algorithms, which have lower cost $O(M^2)$.) The spectral decomposition (2.17) and formation of the auxiliary matrices U and Q all have strictly lower complexity. If we assume that between each of L_p polymorphic sites there is a large stretch of nonpolymorphic sites then the total cost of the algorithm is $O(L_p M^3)$.

[1] Note that in general P and D will not be real matrices, as W is not symmetric. This may seem troublesome since arithmetic over the complex field is a notorious source of inaccuracy in numerical analysis. However, for non-segregating sites we will have $B = I + E_1$ and similarly $T = I + E_2$ for some real-valued perturbation matrices E_i with $\|E_i\| \ll 1$. It follows by the Gershgorin circle theorem that $D = \Re(D) + i\Im(D)$ will have $\big\|\Im(D)\big\| \ll 1$, so that D is "almost" real. In practice we have encountered no problems by truncating the imaginary part of D.

2.7 Extension to multiple populations

In this section we describe how the above conditioned SFS may be extended to handle multiple populations. We focus on the specific case of a "clean split" model in which two subpopulations are descended from a common ancestral population, with no gene flow occurring more recently than t_S generations ago. This section draws heavily on the ideas used by the program `momi` (Kamm, Terhorst, and Song, 2016) to calculate the *unconditioned* joint frequency spectrum.

The "clean split" model we analyze is as follows. Population 1 has instantaneous rate of coalescence $\alpha_1(t)$ and cumulative coalescence function $R_1(t) \stackrel{\text{def}}{=} \int_0^t \alpha_1(s)\,ds$. Going forward in time, at time t_S population 2 splits off from population 1 and follows its own size history with coalescence rate function $\alpha_2(t), 0 \le t < t_S$. Hence the cumulative coalescence rate function for population 2 is

$$R_2(t) = \begin{cases} \int_0^t \alpha_2(s)\,ds, & t \le t_S \\ \int_0^{t_S} \alpha_2(s)\,ds + \int_{t_S}^t \alpha_1(s)\,ds, & t > t_S. \end{cases}$$

Analogously to the CSFS described above, the joint conditioned SFS (JCSFS) is a 4-dimensional tensor which gives describes the sampling distribution of segregating sites in n_1+n_2+2 sampled lineages, where n_i samples are obtained from population i, conditioned on the coalescence time of a pair of distinguished lineages. The distinguished haploid lineages can either arise in the same subpopulation (the "together" case), in which case phased data is not required, or one can be sampled from each subpopulation (the "apart" case).

For the rest of this section we make use of the following notation. The joint CSFS is denoted

$$\text{JCSFS}(\tau, t_S) \in \mathbb{R}^{(a_1+1)\times(n_1+1)\times(a_2+1)\times(n_2+1)}.$$

Here, τ denotes the conditioned coalescence time of the distinguished pair, t_S is the split time, and $n_i(a_i)$ is the number of (un)distinguished lineages sampled in population i. We require that $a_1 + a_2 = 2$; the "together" case corresponds to $a_1 = 2, a_2 = 0$ and the "apart" case, $a_1 = a_2 = 1$. For brevity, we drop the explicit dependence on τ and t_S from now on. The entry JCSFS$[i, j, k, l]$ therefore gives the probability of observing $i + k$ derived alleles in the distinguished lineages, j derived alleles in subpopulation 1, and l derived alleles in subpopulation 2.

The following quantities will be useful below. The *truncated (C)SFS* is defined to be the total expected branch length occurring beneath the split in each subpopulation. It is obtained mathematically by sending the population size above the split to zero (see (Kamm, Terhorst, and Song, 2016)). For n lineages we denote this as (C)SFS$_n(\overline{t_S})$. The *ancestral (J)CSFS* is defined to be expected site frequency spectrum of the ancestral population sampled immediately before the split. It is equivalent to computing the standard (J)CSFS using the cumulative rate function $R'(t) \stackrel{\text{def}}{=} R(t + t_S)$.

Case 1: Together

In this case we write $J[i,j,k]$ for the probability of observing $0 \le i \le 2$ derived alleles in the distinguished pair, $0 \le j \le n_1$ in population 1, and $0 \le k \le n_2$ in population 2.

We distinguish two cases, depending on whether the distinguished lineages coalesce before or after the split. In either case, the total expected branch length beneath the split subtending population 2 (and no lineages in population 1) equals the SFS on n_2 lineages truncated to time t_S:

$$\text{JCSFS}[0,0,k] = \text{SFS}(\overline{t_S})[k].$$

$\tau < t_S$

If the distinguished lineages coalesce before the split, then the total expected branch length subtending population 1(2) and no samples from the other population is simply the JCSFS(τ) (SFS) truncated at time t_S. (For the additional entries $J[2,n_1,0]$ and $J[0,0,n_2]$ see Kamm, Terhorst, and Song, 2016, Lemma 1).

For the branch length above τ subtending we follow the same approach as in Section 2.5, by computing the ancestral JCSFS and using Moran models to transition forwards in time. That is, the total branch length subtending (i,b_1,b_2) at the present is equal to

$$\text{JCSFS}[i,b_1,b_2] = \sum_{s=1}^{n_1+n_2} \sum_{t=(s-n_2)\vee 0}^{s\wedge(n_1+1)} h(t,n_1+n_2+1,s,n_1+1)\cdot \mathcal{A}(s)\cdot \mathbb{P}_{1,i}(s\to b_1)\cdot \mathbb{P}_2(s-t\to b_2).$$

Here, $h(k,M,n,N) = \binom{n}{k}\binom{M-n}{N-k}/\binom{M}{N}$ is the hypergeometric density, and $\mathbb{P}_\alpha(i\to j)$ are transition functions. For $\alpha = 2$, we have

$$\mathbb{P}_2(i\to j) = (e^{t_S \mathbf{M}_{n_2}})_{ij}$$

where $\mathbf{M}_k \in \mathbb{R}^{(k+1)\times(k+1)}$ is the standard Moran rate matrix on k lineages. For $\alpha = (1,0)$ and $\alpha = (1,2)$ the process is inhomogeneous since the dynamics change at time τ. Hence,

$$\mathbb{P}_{(1,0)}(i\to j) = \left(e^{(t_S-\tau)\mathbf{M}_{n_1+1}} e^{\tau \mathbf{M}_{(0,n_1)}}\right)_{ij}$$

where $\mathbf{M}_{(0,n_1)}$ is the modified rate matrix described in Section 2.5, and similarly for $\alpha = (1,2)$.

$\tau \ge t_S$

The total expected branch length in population 1 beneath the split is equal to CSFS($t_S\downarrow$) (see Section 2.5). For the branch length above t_S subtending both populations, we compute ancestral CSFS(τ) on n_1+n_2 lineages (denoted $\mathcal{A}$) and use Moran models to transition to the present:

$$\text{JCSFS}[i,b_1,b_2] = \sum_{s=1}^{n_1+n_2} \sum_{t=(s-n_2)\vee 0}^{s\wedge(n_1+1)} h(t,n_1+n_2,s,n_1)\cdot \mathcal{A}(i,s)\cdot \mathbb{P}_{1,i}(s\to b_1)\cdot \mathbb{P}_2(s-t\to b_2).$$

In this case $\mathbb{P}_2(i \to j)$ is the same as in the preceding section, and $\mathbb{P}_{(1,j)}(i \to j) = (e^{t_S \mathbf{M}_{(j,n_1)}})_{ij}$ where $\mathbf{M}_{(j,n_1)}$ is defined as in Section 2.5.

Case 2: Apart

In this case JCSFS is a 4-dimensional tensor which we will index as JCSFS$[i, j, k, \ell]$. One distinguished lineage is sampled from each population, and no coalescences can occur before t_S. To compute the branch length above t_S subtending the present, we again sample the ancestral CSFS $\mathcal{A}$ at time t_S. Let $\gamma \stackrel{\text{def}}{=} h(t, n_1 + n_2, s, n_1)$ and define the matrix exponentials $T_{ij} \stackrel{\text{def}}{=} e^{\mathbf{M}_{(j,n_i)}}$ as in the preceding section. We distinguish four sub-cases.

1. For $i = k = 1$ we have

$$\text{JCSFS}[1, b_1, 1, b_2] = \sum_{s=1}^{n_1+n_2} \sum_{t=(s-n_2)\vee 0}^{s\wedge(n_1+1)} \gamma \cdot \mathcal{A}(2, s) \cdot T_{11}(t, b_1) \cdot T_{21}(s - t, b_2)$$

2. For $i = 1, k = 0$ we have

$$\text{JCSFS}[1, b_1, 0, b_2] = \sum_{s=1}^{n_1+n_2} \sum_{t=(s-n_2)\vee 0}^{s\wedge(n_1+1)} \frac{1}{2} \cdot \gamma \cdot \mathcal{A}(1, s) \cdot T_{11}(t, b_1) \cdot T_{20}(s - t, b_2).$$

   The case $i = 0, k = 1$ in analogous.

3. For $i = k = 0$ we have

$$\text{JCSFS}[0, b_1, 0, b_2] = \sum_{s=1}^{n_1+n_2} \sum_{t=(s-n_2)\vee 0}^{s\wedge(n_1+1)} \gamma \cdot \mathcal{A}(0, s) \cdot T_{10}(t, b_1) \cdot T_{20}(s - t, b_2)$$

   Finally, the branch length subtending population i beneath t_S is given by the truncated SFS on $n_i + 1$ lineages. For $1 \le j \le n_i$ the total expected branch length subtending 0 distinguished and j undistinguished lineages beneath t_S equals

$$\text{SFS}_{n_i+1}(\overline{t_S}) \frac{j}{n_i + 1},$$

   while the expected branch length subtending 1 distinguished and $j - 1$ undistinguished is

$$\text{SFS}_{n_i+1}(\overline{t_S}) \frac{n_i + 1 - j}{n_i + 1}.$$

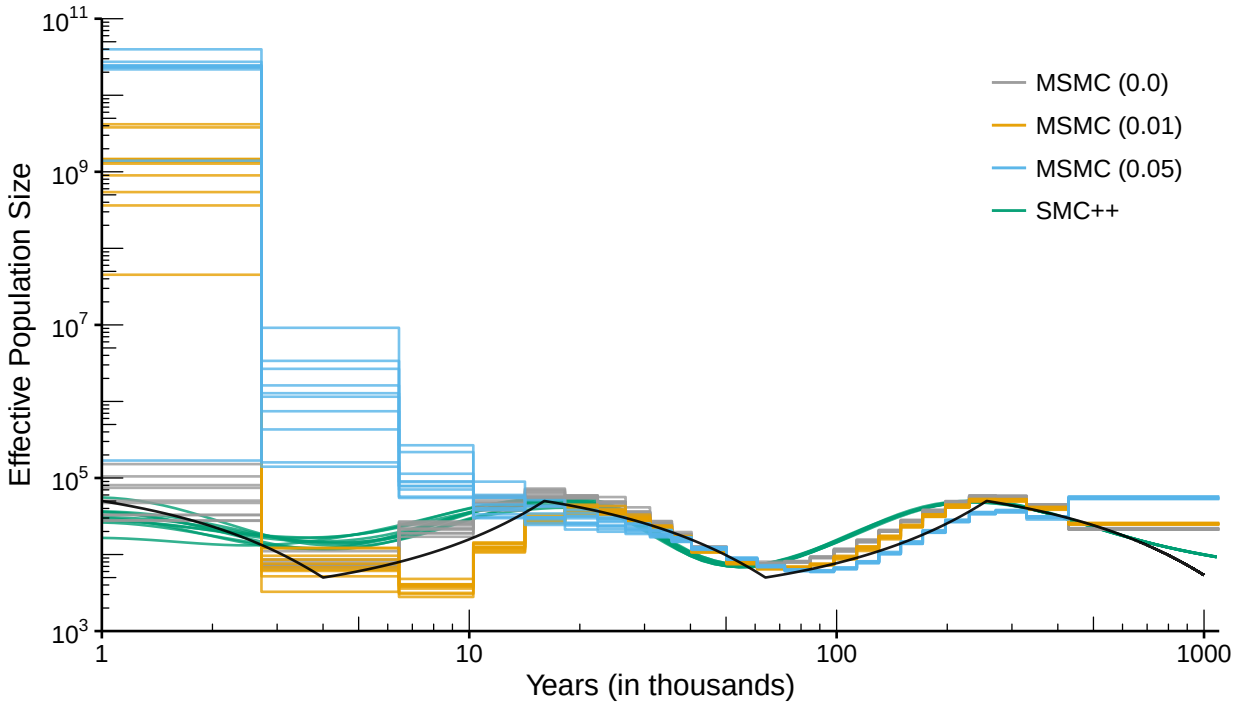


Figure 2.1: The effect of phasing error. The true population size history is indicated by a bold black line, while colored lines indicate inferred histories for ten simulations each with sample size $n = 4$. For MSMC, switch error was introduced at the rate of 0%, 1%, or 5%, indicated in parenthesis in the legend. SMC++ does not require phased data and its results are insensitive to phasing errors. With phasing error, MSMC estimates can be off by orders of magnitude in the recent past. In the absence of phasing error, the accuracy of MSMC is comparable to that of SMC++, with SMC++ producing higher resolution in the recent past.

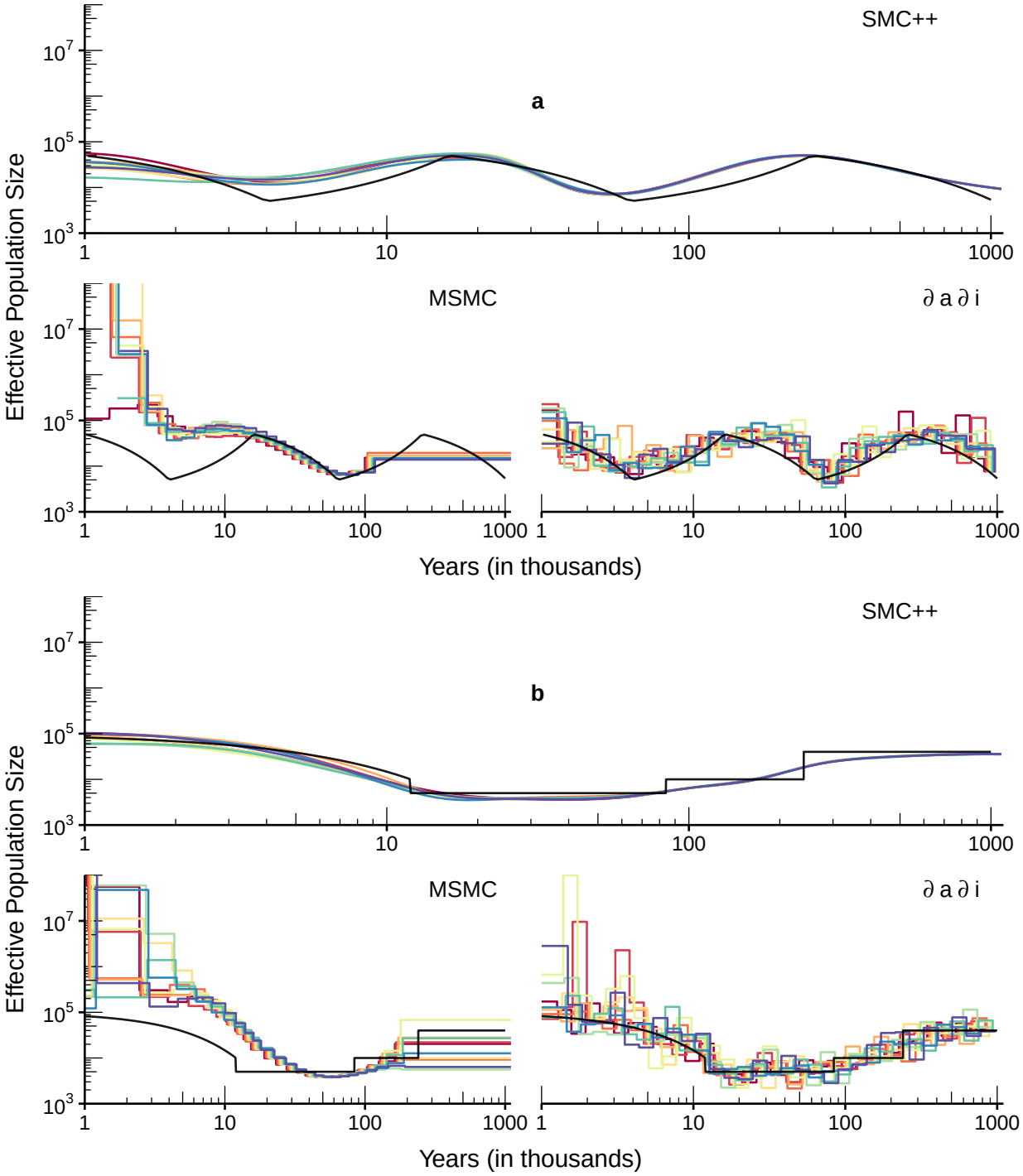


Figure 2.2: Performance of SMC++ compared to MSMC and $\partial a \partial i$. (a) The sawtooth demography. (b) The recent-expansion demography. Each method was used to analyze ten simulated datasets generated according to the demography shown in black. SMC++ was given sequence data from $n = 100$ lineages, and $\partial a \partial i$ analyzed the SFS from that data set. MSMC analyzed $n = 8$ of those lineages, the largest sample size for which it successfully ran. For this simulation, we introduced switch errors at a rate of 1% at segregating sites.

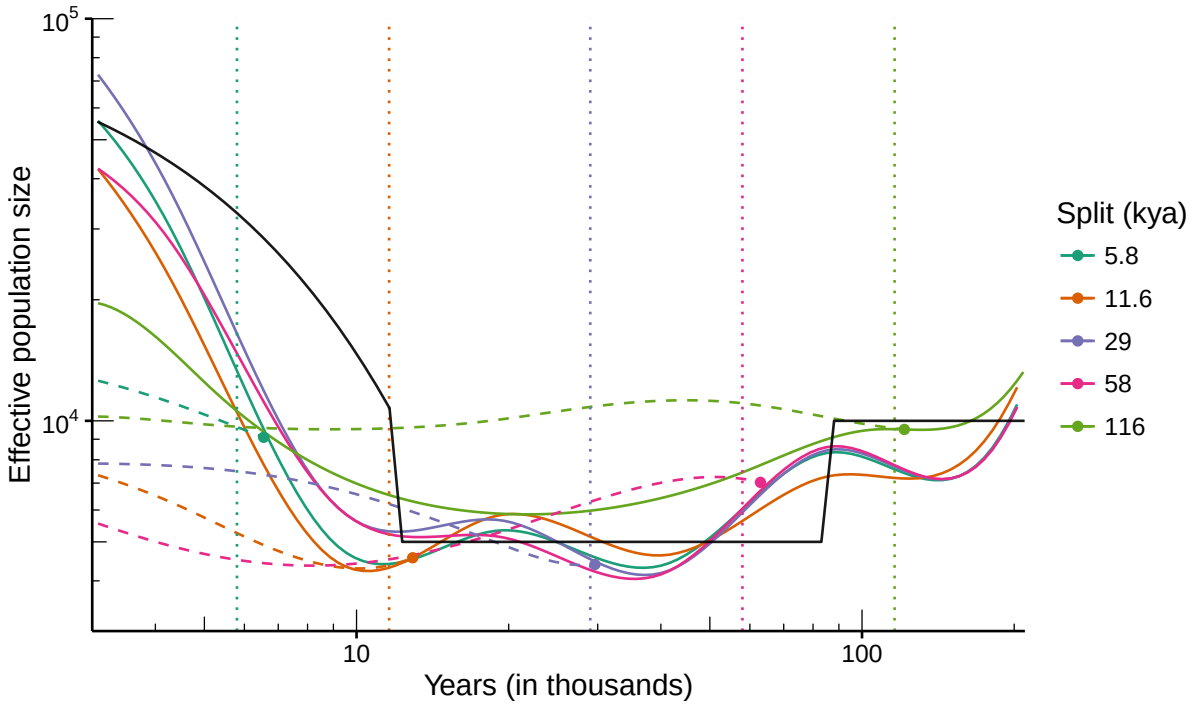


Figure 2.3: SMC++ results of jointly inferring population size histories and divergence times. Two populations were simulated under the "recent-expansion" demography described above. Each population consisted of $n = 10$ lineages. Different colors correspond to different divergence times. From the point of divergence until present, population 2 maintains a constant effective population size equal to the value it had at the time of the split. The solid colored lines indicate the inferred demography for population 1, which should follow the solid black line indicating the simulated demography. The dashed colored lines indicate the inferred demography for population 2, which should be flat from the time of the split onwards. The vertical dotted lines represent the true value of the split, whereas solid dots in corresponding color represent the value of the inferred split time. This result shows that our method is able to infer divergence times with low error over a wide range of split times, spanning approximately 6–120 kya.

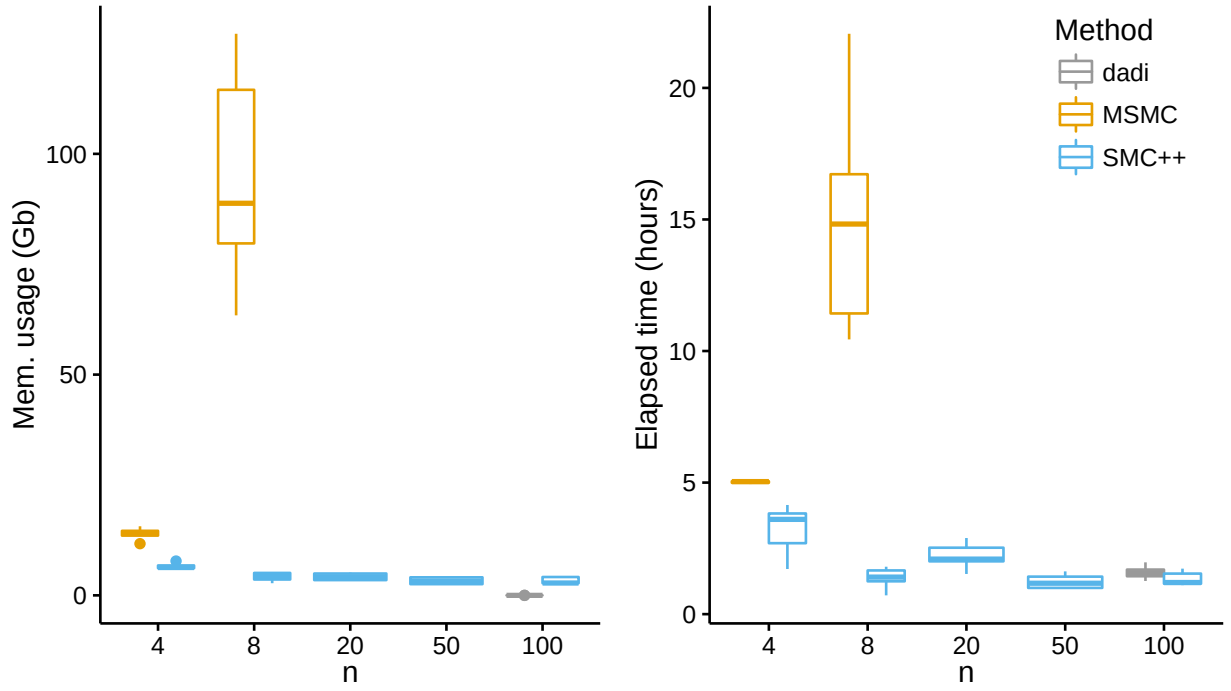


Figure 2.4: Computational performance of SMC++, MSMC, and $\partial a \partial i$. The plots show median memory usage and runtime; error bars denote interquartile range. Each datum comprises ten repetitions on 3Gb of simulated data. The largest sample size for which we were able to successfully run MSMC was $n = 8$. For large sample sizes ($n \geq 8$), SMC++ requires orders of magnitude less memory and time than does MSMC. The lower and upper hinges represent 25th and 75th percentiles; the middle line is the median. Whiskers extend to the nearest observation less than 1.5IQR beyond the corresponding hinges.

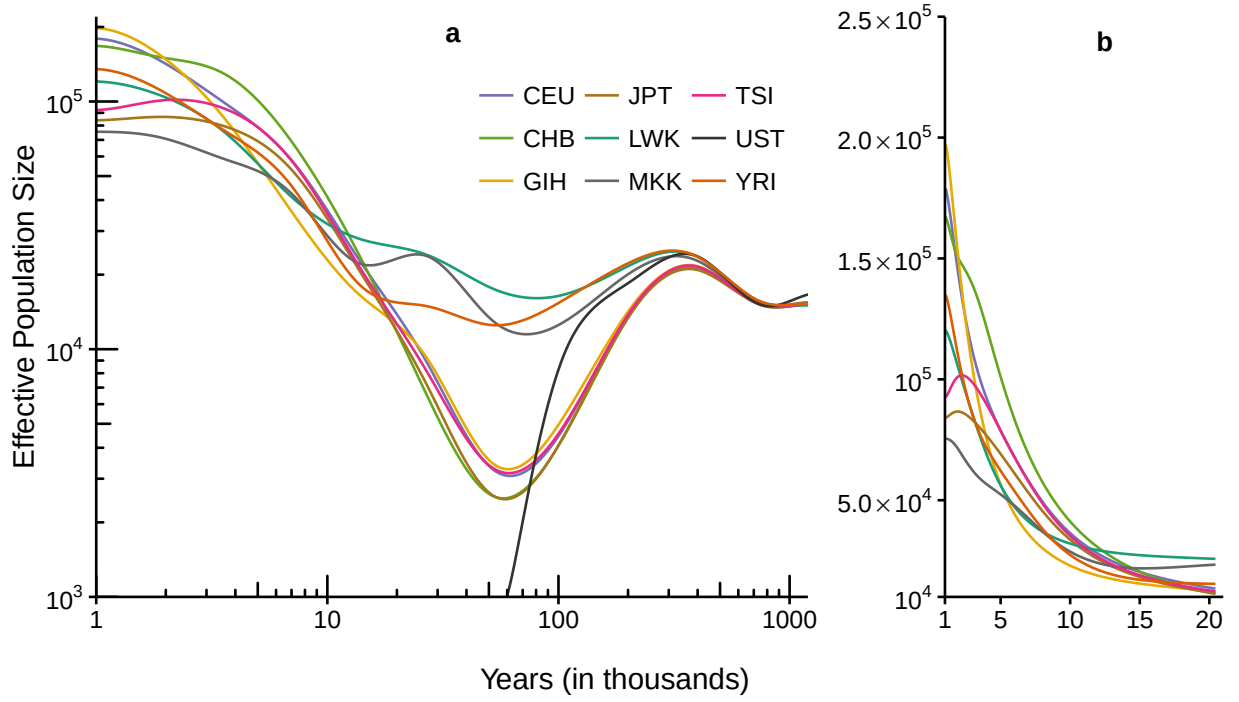


Figure 2.5: Results of effective population size inference across eight extant human populations and an ancient Ust'-Ishim individual. A generation time of 29 years was used to convert the coalescent scaling to calendar time. (a) Results for all populations populations on a log-log scale. Plot assumes that the Ust'-Ishim individual lived until 45 kya. (b) Results for present-day populations on a linear scale over the past 20 ky. See Table 2.2 for a description of the populations and sample sizes.

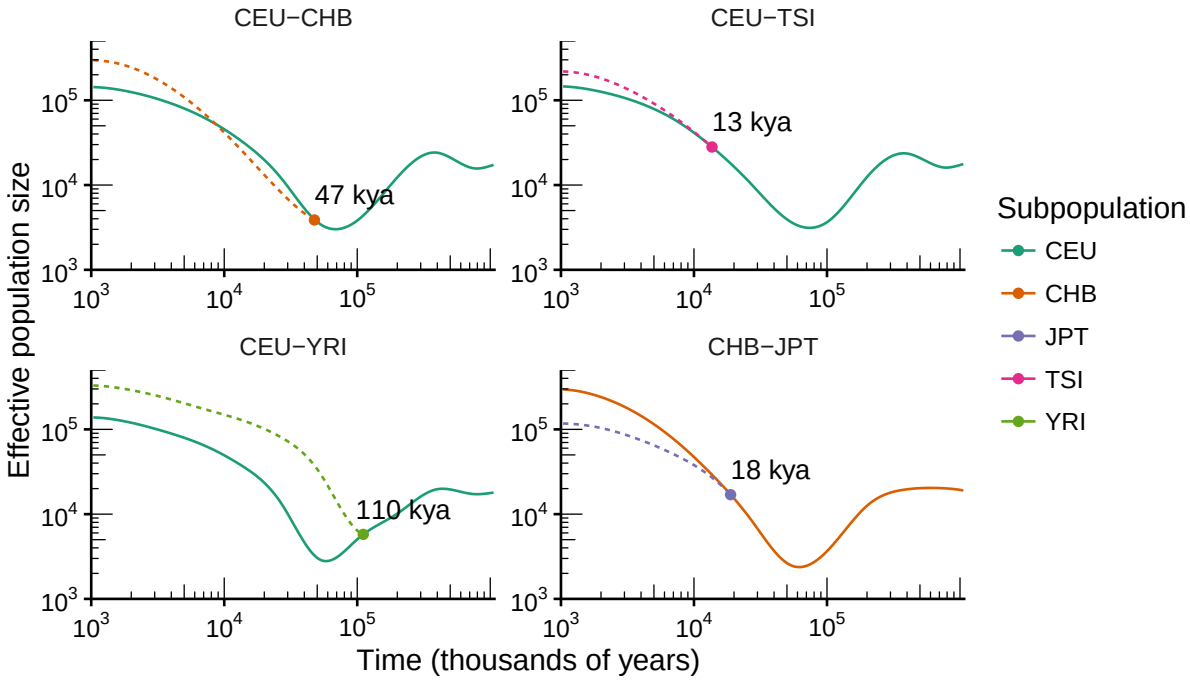


Figure 2.6: Inference of split times in modern humans. Results of jointly estimating population size histories and split times in a two-population model. The same data and generation times as in Figure 2.5 were used to generate the plot.

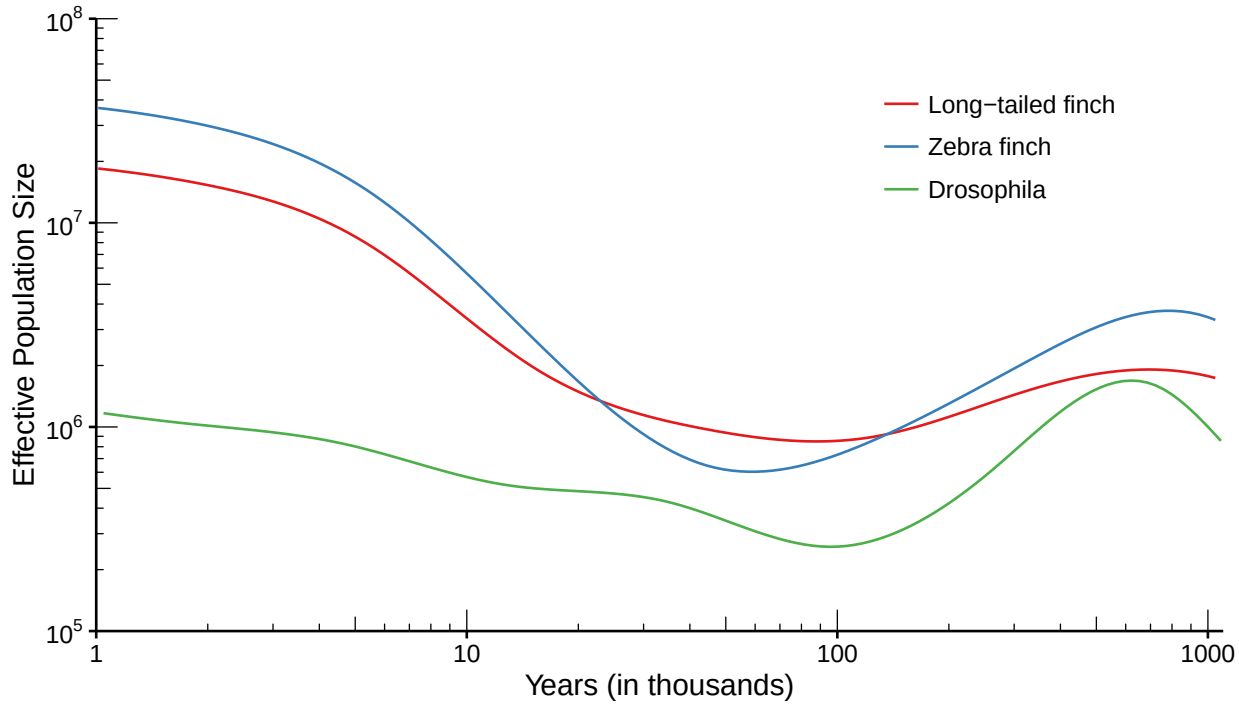


Figure 2.7: Results of effective population size inference for two finch species and *D. melanogaster*. Generation times of 3 months (finch) and 1 month (*D. melanogaster*) were used to convert the coalescent scaling to calendar time. See Table 2.3 for a description of the populations and sample sizes.

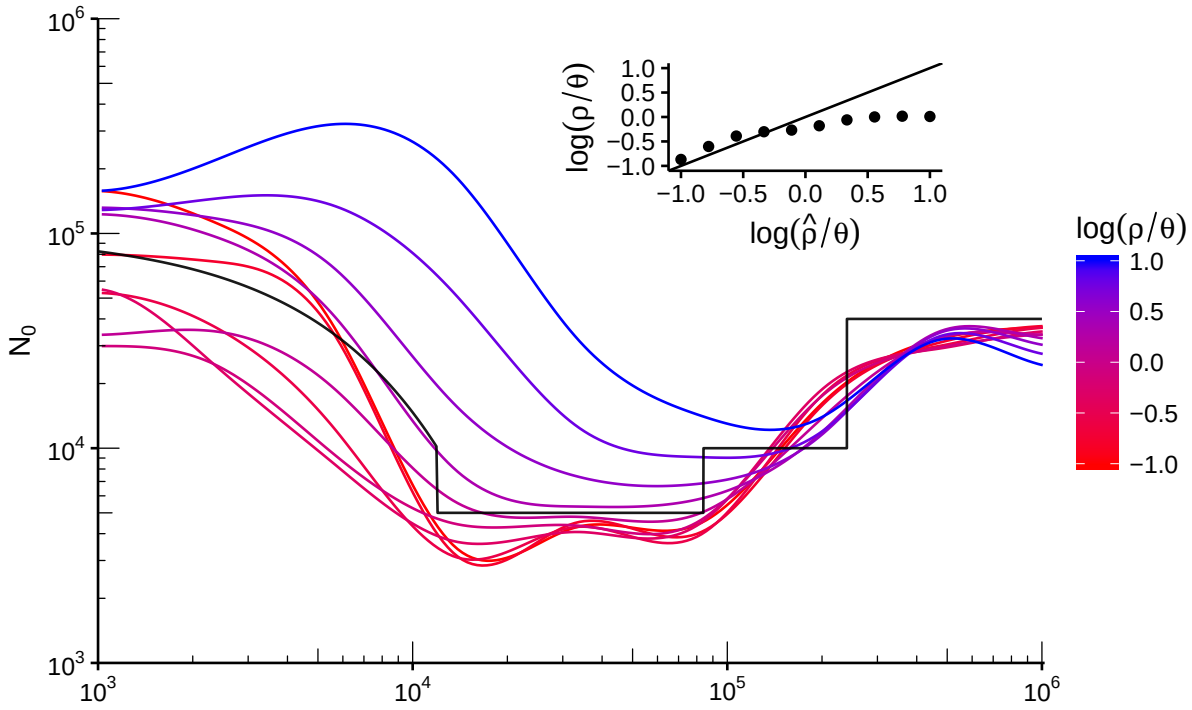


Figure 2.8: Results of demographic inference when ρ is not known. Each step plot represents inference on a single simulated data set with sample size $n = 50$. The colors of the estimated size histories indicate the ratio of recombination to mutation used in each simulation, which was not known to SMC++ during model fitting. The ratio ranged from $1:10$ (black) to $10:1$ (light blue). The true demography used for simulation is indicated in bold black. The nested scatter plot compares the true versus estimated ratio of recombination to mutation rates. The mutation rate $\theta/2$ was assumed to be known. SMC++ is able to fairly accurately estimate the recombination rate over two orders of magnitude with respect to the mutation rate, and is most accurate when the mutation and recombination rates are approximately equal.

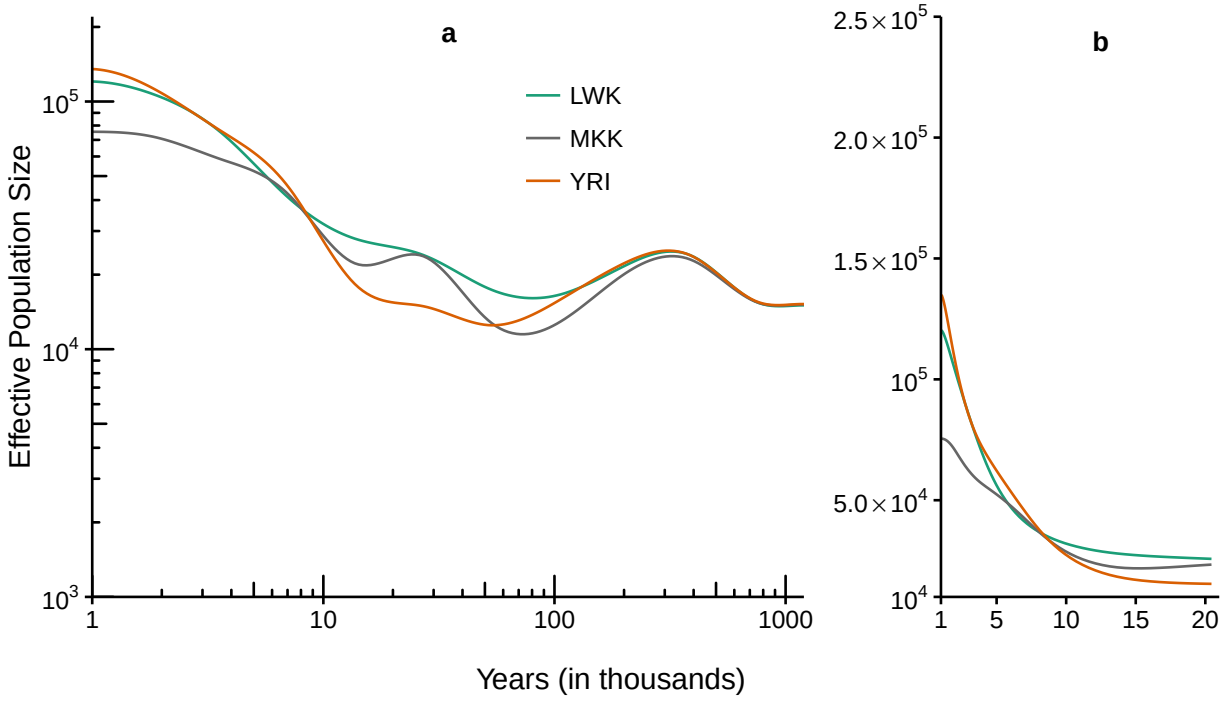


Figure 2.9: Results of demographic inference across three African subpopulations.

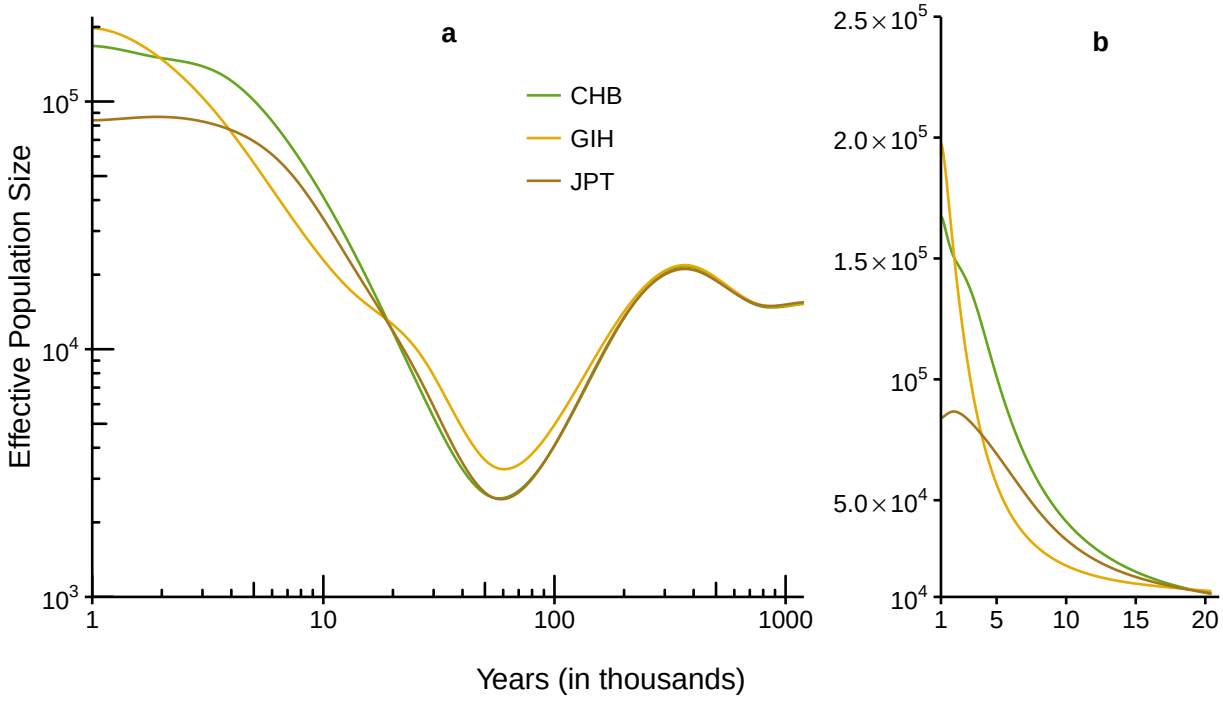


Figure 2.10: Results of demographic inference across three Asian subpopulations.

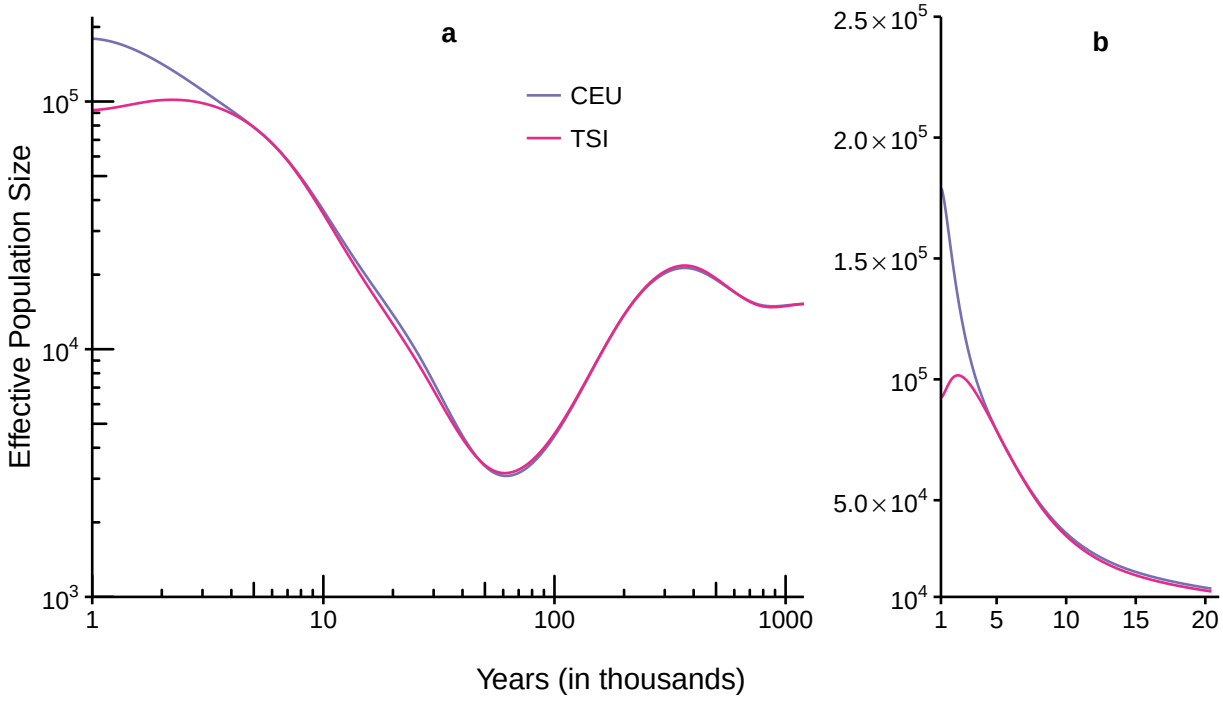


Figure 2.11: Results of demographic inference across two European subpopulations.

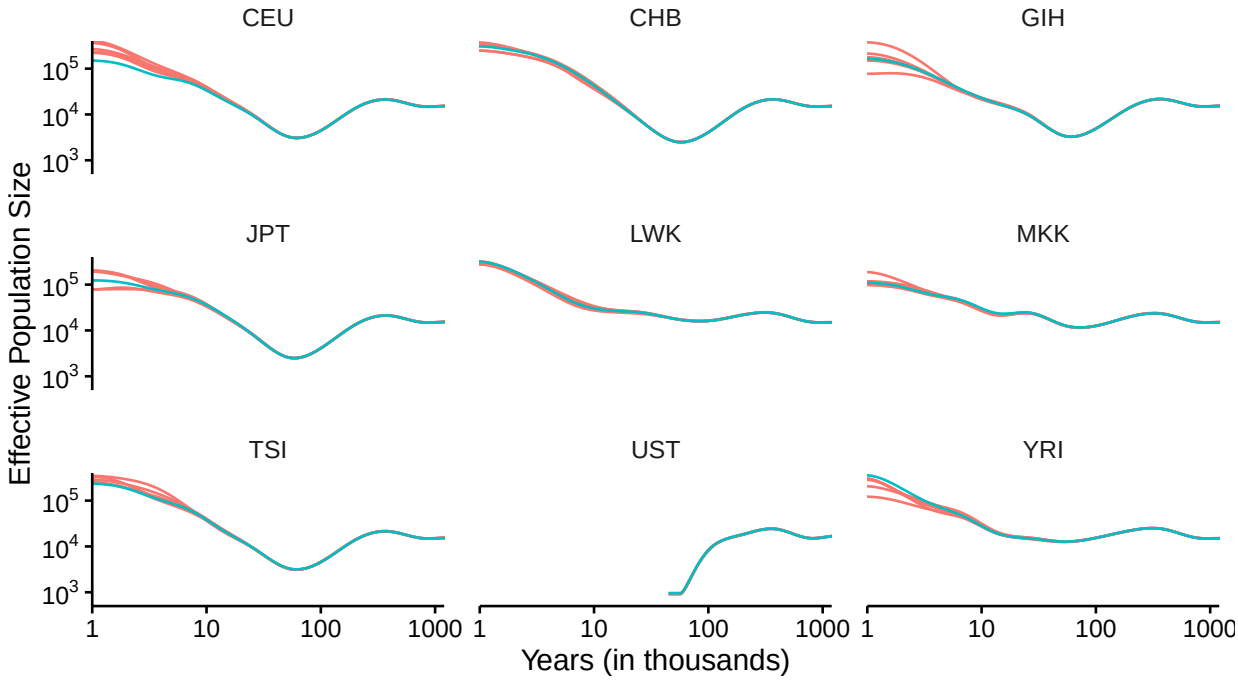


Figure 2.12: Sensitivity analysis for human demographic inference. Blue lines are reproduced from Figure 2.5. Red lines represent the result of randomly down-sampling the data to contain 90% of the original set of chromosomes and re-running the analysis.

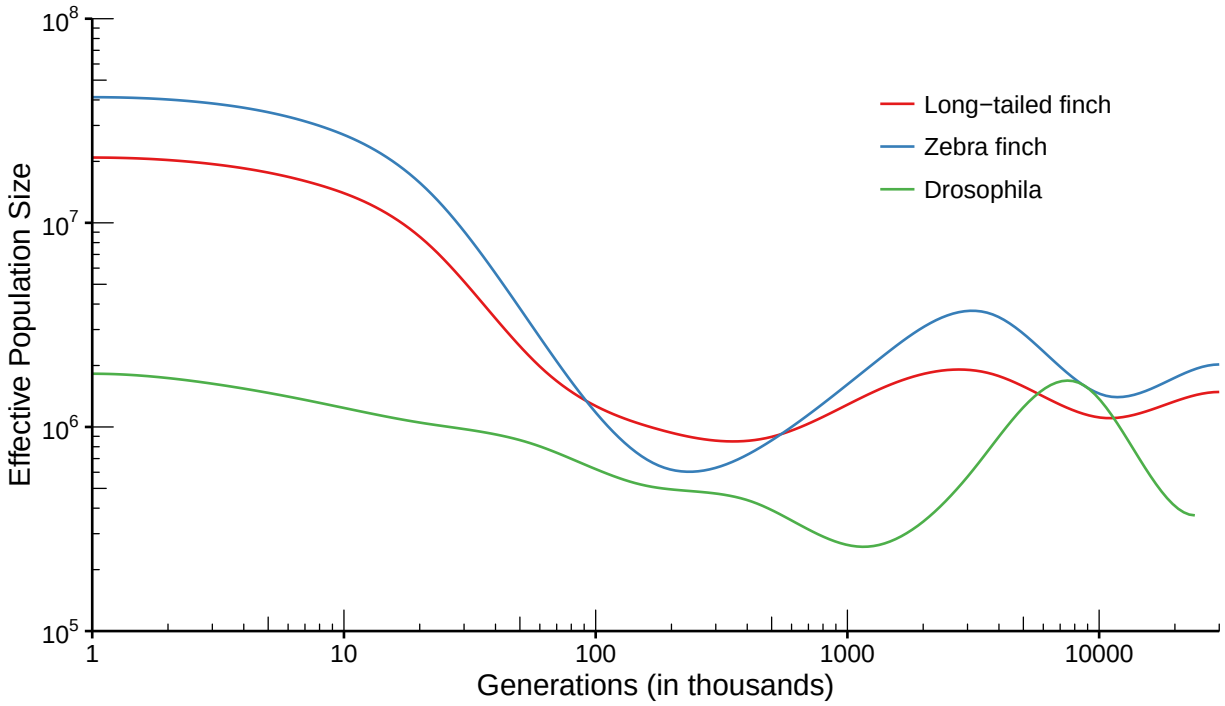


Figure 2.13: Results of analyzing non-human species, in generations.

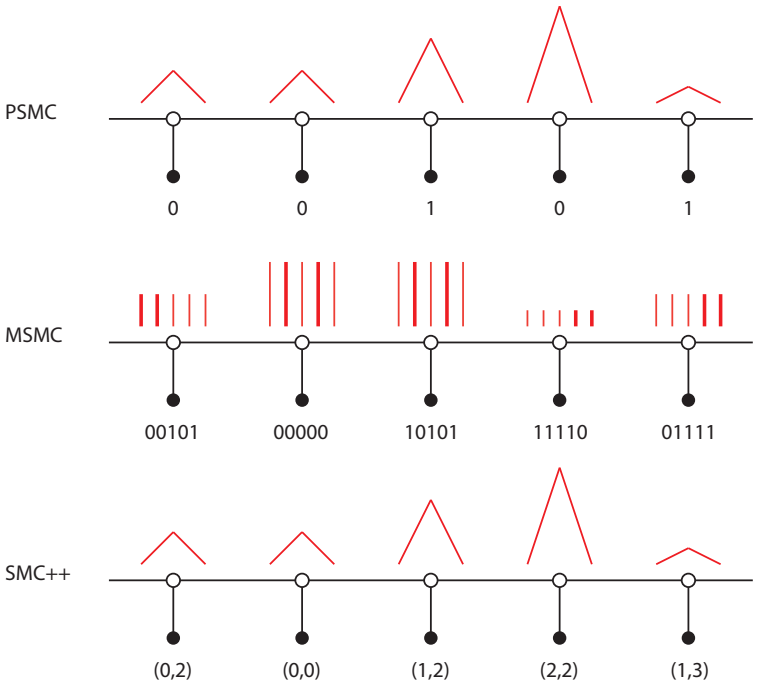


Figure 2.14: Schematic of the differences between PSMC, MSMC, and SMC++. The HMM used in PSMC tracks the hidden TRMCA of a pair of haploid lineages, and emits binary symbols based on the heterozygosity of this pair at each block of sites. MSMC tracks the hidden time to first coalescence among several haploid lineages, as well as the identity (denoted by the bolded bars) of this two lineages which coalesce first. It considers as emissions the allelic state of all lineages in the sample. SMC++, like PSMC, tracks the TMRCA in only a pair of individuals, and emits 2-tuples whose distribution is given by the conditioned SFS (Section 2.5).

Table 2.1: Effect of data quality and sample size on the accuracy of posterior decoding a 10Mb region. The first three columns indicate the ratio of recombination to mutation, error rate, and fraction of missing data, respectively. The table entries are the ratio of posterior mean-squared error at the indicated sample size compared to sample size $n = 2$, averaged over ten simulations. (Hence, lower is better.) Entries marked with a "*" are significantly different from 1.0 at the 5% level using a two-sided t-test.

			n		
ρ/μ	$-\log_{10}(err)$	Miss. %	**5**	**10**	**25**
10.0	3	0	0.96*	0.95*	0.95*
		10	0.99*	0.99*	0.99*
		20	0.98*	0.98*	0.98*
	4	0	0.99*	0.99*	0.99*
		10	0.99*	0.99*	0.99*
		20	0.98*	0.98*	0.98*
	5	0	0.99*	0.99*	0.99*
		10	0.99*	0.99*	0.98*
		20	0.98*	0.98*	0.98*
1.0	3	0	0.82*	0.74*	0.68*
		10	0.96	0.93	0.94
		20	0.95*	0.94*	0.94*
	4	0	0.98*	0.97*	0.97*
		10	0.98*	0.98*	0.98*
		20	0.96*	0.95*	0.95*
	5	0	0.92	0.91	0.92
		10	0.97*	0.97*	0.97*
		20	0.95*	0.94*	0.94*
0.1	3	0	0.66*	0.55*	0.44*
		10	1.10	0.82	0.82
		20	0.86*	0.83*	0.81*
	4	0	0.92*	0.91*	0.88*
		10	0.99	0.98	0.99
		20	0.92*	0.85*	0.84
	5	0	1.04	0.99	0.94
		10	0.91	0.92	0.94*
		20	0.88*	0.85*	0.84*

Table 2.2: Description of modern human data sets analyzed.

Population	Description	Sample Size (n)	
		Complete Genomics	1000 Genomes
CEU	Utah residents with European ancestry	12	112
CHB	Han Chinese in Beijing, China	8	198
GIH	Gujarati Indian in Houston, Texas, USA	8	198
JPT	Japanese in Tokyo, Japan	8	200
LWK	Luhya in Webuye, Kenya	8	190
MKK	Maasai in Kinyawa, Kenya	8	0
TSI	Toscans in Italy	10	212
YRI	Yoruba in Ibadan, Nigeria	14	202

Table 2.3: Description of other data sets analyzed.

Organism	Sample Size (n)	Source
Long-tailed finch	40	(Singhal et al., 2015)
Zebra finch	40	*ibid.*
D. melanogaster	197	(Lack et al., 2015)
Ust'-Ishim (human)	2	(Q. Fu et al., 2014)

Chapter 3

A fundamental limit on the accuracy of demographic inference based on the sample frequency spectrum

In this chapter we derive a minimax lower bound on the performance of any size history estimator which operates on the site frequency spectrum. This work appears in Terhorst and Song (2015).

3.1 Motivation

The past decade has seen a revolution in our ability to interrogate the genome at the molecular level. Fueled by technological advances like next-generation sequencing, studies now routinely query hundreds or thousands of whole genomes in order to better understand disease susceptibility, heritability, population history, and other phenomena. In most cases, the conclusions of these studies come in the form of statistical estimates obtained from models which relate the effect of interest to mutation patterns arising in sampled DNA. As genetic sample sizes explode, it is natural to wonder how additional data improves the quality of these estimates. While this general question has received intense focus in theoretical statistics, certain aspects of the genetics setting (for example, non-Gaussianity and lack of independence among samples) complicate efforts to study such models using classical techniques. New methods are needed to theoretically characterize some common models in statistical genetics.

Here, we address this need for a specific estimation problem in population genetics known as *demographic inference*. As we explain in further detail below, the aim of this problem is to reconstruct the sequence of historical events—including size changes, migration, and admixture—which gave rise to present-day populations, on the basis of DNA samples obtained from those populations. We focus on the simplest problem of estimating the size history of a single population backwards in time. The main result of this chapter is to show that, for a common class of estimators which analyze a summary statistic known as the *sample frequency*

spectrum (SFS; see below), there is a fundamental limit on their accuracy as a function of the sample size. More precisely, we show that, under a standard statistical error metric known as *minimax error*, the rate at which these estimators converge to the truth for certain populations is at best inversely logarithmic in the number of independent segregating sites analyzed, and does not depend at all on the number of individuals sampled. Compared to other types of statistical estimation problems (for example, linear regression) this is an extremely slow rate of convergence. Our proof is information-theoretic in nature and applies to *any* estimator which operates solely on the sample frequency spectrum. This is the first result we are aware of which characterizes the convergence rate of demographic history estimates.

The remainder of this chapter is organized as follows. In Preliminaries we formally define our notation and model, and in Related work we describe related work on SFS-based estimators. In Main Results we state our main theoretical results, followed by a discussion of their practical implications in Discussion. To streamline our exposition, all mathematical proofs are deferred until Proofs.

3.2 Preliminaries

The stochastic process underlying the inference procedure we consider is Kingman's coalescent (Kingman, 1982c; Kingman, 1982b; Kingman, 1982a), which evolves backward in time and describes the genealogy of a collection of chromosomes randomly sampled from a population. The population size is assumed to change deterministically over time and is described by a function $\eta : [0, \infty) \to (0, \infty)$, with $\eta(t)$ being the population size at time t in the past. The instantaneous rate of coalescence between any pair of lineages at time t is $1/\eta(t)$.

As in the standard infinite-sites model of mutation (Kimura, 1969), we assume that every dimorphic site (i.e., a site with exactly two observed allelic types) has experienced mutation exactly once in the evolutionary history relating the sample. Further, for each such site, we assume that it is known which allele is the ancestral type versus the mutant type. In what follows, we use the terms dimorphic and segregating interchangeably.

A population size function $\eta(t)$ induces a probability distribution on the number of derived alleles found at a particular segregating site. Specifically, for a sample of $n \geq 2$ randomly sampled individuals, let $\xi^{(\eta)}_{n,b}$, for $1 \leq b \leq n-1$, denote the probability that a segregating site contains b mutant alleles in a sample of n individuals under model η. The vector $\xi^{(\eta)}_{n} \stackrel{\text{def}}{=} (\xi^{(\eta)}_{n,1}, \ldots, \xi^{(\eta)}_{n,n-1})$ is called the expected SFS. In the coalescent setting, a general expression for $\xi^{(\eta)}_{n,b}$ is given by R. Griffiths and Tavaré (1998):

$$\xi^{(\eta)}_{n,b} \propto \sum_{k=2}^{n-b+1} \frac{\binom{n-b-1}{k-2}}{\binom{n-1}{k-1}} \cdot k \cdot \mathbb{E}T^{(\eta)}_{n,k},$$

where $\mathbb{E}T^{(\eta)}_{n,k}$ denotes the amount of time (in coalescent units) during which the genealogy of the sample contained k lineages under model η. The expected waiting time $\mathbb{E}T^{(\eta)}_{m,m}$ to the first

coalescence in a sample of m individuals is given by

$$c_m^{(\eta)} \stackrel{\text{def}}{=} \mathbb{E}T_{m,m}^{(\eta)} = \int_0^\infty t \frac{a_m}{\eta(t)} \exp\{-a_m R_\eta(t)\}\, \mathrm{d}t, \tag{3.1}$$

where $a_m \stackrel{\text{def}}{=} \binom{m}{2}$ and $R_\eta(t) \stackrel{\text{def}}{=} \int_0^t \frac{1}{\eta(s)} \mathrm{d}s$ is the cumulative rate of coalescence up to time t. It turns out (Polanski, Bobrowski, and Kimmel, 2003) that there is an invertible linear transformation which relates $(\mathbb{E}T_{n,2}^{(\eta)}, \mathbb{E}T_{n,3}^{(\eta)}, \ldots, \mathbb{E}T_{n,n}^{(\eta)})$ to

$$\mathbf{c}^{(\eta)} \stackrel{\text{def}}{=} (c_2^{(\eta)}, c_3^{(\eta)}, \ldots, c_n^{(\eta)}). \tag{3.2}$$

Using this relation, the quantity $\xi_{n,b}^{(\eta)}$ can be written as (Polanski and Kimmel, 2003)

$$\xi_{n,b}^{(\eta)} = \frac{\left\langle \mathbf{c}^{(\eta)}, \mathbf{W}_{n,b} \right\rangle}{\left\langle \mathbf{c}^{(\eta)}, \mathbf{V}_n \right\rangle}, \tag{3.3}$$

where $\mathbf{W}_{n,b} = (W_{n,b,2}, \ldots, W_{n,b,n})$ and $\mathbf{V}_n = (V_{n,2}, \ldots, V_{n,n})$ are vectors of universal constants that do not depend on the population size function η, and $\langle \cdot, \cdot \rangle$ denotes the l_2-inner product. Under model η, the quantity $\langle \mathbf{c}^{(\eta)}, \mathbf{W}_{n,b} \rangle$ is the total expected length of edges subtending b out of n individuals sampled at time 0, while the quantity $\langle \mathbf{c}^{(\eta)}, \mathbf{V}_n \rangle$ is the total expected tree length for a sample of size n. Both quantities are positive for all population size functions η. For an arbitrary population size function η, we have $\sum_{b=1}^{n-1} W_{n,b,m} = V_{n,m}$ for all $2 \le m \le n$, which implies

$$\sum_{b=1}^{n-1} \left\langle \mathbf{c}^{(\eta)}, \mathbf{W}_{n,b} \right\rangle = \left\langle \mathbf{c}^{(\eta)}, \mathbf{V}_n \right\rangle. \tag{3.4}$$

For a constant function $\eta(t) \equiv N$,

$$c_m^{(\eta)} = \frac{N}{a_m},$$

$$\langle \mathbf{c}^{(\eta)}, \mathbf{W}_{n,b} \rangle = \frac{2}{b} N, \tag{3.5}$$

$$\langle \mathbf{c}^{(\eta)}, \mathbf{V}_n \rangle = 2NH_{n-1}, \tag{3.6}$$

where $H_{n-1} \stackrel{\text{def}}{=} \sum_{b=1}^{n-1} \frac{1}{b}$.

To formulate the problem, we employ the following notation. We suppose that a sample of $n \ge 2$ randomly sampled individuals has been typed at s unlinked segregating sites. These data are used to form the *empirical* sample frequency spectrum, which is an $(n-1)$-tuple $(\hat{\xi}_{n,1}, \ldots, \hat{\xi}_{n,n-1})$, where $\hat{\xi}_{n,b}$ denotes the proportion of segregating sites with b copies of the mutant allele and $n-b$ copies of the ancestral allele. A *frequency-based estimator* is any statistic $\hat{\eta}$ which maps an empirical SFS to a population size history.

3.3 Related work

Although the SFS is often employed in empirical studies (Nielsen, 2000; Gutenkunst et al., 2009; Coventry et al., 2010; Gazave et al., 2014; Gravel et al., 2011; Nelson et al., 2012; Excoffier et al., 2013; Bhaskar, Wang, and Song, 2015), there have been fewer attempts to understand SFS estimation from a theoretical perspective. The paper which most closely relates to the present work is by Kim *et al.* (Kim et al., 2015), who obtain lower bounds on the amount of *exact* coalescence time data necessary to distinguish between size histories in a hypothesis testing framework. Since coalescence times are never observed and must be estimated from data, these bounds place a limit on the accuracy with which a population size function can be inferred. The authors also describe an estimator which uses coalescence times (again observed without noise) to accurately recover the underlying population size function with high probability, at a rate which roughly matches the lower bound. The results presented here are complementary to those of Kim *et al.* in the following sense: access to coalescence times is equivalent to knowing the site frequency spectrum, so that their results apply to SFS-based estimators as well as other types of procedures which also incorporate e.g. linkage information. By focusing specifically on SFS-based estimators, we are able to more precisely quantify the effects of sample size and underlying demography on the ability to accurately reconstruct a population size function. Ultimately, we derive minimax lower bounds on the error of estimating a size history using the SFS.

Another line of work centers around the *identifiability* of the parameter $\eta(t)$ using the SFS. Roughly speaking, a family of statistical models $\{P_\theta\}_{\theta\in\Theta}$ defined over a parameter space Θ is identifiable if for any $\theta_1, \theta_2 \in \Theta$ with $\theta_1 \neq \theta_2$, the sampling distributions induced by P_{θ_1} and P_{θ_2} are different. In our context this simply says that, for all n, $\xi_n^{(\eta_1)} \neq \xi_n^{(\eta_2)}$ unless $\eta_1 = \eta_2$ almost everywhere. Standard desiderata for statistical estimators (e.g., consistency or unbiasedness) are impossible without identifiability, so it is the weakest possible regularity condition one can impose on a useful family of models.

Perhaps surprisingly, the sample frequency spectrum is non-identifiable in general. Indeed, for any given $\eta(t)$, it has been shown that an infinite number of smooth functions $F(t)$ exist such that $\xi_n^{(\eta)} = \xi_n^{(\eta+F)}$. Moreover, explicit examples can be constructed which demonstrate this phenomenon (Myers, Fefferman, and Patterson, 2008). On the other hand, these counter-examples consist of functions which exhibit highly oscillatory behavior, which is perhaps unrealistic when modeling naturally occurring populations. More recently, it has been shown that identifiability holds for many classes of population size functions employed by practitioners (including piecewise-constant, -exponential and -generalized-exponential). Furthermore, the number of samples n required for identifiability can be explicitly given and is a function of the complexity of the underlying class of models being studied (Bhaskar and Song, 2014).

Identifiability asserts that, given an *infinite* amount of data (specifically, taking the number of segregating sites $s \to \infty$), the model parameter $\eta(t)$ can be uniquely recovered. In practice, s is finite and only a perturbed version of the expected frequency spectrum, say $\hat{\xi}_n^{(\eta)}$, is ob-

served. From a practical standpoint, it is important to understand how these perturbations ultimately affect the parameter estimate $\hat{\eta}(t)$. It is this question that forms the starting point for the present work.

3.4 Main Results

Here, we establish a minimax lower bound on the ability of *any* estimator $\hat{\eta}$ to accurately reconstruct population size functions.

A general bound on the KL divergence between two SFS distributions

Abusing notation, we use $D(\eta \,\|\, \eta')$ to denote the Kullback-Leibler (KL) divergence between the probability distributions $\xi_n^{(\eta)}$ and $\xi_n^{(\eta')}$. In Proofs, we prove the following general upper bound on the KL divergence between two SFS distributions:

Theorem 3.1. *Let $\mathcal{M}$ denote a general space of population size functions and suppose $\eta, \eta' \in \mathcal{M}$ satisfy $\eta(t) = \eta'(t)$ for all $0 \le t \le t_c$ and $\max_{t>t_c} \eta(t) \le \min_{t>t_c} \eta'(t)$. Then,*

$$D(\eta \| \eta') \le \frac{\left\langle \mathbf{c}^{(\eta')} - \mathbf{c}^{(\eta)}, \mathbf{V}_n \right\rangle}{\left\langle \mathbf{c}^{(\eta)}, \mathbf{V}_n \right\rangle}. \tag{3.7}$$

Bounds for a family of piecewise constant models

We now focus on a particular class of population size functions which are easier to analyze and popular in the literature (Bhaskar and Song, 2014; Li and Durbin, 2011; Bhaskar, Wang, and Song, 2015). For a fixed positive integer $K > 1$, let $\mathcal{M}_K \subset \mathcal{M}$ denote the space of piecewise-constant size functions with exactly K pieces. A population size function η is a member of $\mathcal{M}_K$ if and only if there exist positive real numbers $t_1 < \cdots < t_{K-1}$ and $N_1, N_2, \ldots, N_K$ such that

$$\eta(t) = \sum_{k=1}^{K} N_k \mathbf{1}\{t_{k-1} \le t < t_k\}, \tag{3.8}$$

where by convention we define $t_0 = 0$ and $t_K = \infty$. For such an η, define

$$S_k^{(\eta)} \stackrel{\text{def}}{=} \sum_{j=1}^{k} \frac{t_j - t_{j-1}}{N_j}. \tag{3.9}$$

For $\eta \in \mathcal{M}_K$, the expected waiting time $c_m^{(\eta)}$ defined in (3.1) is given by

$$c_m^{(\eta)} = \frac{1}{a_m} \sum_{k=1}^{K} N_k (e^{-a_m S_{k-1}^{(\eta)}} - e^{-a_m S_k^{(\eta)}}). \tag{3.10}$$

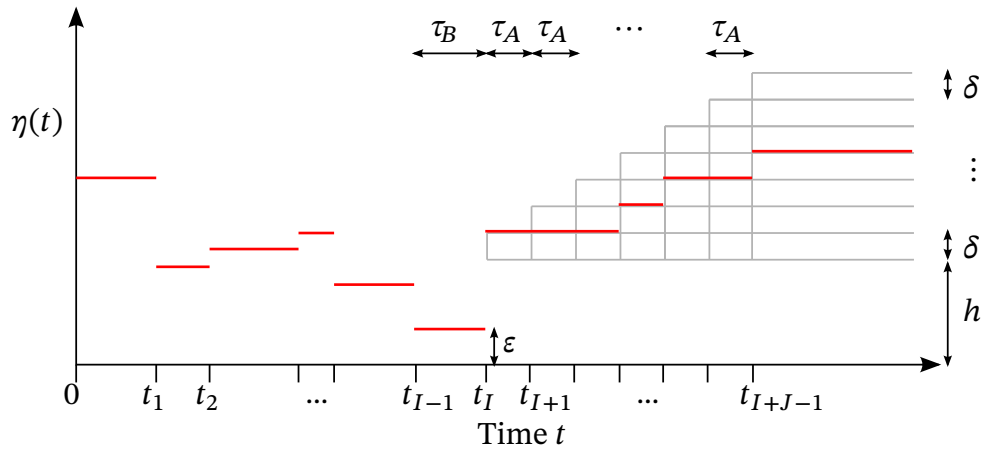


Figure 3.1: A family $\mathcal{U}_{I,J}$ of piecewise-constant models with $K = I + J$ epochs.

Note that since $t_K = \infty$,

$$e^{-a_m S_K^{(\eta)}} \equiv 0, \quad \text{for all } \eta \in \mathcal{M}_K. \tag{3.11}$$

To formulate our result, we let I, J denote positive integers which satisfy $I + J = K$, and introduce a sub-family $\mathcal{U}_{I,J} \subset \mathcal{M}_K$ of piecewise-constant functions defined as follows. See Figure 3.1 for illustration. We assume that all change points $t_1 < \cdots < t_{I+J-1}$ are fixed and that the sizes $N_1, \ldots, N_I$ of the first I epochs are also fixed, with N_I being the smallest size. So, all functions in $\mathcal{U}_{I,J}$ are identical to each other for the first I epochs, and there is a population bottleneck in the last epoch. Then, for $t \geq t_I$, every function $\eta \in \mathcal{U}_{I,J}$ undergoes jumps according to the following rules:

1. For the interval $t_I \leq t < t_{I+1}$, $\eta(t)$ takes a constant value of either h or $h + \delta$, where $h > N_I$ and $\delta > 0$.

2. At later change points $\{t_{I+1}, \ldots, t_{J-1}\}$, η either stays the same or jumps upward by δ.

Hence, $\mathcal{U}_{I,J}$ consists of 2^J distinct piecewise-constant functions that are non-decreasing functions of t for $t \geq t_I$. Note that $\min_t \eta(t) = N_I$ for all $\eta \in \mathcal{U}_{I,J}$. For ease of notation, we use $\varepsilon \stackrel{\text{def}}{=} N_I$ to denote the bottleneck size and $\tau_B \stackrel{\text{def}}{=} t_I - t_{I-1}$ to denote the bottleneck duration. To facilitate analysis later, we fix $t_{I+j} - t_{I+j-1}$ to some positive constant τ_A for all $j = 1, \ldots, J-1$. Finally, we let $\mathcal{U}_{I,J}^B = \{\eta \in \mathcal{U}_{I,J} : \|\eta\|_\infty < B\}$ be the set of all models in $\mathcal{U}_{I,J}$ which are bounded by some constant B.

For any two models in $\mathcal{U}_{I,J}$, we obtain the following bound on the difference of their waiting times to the first coalescence:

Lemma 3.2. *For all $\eta, \eta' \in \mathcal{U}_{I,J}$,*

$$\left|c_m^{(\eta)} - c_m^{(\eta')}\right| \leq J \frac{\delta}{a_m} e^{-a_m \tau_B/\varepsilon}. \tag{3.12}$$

Then, together with Theorem 3.1, this lemma can be used to show

Theorem 3.3. *Let $\eta, \eta' \in \mathcal{U}_{I,J}$ which satisfy $\max_{t \geq t_I} \eta(t) \leq \min_{t \geq t_I} \eta'(t)$. Then,*

$$D(\eta \,||\, \eta') \leq J \frac{\delta}{\varepsilon} e^{-\tau_B/\varepsilon}. \tag{3.13}$$

Our proofs of the above results are deferred to Proofs. It is interesting that the above bound does not depend on the number n of sampled individuals.

Minimax lower bound

Before using the above results to obtain a minimax lower bound, we first note a subtle fact. Given any population size function η, consider a function ζ which satisfies $\zeta(t) = \kappa \cdot \eta(t/\kappa)$ for all $t \in [0, \infty)$, where κ is some positive constant. Such functions are equivalent, as it turns out that $\xi^{(\zeta)}_{n,b} = \xi^{(\eta)}_{n,b}$ for all $n \geq 2$ and $1 \leq b \leq n-1$. To mod out by this equivalence, we assume that every $\eta \in \mathcal{M}$ satisfies $\eta(0) = N_{\text{fix}}$, where N_{fix} is some fixed positive constant.

Let $\|\cdot\|_*$ denote a generic norm (specific examples will be given later) and let $\mathbb{E}_\eta(\cdot)$ denote expectation with respect to the SFS distribution $\xi^{(\eta)}_n = (\xi^{(\eta)}_{n,1}, \ldots, \xi^{(\eta)}_{n,n-1})$ induced by population size function η. Then, note that

$$\inf_{\hat{\eta}} \sup_{\eta \in \mathcal{M}} \mathbb{E}_\eta \left\|\hat{\eta} - \eta\right\|_* \geq \inf_{\hat{\eta}} \sup_{\eta \in \mathcal{M}_K} \mathbb{E}_\eta \left\|\hat{\eta} - \eta\right\|_* \geq \inf_{\hat{\eta}} \sup_{\eta \in \mathcal{U}_{I,J}} \mathbb{E}_\eta \left\|\hat{\eta} - \eta\right\|_*.$$

In what follows, we will put a lower bound on the last quantity. We first fix a sensible distance metric on $\mathcal{M}$. An intuitive way to measure distance between two population size functions is their L_1 distance, $\|\eta_a - \eta_b\|_1 = \int_0^\infty |\eta_a(t) - \eta_b(t)| \mathrm{d}t$, but this is unreasonably stringent in that $\left\|\eta_a - \eta_b\right\|_1 = \infty$ if η_a and η_b do not agree infinitely far back into the past. Instead we will focus on the following truncated L_1 distance: $\|\eta_a - \eta_b\|_{1,T} \stackrel{\text{def}}{=} \int_0^T |\eta_a(t) - \eta_b(t)| \mathrm{d}t$, which measures the discrepancy between η_a and η_b back to some fixed time T in the past.

In Proofs, we obtain the following main result of this chapter:

Theorem 3.4. *Suppose $J > 8$ and $T \geq t_{I+J-1} + \tau_A$, and let $\hat{\eta}$ be any estimator of the population size function which operates on a sample of s unlinked segregating sites obtained from a sample of n randomly sampled individuals. Then for all s,*

$$\inf_{\hat{\eta}} \sup_{\eta \in \mathcal{U}^B_{I,J}} \mathbb{E}_\eta \left\|\hat{\eta} - \eta\right\|_{1,T} \geq C \frac{(J-8)^2}{J} \frac{\tau_B \tau_A}{\log s}, \tag{3.14}$$

where C is a positive constant.

By specializing $\mathcal{U}_{I,J}$, a simplified form Theorem 3.4 can be obtained.

Corollary 3.5. *Let* $\mathcal{U}^B_{I,\star} = \bigcup_{J\geq 1} \mathcal{U}^B_{I,J}$. *Then*

$$\inf_{\hat{\eta}} \sup_{\eta\in\mathcal{U}^B_{I,\star}} \mathbb{E}_\eta \left\|\hat{\eta} - \eta\right\|_{1,T} \geq C'(T - t_I)\frac{\tau_B}{\log s}, \tag{3.15}$$

where C' *is a positive constant.*

The preceding results are obtained by setting $\varepsilon = \tau_B/\log s$ and $\delta \propto \varepsilon$, so that in particular $\varepsilon \to 0$ as $s \to \infty$. In some scenarios in may be undesirable to let the bottleneck size ε tend to zero, though we argue in the next section that this rate is slow enough to make the result apply to many natural populations. Also, in nature the number of segregating sites s is bounded, so a result which is non-asymptotic in s may be of interest. We record such a result below.

Theorem 3.6. *Consider the sub-family* $\mathcal{U}_{I,J}$ *of models described above, and suppose* $J > 8$ *and* $T \geq t_{I+J-1}+\tau_A$. *Let* $\hat{\eta}$ *be any estimator of the population size function which operates on a sample of* s *unlinked segregating sites obtained from a sample of* n *randomly sampled individuals. Then,*

$$\inf_{\hat{\eta}} \sup_{\eta\in\mathcal{U}_{I,J}} \mathbb{E}_\eta \left\|\hat{\eta} - \eta\right\|_{1,T} \geq C''\tau_A\frac{(J-8)^2}{J}\frac{\varepsilon}{s}e^{\tau_B/\varepsilon}, \tag{3.16}$$

where C'' *is a positive constant.*

Bottleneck followed by exponential growth

A slight modification of our above results permits us to analyze the following model class, which is of interest in, for example, human genetics: let

$$\mathcal{E}_J \stackrel{\text{def}}{=} \left\{\eta_0 e^{-\beta(\eta_0)t}\mathbf{1}_{\{t\leq t_1\}} + u(t-t_1)\mathbf{1}_{\{t>t_1\}} : u \in \mathcal{U}_{1,J}\right\}.$$

Here, $\beta(\eta_0) \stackrel{\text{def}}{=} \log(\frac{\eta_0}{\gamma\varepsilon})/t_1$ is a rate-of-decay function defined so that $\eta(t_1) = \gamma\varepsilon$ for all $\eta \in \mathcal{E}_J$, and $\gamma \geq 1$ is a multiplier. The class $\mathcal{E}_J$ consists of size histories which are piecewise constant before the bottleneck, and then jump to some level $\gamma\varepsilon$ and undergo (identical) exponential growth from time t_1 to present.

Theorem 3.7. *Consider the sub-family* $\mathcal{E}_J$ *of models described above, and suppose* $J > 8$ *and* $T \geq t_J + \tau_A$. *Let* $\hat{\eta}$ *be any estimator of the population size function which operates on a sample of* s *unlinked segregating sites obtained from a sample of* n *randomly sampled individuals. Then,*

$$\inf_{\hat{\eta}} \sup_{\eta\in\mathcal{E}_J} \mathbb{E}_\eta \left\|\hat{\eta} - \eta\right\|_{1,T} \geq C''\tau_A\frac{(J-8)^2}{J}\frac{\varepsilon}{s}\exp\left\{\tau_B/\varepsilon + t_1\frac{1/\eta_0 - 1/\gamma\varepsilon}{\log(\gamma\varepsilon) - \log(\eta_0)}\right\}. \tag{3.17}$$

Theorem 3.6 is a measure of how (a lower bound on) estimation error improves due to growth following a bottleneck. The two extremes and $\eta_0 \to \infty$ and $\eta_0 \to \gamma\varepsilon$ have intuitive interpretations. For large η_0 the bound in (3.17) tends to the corresponding bound given by

Theorem 3.6, as expected since coalescences become increasingly less likely in the first time period. Similarly, small η_0 has the effect of "prolonging" the bottleneck; in particular, if $\gamma = 1$ then $t_1 \frac{1/\eta_0 - 1/\gamma\varepsilon}{\log(\gamma\varepsilon) - \log(\eta_0)} \to \frac{t_1}{\varepsilon}$, so that the minimax estimation effect of low population growth is to simply prolong the bottleneck effect by an additional t_1 time periods.

3.5 Discussion

We have shown that the minimax error rate for estimating the piecewise-constant demography of a single population is at least $O(1/\log s)$, where s is the number of independent segregating sites analyzed. In contrast, the minimax error for many classical estimation problems in statistics (for example, non-parametric regression or density estimation) decays inverse-polynomially in the sample size (Tsybakov, 2009). Compared with these problems, exponentially more samples would be required to estimate a population size history function to within a similar magnitude of error.

As is standard procedure when obtaining this type of bound, our result relies on a specifically constructed class of inference problems for which estimation is difficult. As such, it represents a "worst-case" type of analysis. In general, better (less pessimistic) rates may be achievable in instances where additional structure can be placed on the problem. However, in situations where no nothing is known *a priori* concerning the population size history function to be estimated, our results establish the existence of a class of size history functions which are fairly distinct while being statistically indistinguishable from one another. For users who find themselves in such a setting, this should serve as a cautionary note.

A single population evolving under a piecewise-constant demography is a special case of many richer classes of demographic models. For example, it is a (limiting) member of the family of exponential growth models, seen by taking each exponential growth parameter to zero. In the multi-species coalescent setting (Chen, 2012; Excoffier et al., 2013), multiple population size histories must be estimated, and the error of that estimate must necessarily be lower bounded by that of estimating a single such history. Thus, our result can be expected to apply to a broader class of models than the one we have studied here.

As detailed in Proofs, the result in Theorem 3.4 follows from setting $\varepsilon = \tau_B/\log s$ and $\delta \propto \varepsilon$ in the sub-family $\mathcal{U}_{I,J}$. The size $\tau_B/\log s$ is in coalescent units. In terms of the number of individuals, it is proportional to $g_B/\log s$, where g_B is the number of generations corresponding to duration τ_B in the coalescent limit. Intuitively, as the severity of the bottleneck increases, the population is increasingly likely to find its most recent common ancestor (MRCA) during that time; further back in time than the MRCA, no information is conveyed concerning the demographic events experienced by the population.

One might object to considering models with a bottleneck size that scales inversely with the number s of segregating sites in the data, and it is indeed possible that a better convergence rate may be achievable for populations which are known not to contain a bottleneck. On the other hand, we note that $1/\log s$ decreases sufficiently slowly with s that our result can be expected

to apply to many real-world examples. For example, for $s \approx 10^8$, which is a conservative upper bound for most organisms, $g_B/\log s \approx 0.054 g_B$. This implies that for populations which have experienced roughly an order-of-magnitude increase in effective population size during their history, accurate estimation of demographic events which occurred before this expansion is difficult using SFS-based methods. Additionally, an interesting aspect of our result is that the lower bound (3.14) does not depend on the number of sampled individuals n; increasing the sample size is not enough to overcome the information barrier imposed by the presence of a bottleneck. This is intuitively plausible since, as n increases, the $(n+1)$-th sampled lineage becomes more likely to coalesce early on.

An interesting question which we have not attempted to analyze is whether the $O(1/\log s)$ rate is optimal, i.e. whether there exists some estimator $\hat{\eta}(t)$ which achieves the minimax lower bound established here. In practice, from equations (3.3), (3.9), and (3.10), it can be seen that naively maximizing the likelihood of the observed SFS with respect to $\eta(t)$ requires solving a non-convex optimization problem, so that convergence to the global maximum is not even guaranteed. Computational issues aside, finding such an estimator remains an open theoretical challenge.

In closing, we stress that our result is specific to SFS-based estimators, which analyze only independent unlinked sites. The main allure of these estimators is their mathematical tractability, rather than their realism. In fact, a rich source of additional information exists in the correlation structure found among linked sites in the genome. Methods which seek to exploit this structure by modeling the action of recombination and gene conversion pose greater mathematical and computational difficulties, but there has been recent progress in this area (Li and Durbin, 2011; Sheehan, Harris, and Song, 2013; M. D. Rasmussen et al., 2014). Our result serves to underscore the importance of pursuing ever more realistic models of genomic evolution, challenging though they may be.

3.6 Proofs

Proof of Theorem 3.1. To simplify the notation, we write $\mathbf{c} = \mathbf{c}^{(\eta)}$ and $\mathbf{c}' = \mathbf{c}^{(\eta')}$. Then, using (3.3) we can write

$$D(\eta \| \eta') = \sum_{b=1}^{n-1} \xi_{n,b}^{(\eta)} \log \frac{\xi_{n,b}^{(\eta)}}{\xi_{n,b}^{(\eta')}} = \sum_{b=1}^{n-1} \xi_{n,b}^{(\eta)} \left[\log\left(\frac{\langle \mathbf{c}, \mathbf{W}_{n,b} \rangle}{\langle \mathbf{c}', \mathbf{W}_{n,b} \rangle} \right) + \log\left(\frac{\langle \mathbf{c}', \mathbf{V}_n \rangle}{\langle \mathbf{c}, \mathbf{V}_n \rangle} \right) \right].$$

The assumption $\min_{t>t_c} \eta'(t) \geq \max_{t>t_c} \eta(t)$ implies that, for all times $t, t' > t_c$, the instantaneous rate of coalescence at time t in model η is $\geq$ the instantaneous rate of coalescence at time t' in model η'. Hence, this assumption together with $\eta(t) = \eta'(t)$ for all $0 \leq t \leq t_c$ implies $\langle \mathbf{c} - \mathbf{c}', \mathbf{W}_{n,b} \rangle \leq 0$ for all $1 \leq b \leq n-1$; equivalently, $\log\left(\frac{\langle \mathbf{c}, \mathbf{W}_{n,b} \rangle}{\langle \mathbf{c}', \mathbf{W}_{n,b} \rangle} \right) < 0$. Additionally,

$\frac{\langle \mathbf{c}'-\mathbf{c},\mathbf{V}_n\rangle}{\langle \mathbf{c},\mathbf{V}_n\rangle} > -1$ and $\log(1+x) \le x$ for all $x \ge -1$. Combining these facts, we obtain

$$D(\eta\|\eta') \le \sum_{b=1}^{n-1} \xi_{n,b}^{(\eta)} \log\left(\frac{\langle \mathbf{c}', \mathbf{V}_n\rangle}{\langle \mathbf{c}, \mathbf{V}_n\rangle}\right) \le \sum_{b=1}^{n-1} \xi_{n,b}^{(\eta)} \frac{\langle \mathbf{c}'-\mathbf{c}, \mathbf{V}_n\rangle}{\langle \mathbf{c}, \mathbf{V}_n\rangle} = \frac{\langle \mathbf{c}'-\mathbf{c}, \mathbf{V}_n\rangle}{\langle \mathbf{c}, \mathbf{V}_n\rangle},$$

where we have used $\sum_{b=1}^{n-1} \xi_{n,b}^{(\eta)} = 1$ in the final equality. □

Proof of Lemma 3.2. We distinguish two particular models $\eta_\ell, \eta_u \in \mathcal{U}_{I,J}$ which are the *lower* and the *upper* envelopes of $\mathcal{U}_{I,J}$. The function η_ℓ stays constant at h for all $t \ge t_I$, while η_u jumps upward by δ at every change point $t_I, \dots, t_{I+J-1}$. Hence, $\eta_\ell \le \eta \le \eta_u$ pointwise for all $\eta \in \mathcal{U}_{I,J}$. The two enveloping functions will form the basis of subsequent analysis.

Fix η and η' and assume without loss of generality that $\eta(t) \le \eta'(t)$ for all t. Then for all t,

$$\eta_\ell(t) \le \eta(t) \le \eta'(t) \le \eta_u(t),$$

which implies

$$c_m^{(\eta_\ell)} \le c_m^{(\eta)} \le c_m^{(\eta')} \le c_m^{(\eta_u)},$$

for all $m = 2, \dots, n$. Using these inequalities, we conclude

$$c_m^{(\eta')} - c_m^{(\eta)} \le c_m^{(\eta_u)} - c_m^{(\eta_\ell)},$$

so it suffices to demonstrate (3.12) for $c_m^{(\eta_u)} - c_m^{(\eta_\ell)}$. Now, by equation (3.10) and the definition of η_ℓ,

$$\begin{aligned} a_m c_m^{(\eta_\ell)} &= \sum_{i=1}^{I} N_i \left[e^{-a_m S_{i-1}^{(\eta_\ell)}} - e^{-a_m S_i^{(\eta_\ell)}}\right] + \sum_{j=1}^{J} h \left[e^{-a_m S_{I+j-1}^{(\eta_\ell)}} - e^{-a_m S_{I+j}^{(\eta_\ell)}}\right] \\ &= \sum_{i=1}^{I} N_i \left[e^{-a_m S_{i-1}^{(\eta_\ell)}} - e^{-a_m S_i^{(\eta_\ell)}}\right] + h e^{-a_m S_I^{(\eta_\ell)}}, \end{aligned}$$

where we have used equation (3.11). Similarly,

$$\begin{aligned} a_m c_m^{(\eta_u)} &= \sum_{i=1}^{I} N_i \left[e^{-a_m S_{i-1}^{(\eta_u)}} - e^{-a_m S_i^{(\eta_u)}}\right] + \sum_{j=1}^{J} (h + j\delta) \left[e^{-a_m S_{I+j-1}^{(\eta_u)}} - e^{-a_m S_{I+j}^{(\eta_u)}}\right] \\ &= \sum_{i=1}^{I} N_i \left[e^{-a_m S_{i-1}^{(\eta_u)}} - e^{-a_m S_i^{(\eta_u)}}\right] + h e^{-a_m S_I^{(\eta_u)}} + \sum_{j=1}^{J} j\delta \left[e^{-a_m S_{I+j-1}^{(\eta_u)}} - e^{-a_m S_{I+j}^{(\eta_u)}}\right]. \end{aligned}$$

Now, using the fact that η_ℓ and η_u agree on the first I epochs, we obtain

$$\begin{aligned} a_m[c_m^{(\eta_u)} - c_m^{(\eta_\ell)}] &= \sum_{j=1}^{J} j\delta \left[e^{-a_m S_{I+j-1}^{(\eta_u)}} - e^{-a_m S_{I+j}^{(\eta_u)}}\right] \\ &= \delta \sum_{j=1}^{J} e^{-a_m S_{I+j-1}^{(\eta_u)}} \le J\delta e^{-a_m \tau_B/\varepsilon}, \end{aligned} \tag{3.18}$$

where the inequality follows from the fact that $\frac{\tau_B}{\varepsilon} \leq S^{(\eta_u)}_{I+j-1}$ for all $j = 1, \ldots, J$. □

Proof of Theorem 3.3. For ease of notation, define $\mathbf{c} = \mathbf{c}^{(\eta)}$ and $\mathbf{c}' = \mathbf{c}^{(\eta')}$. By Lemma 3.2,

$$\begin{aligned}\langle \mathbf{c}' - \mathbf{c}, \mathbf{V}_n \rangle = \sum_{m=2}^{n} (c'_m - c_m) V_{n,m} &\leq J\delta \sum_{m=2}^{n} \frac{V_{n,m}}{a_m} e^{-a_m \tau_B/\varepsilon} \\ &\leq J\delta e^{-\tau_B/\varepsilon} \sum_{m=2}^{n} \frac{V_{n,m}}{a_m},\end{aligned}$$

where the second inequality follows from $e^{-a_m\tau_B/\varepsilon} \leq e^{-\tau_B/\varepsilon}$ for all $m = 2, \ldots, n$. Now, noting that $\sum_{m=2}^{n} \frac{V_{n,m}}{a_m}$ corresponds to the total tree length for the constant population size function $\eta \equiv 1$ and using (3.6), we obtain

$$\langle \mathbf{c}' - \mathbf{c}, \mathbf{V}_n \rangle \leq J\delta e^{-\tau_B/\varepsilon} 2H_{n-1}. \tag{3.19}$$

To finish the proof, recall that $\langle \mathbf{c}, \mathbf{V}_n \rangle$ is the total expected branch length of the coalescent tree under model η. Since $\min_t \eta(t) = \varepsilon$, we have that $\langle \mathbf{c}, \mathbf{V}_n \rangle$ is at least as large as the corresponding quantity under a model with constant population size ε. By (3.6), the total expected tree length under the latter model equals $2\varepsilon H_{n-1}$. Thus, $\langle \mathbf{c}, \mathbf{V}_n \rangle \geq 2\varepsilon H_{n-1}$ and combining this result with (3.19) gives

$$\frac{\langle \mathbf{c}' - \mathbf{c}, \mathbf{V}_n \rangle}{\langle \mathbf{c}, \mathbf{V}_n \rangle} \leq J \frac{\delta}{\varepsilon} e^{-\tau_B/\varepsilon}.$$

Finally, (3.13) follows from this inequality and Theorem 3.1. □

Proof of Theorem 3.6. Our proof uses a generalized form of Fano's inequality (B. Yu, 1997). Adapted to our setting and notation, the method reads as follows.

Theorem 3.8 (Fano's method). *Consider a space $\mathcal{M}$ of population size models. Let $r \geq 2$ be an integer, and let $S^r_n = \{\eta_1, \eta_2, \ldots, \eta_r\} \subset \mathcal{M}$ contain r population size functions such that for all $a \neq b$, $\|\eta_a - \eta_b\|_* \geq \alpha_r$ and $D(\xi^{(\eta_a)}_n \| \xi^{(\eta_b)}_n) \leq \beta_r$. Let $\hat{\eta}^{(n,s)} = \hat{\eta}^{(n,s)}(X_1, \ldots, X_s)$ be an estimator of η based on the SFS data $X_1, \ldots, X_s$ sampled independently from $\xi^{(\eta)}_n$; i.e., $X_1, \ldots, X_s$ are SFS data for n individuals at s unlinked segregating sites. Then,*

$$\inf_{\hat{\eta}} \sup_{\eta \in \mathcal{M}} \mathbb{E}_\eta \|\hat{\eta}^{(n,s)} - \eta\|_* \geq \frac{\alpha_r}{2}\Big(1 - \frac{s \cdot \beta_r + \log 2}{\log r}\Big). \tag{3.20}$$

This theorem places a lower bound on the minimax rate of convergence of a population size history estimator based on the SFS.

For $\eta \in \mathcal{U}_{I,J}$, let w_j denote the variable $\in \{0, 1\}$ indicating whether η jumps by δ at change point t_{I+j}. Let $\mathcal{Y} = \{\mathbf{w} = (w_0, \ldots, w_{J-1}) \mid w_i \in \{0, 1\}\}$, where $J \geq 8$. By the Varshamov-Gilbert lemma (see Massart, 2007, Lemma 4.7), there exist $\mathcal{X} = \{\mathbf{w}^0, \ldots, \mathbf{w}^M\} \subset \mathcal{Y}$ such (i)

$\mathbf{w}^0 = (0, \dots, 0)$, (ii) $M \geq 2^{J/8}$, and (iii) $H(\mathbf{w}^i, \mathbf{w}^j) \geq J/8$, where $H(\cdot, \cdot)$ denotes the Hamming distance.

Let $\mathcal{U}_{I,J}^{\mathcal{X}}$ denote the subset of $2^{J/8} + 1$ functions in $\mathcal{U}_{I,J}$ with the indicator variable for δ-jumps at $t_I, \dots, t_{I+J-1}$ given by $\mathbf{w} \in \mathcal{X}$. Then, for any two $\eta_a \neq \eta_b \in \mathcal{U}_{I,J}^{\mathcal{X}}$, we have

$$\|\eta_a - \eta_b\|_{1,T} \geq \frac{J}{8} \cdot \tau_A \cdot \delta. \tag{3.21}$$

Using Theorem 3.8 via (3.21) and Theorem 3.3, we obtain

$$\begin{aligned} \inf_{\hat{\eta}} \sup_{\eta \in \mathcal{U}_{I,J}} \mathbb{E}_\eta \|\hat{\eta}^{(n,s)} - \eta\|_{1,T} &\geq \frac{J \cdot \tau_A \cdot \delta}{16} \left[1 - \frac{sJ\frac{\delta}{\varepsilon} e^{-\tau_B/\varepsilon} + \log 2}{\log(2^{J/8} + 1)} \right] \\ &\geq \frac{J \cdot \tau_A \cdot \delta}{16} \left[1 - \frac{sJ\frac{\delta}{\varepsilon} e^{-\tau_B/\varepsilon} + \log 2}{\frac{J}{8} \log 2} \right]. \end{aligned} \tag{3.22}$$

We now optimize the bound with respect to δ. A straightforward calculation shows that the maximum is attained at

$$\delta^* = \frac{(J-8)\log 2}{16J} \frac{\varepsilon}{s} e^{\tau_B/\varepsilon} \tag{3.23}$$

and setting $\delta = \delta^*$ in (3.22) yields the result. □

Proof of Theorem 3.4. The result is obtained by scaling ε with the number of segregating sites s. Denote this scaling by $\varepsilon(s)$; we will determine $\varepsilon(s)$ which produces the largest possible lower bound. Starting from Equation (3.23) in the proof of Theorem 3.6, note that δ^* scales as $\frac{\varepsilon}{s} e^{\tau_B/\varepsilon} =: f(\varepsilon)$. In order to satisfy the constraint that $\|\eta\|_\infty < B$ for all $\eta \in \mathcal{U}_{I,J}^B$ and s, the condition

$$\limsup_{s \to \infty} \max\left\{ \frac{\varepsilon(s)}{s} e^{\tau_B/\varepsilon(s)}, \varepsilon(s) \right\} < \infty \tag{3.24}$$

must therefore hold. This implies that $\varepsilon(s) s^\beta \to \infty$ for all $\beta > 0$. Suppose that $\alpha = \liminf_s \varepsilon(s) \log s < \tau_B$; note that $\varepsilon(s) > 0 \implies \alpha > 0$. Then there exists a diverging sequence $s_1, s_2, \dots \to \infty$ with $\log(s_i) < \frac{1+\alpha}{2} \frac{\tau_B}{\varepsilon(s_i)}$ for all i, whence

$$\limsup_{s \to \infty} \frac{\varepsilon(s)}{s} e^{\tau_B/\varepsilon(s)} \geq \limsup_{i \to \infty} \frac{\varepsilon(s_i)}{s_i} e^{\frac{2}{1+\alpha} \log(s_i)} = \limsup_{i \to \infty} \varepsilon(s_i) s_i^{\frac{1-\alpha}{1+\alpha}} = \infty.$$

From this it follows that $\varepsilon(s) \geq \tau_B / \log s$ for sufficiently large s. Now, on the interval $(0, \infty)$ the function $f(\varepsilon)$ is convex with a unique minimum at $\varepsilon = \tau_B$. Let ε' be a point where $f(\varepsilon') > f(\tau_B / \log s) = \tau_B / \log s$. Then $\varepsilon' \notin [\tau_B / \log s, \tau_B]$. If $\varepsilon' > \tau_B$ then $f(\varepsilon') < \frac{\varepsilon'}{s} e$, which implies that $\varepsilon > \frac{s\tau_B}{e \log s}$.

In summary, we see the largest possible lower bound which obeys (3.24) must have $f(\varepsilon)$ asymptotically $\leq \tau_B / \log s$, and that this bound is achieved by setting $\varepsilon(s) = \tau_B / \log s$. Plugging this in to equation (3.20) yields the claim. □

Proof of Corollary 3.5. Let $c \in (0,1)$ and pick J large enough that $(J-8)(\log 2)/J > c$. Fix τ_A so that $T = t_I + J\tau_A$. Then $(J-8)(\log 2)\tau_A \geq cJ\tau_A = c(T - t_I)$ and $(J-8)(\log 2)/J \geq c$. Letting $C' = Cc^2$ and substituting these inequalities into equation (3.14) yields the claim. □

Proof of Theorem 3.7. The theorem is obtained by suitably modifying the preceding results to account for the effect of exponential growth in the first period. Let η'_u, η'_ℓ be the analogously-defined upper and lower envelope functions for $\mathcal{E}_J$. Then

$$\begin{aligned}\int_0^{t_{j+1}} \frac{ds}{\eta'_u(s)} &= \frac{e^{\beta(\eta_0)t_1} - 1}{\beta(\eta_0)\eta_0} + S_j^{(\eta_u)} \\ &= t_1 \frac{1/\eta_0 - 1/\gamma\varepsilon}{\log(\gamma\varepsilon) - \log(\eta_0)} + S_j^{(\eta_u)},\end{aligned}$$

where we have used the definition of $\beta(\eta_0)$ in the second inequality. Since all size histories in $\mathcal{E}_J$ are equal up to period t_2, the steps of Lemma 3.2 all go through unchanged. Starting from equation (3.18), we obtain the modified bound

$$a_m[c_m^{(\eta'_u)} - c_m^{(\eta'_\ell)}] \leq J\delta \exp\Big\{ - a_m t_1 \frac{1/\eta_0 - 1/\gamma\varepsilon}{\log(\gamma\varepsilon) - \log(\eta_0)} \Big\} e^{-a_m \tau_B/\varepsilon}. \tag{3.25}$$

Propagating the modified bound (3.25) through Theorems 3.3 and 3.6 ultimately yields the claim. □

Chapter 4

Demographic inference: a nonparametric perspective

In this chapter we continue our study of the demographic inference problem for large samples using ideas from the theory of nonparametric statistical estimation. As in the previous chapter, we focus on inferring the size history of a population from the site frequency spectrum. Under the additional assumption that the per-generation mutation rate is known, we show how the size history function is approximately related to the SFS by a certain linear operator. Thus, in this case demographic inference reduces to the problem of inverting this operator, which is a well studied problem.

4.1 The site frequency spectrum: diffusion argument

In this section we derive the expected site frequency spectrum ξ[1] conditional on the demography $\eta(t)$. We have already shown in Chapter 3 that it can be obtained using coalescent theory by multiplying the vector of expected first coalescence times, $\mathbf{c}^{(\eta)}$ (3.2) by a certain combinatorial matrix $\mathbf{W}$ (see equation (3.3) and surrounding discussion). That matrix is defined by a system of recursions first discovered by Polanski and Kimmel (2003). Our goal will eventually be to bound the entries of this matrix, for which the recursive definition is somewhat ill-suited. In this section we will show how the same matrix can be obtained analytically, i.e. its entries are given by a certain explicit integral which we will define below. In order to do so we first review the diffusion-theoretic derivation of the likelihood of the site frequency spectrum. These results are well known in the literature and are provided for completeness. The exposition closely follows Myers, Fefferman, and Patterson (2008).

We will derive the likelihood of observing $b = 1, \dots, n-1$ mutant alleles in a random sample of n individuals from a large (effectively infinite) population. A population-genetic prior density $\mathbb{P}_\eta(\mathrm{d}x)$ is placed on the unknown frequency of the mutant allele in the overall popu-

[1] To simplify the notation in the chapter, we suppress dependence on the sample size n throughout.

lation, so that

$$\mathbb{P}_\eta(B = b) = \int_0^1 \binom{n}{b} x^b (1-x)^{n-b}\, \mathbb{P}_\eta(\mathrm{d}x). \tag{4.1}$$

To derive the prior $\mathbb{P}_\eta(\mathrm{d}x)$ we consider the classical *Kimura diffusion* (Kimura, 1955), which characterizes the trajectory of a neutral mutation in a panmictic population of (large) constant size. Kimura showed that in a population of N diploid individuals, the probability $K(x, y; \tau)$ that the mutation evolves from frequency x to y in $2N\tau$ generations is, for large N, approximately

$$K(x, y; \tau) = x(1-x) \sum_{i=0}^{\infty} J_i(x) J_i(y) e^{-\lambda_i \tau}. \tag{4.2}$$

Here, $\lambda_i \stackrel{\text{def}}{=} \binom{i}{2}$ and $J_i(x)$ denotes the Jacobi-$(1,1)$ polynomial of degree i (see Section 4.6) rescaled and shifted to the unit interval:

$$J_i(x) \stackrel{\text{def}}{=} (-1)^i (i+2)(2i+3) P_i^{(1,1)}(2x-1). \tag{4.3}$$

Expanding (4.2) about $x = 0$, we obtain

$$\begin{aligned} K(\epsilon, y; \tau) &= \underbrace{\epsilon \sum_{i=0}^{\infty} J_i(0) J_i(y) e^{-\lambda_i \tau}}_{Q(y;\tau)} + O(\epsilon^2) \\ &= \epsilon Q(y; \tau) + O(\epsilon^2). \end{aligned} \tag{4.4}$$

Thus, if mutations occur in the population at a constant rate per unit time and, conditional on occurrence, have frequency $1/2N$, the probability that a mutation has frequency y in the present is

$$\mathbb{P}(y) \propto \int_0^\infty 2NK(1/2N, y; \tau)\, \mathrm{d}\tau = \int_0^\infty Q(y; \tau)\, \mathrm{d}\tau + O(1/N). \tag{4.5}$$

Variable population size

The above argument holds for constant population size. To generalize to the case of variable population size, we rely on the following time-change argument again due to Kimura (1955). Let $\tau(t) = \int_0^t 1/(2\eta(s))\mathrm{d}s$, which is monotone increasing and hence invertible. Define $\tilde{\eta}(\tau)$ implicitly by

$$\tilde{\eta}(\tau(t)) = \eta(t). \tag{4.6}$$

Thus if $t(\tau) = \int_0^\tau 2\eta(t)\,\mathrm{d}t$, then $\eta(t(\tau)) = \tilde{\eta}(\tau)$. Kimura showed that the transition density under the variable size model η, say $K_\eta(x, y; t) = K(x, y; \tau(t))$ in the limit as $N \to \infty$, where $K(x, y, \tau)$ is the transition density under constant size. Modifying equation (4.5) to account for

the variable size, we obtain by change of variables $\tau(t) \to \tau$,

$$\begin{aligned}\mathbb{P}_\eta(y) &\propto \int_0^\infty 2\eta(t)K_\eta(1/(2\eta(t)), y; t)\,\mathrm{d}t \\ &= \int_0^\infty 2\tilde{\eta}(\tau(t))K(1/(2\tilde{\eta}(\tau(t))), y; \tau(t))\,\mathrm{d}t \\ &= \int_0^\infty 2\eta(t(\tau))\tilde{\eta}(\tau)K(1/(2\tilde{\eta}(\tau)), y; \tau)\,\mathrm{d}\tau \\ &\approx \int_0^\infty \tilde{\eta}(\tau)\,Q(y,\tau)\,\mathrm{d}\tau. \end{aligned} \tag{4.7}$$

Combining (4.4) and (4.7), we obtain

$$\mathbb{P}_\eta(y) \propto \sum_{i=0}^\infty J_i(y) \int_0^\infty \tilde{\eta}(\tau)e^{-\lambda_i\tau}\,\mathrm{d}\tau \stackrel{\text{def}}{=} \sum_{i=0}^\infty c_i^{(\eta)} J_i(y)$$

so that, by (4.1),

$$\mathbb{P}_\eta(b) \propto \sum_{i=0}^\infty c_i^{(\eta)} \int_0^1 \binom{n}{b} y^b(1-y)^{n-b} J_i(y)\,\mathrm{d}y. \tag{4.8}$$

Since

$$\int y^b(1-y)^{n-b}J_i(y)\,\mathrm{d}y = \int y^{b-1}(1-y)^{n-b-1}y(1-y)J_i(y)\,\mathrm{d}y,$$

the integral represents the coordinates of the $(n-2)$ degree polynomial $y^{b-1}(1-y)^{n-b-1}$ in the Jacobi-$(1,1)$ basis. Hence, it is zero for $i > n-2$. Also, through a change of variables and integration by parts, the numbers $c_i^{(\eta)}$ have an alternative expression given by (3.10), which shows that they are the first coalescence times under the model η.

Comparing expressions (4.8) and (3.3), we see that the entries of the matrix $\mathbf{W}$ are precisely

$$W_{bj} = \int_0^1 \binom{n}{b} x^b(1-x)^{n-b} J_{j-2}(x)\,\mathrm{d}x. \tag{4.9}$$

(Our convention is to index j from 2 to n so that W_{bj} multiplies the first coalescence time from that number of individuals.)

4.2 Approximate likelihood of the SFS

Equation (4.8) shows that the likelihood of observing b mutants in a sample of n individuals is

$$\mathbb{P}_\eta(b) \propto (WER\eta)_b, \tag{4.10}$$

where $W : \mathbb{R}^{(n-1)} \to \mathbb{R}^{(n-1)}$ is a linear transformation given by the matrix $\mathbf{W}$[2] (4.9), $E : L^2(0, \infty, e^{-x}) \to \mathbb{R}^{n-1}$ is an operator defined by

$$\tilde{\eta} \mapsto \int_0^\infty \tilde{\eta}(t) e^{-\lambda_i t} \, \mathrm{d}t, \quad i = 2, \ldots, n,$$

and $R\eta = \tilde{\eta}$ is the time change map defined by equation (4.6). The operator is linear in $\tilde{\eta}$ but nonlinear in η due to the action of R. Since R is well conditioned and easily inverted, for the rest of this analysis we focus estimating $\tilde{\eta}$, from which the estimate $\hat{\eta} = R^{-1}\left(\hat{\tilde{\eta}}\right)$ can easily be formed.

Equation (4.10), after normalization, gives the expected normalized SFS. Let us now consider the empirical unnormalized SFS, denoted by $\hat{\upsilon}$. Define the population-scaled mutation rate $\theta = 2N_0\mu$, where N_0 is the reference population size. Under a Poisson mutation model, the probability, conditional on the coalescent tree at a particular site, that a mutation has k descendents is $\theta S_{n,k} + O(\theta S_{n,k})^2$, where $S_{n,k}$ denotes the total branch length (in coalescent units) subtending k individuals in a sample of size n (cf. Section 3.2).

From the preceding chapter we recognize that the k-th entry of $WE\tilde{\eta}$ is precisely $\mathbb{E}_\eta S_{n,k}$. Thus, if we assume that $\theta S_{n,k} \ll 1$ and the number G of independent sites in the genome is large, then

$$\hat{\upsilon} = \upsilon + \Xi$$

where $\upsilon = \theta WE\tilde{\eta}$ and the noise term Ξ has approximately a Gaussian distribution with zero mean and covariance

$$\Sigma = G(\mathrm{diag}(\upsilon)(I - \mathrm{diag}(\upsilon)) - \upsilon\upsilon^T) \approx G\,\mathrm{diag}(\upsilon), \tag{4.11}$$

where we have dropped terms of order $O(\theta^2)$. The covariance depends on the unknown parameter $\tilde{\eta}$, but if η is bounded, say $B^{-1} \le \eta \le B$, then by the argument stated in the proof of Lemma 3.2 in Chapter 3,

$$\frac{2}{Bk} \le \mathbb{E}_\eta S_{n,k} \le \frac{2B}{k}.$$

Defining $L = 2\theta\,\mathrm{diag}(1/2, 1/3, \ldots, 1/n)$, we have

$$\frac{\theta}{B} I \lesssim L^{-1/2}\Sigma L^{-1/2} \lesssim B\theta I,$$

where the notation $C \lesssim D$ means that $D - C$ is positive semidefinite. Thus, given an empirical unnormalized SFS $\hat{\upsilon}$, we can whiten it and then obtain a likelihood maximizing solution $\hat{\tilde{\eta}}$ under this model by solving the estimating equation

$$L^{-1/2}\hat{\upsilon} = F\hat{\tilde{\eta}} \tag{4.12}$$

by least squares, where

$$F \stackrel{\text{def}}{=} \theta L^{-1/2} WE \tag{4.13}$$

shall be called the *SFS operator* henceforth. Our goal then becomes to understand the statistical properties of this estimator.

[2]We follow the convention of setting an operator T in Roman type and its matrix representation $\mathbf{T}$ in bold.

Remark (Identifiability). It is important to point out that F—in particular the operator E—is *not* invertible for generic functions η (Myers, Fefferman, and Patterson, 2008), though it is for certain restricted function classes (Bhaskar and Song, 2014). This is due a fundamental result in functional analysis known as the Müntz-Szàsz theorem: let $S = \{s_1, s_2, \dots\} \subset \mathbb{N}$. Then

$$\operatorname{cl} \langle x^{s_1}, x^{s_2}, \dots \rangle = C[0,1] \iff \sum_{s \in S} \frac{1}{s} = \infty.$$

After a change of variables, the operator E can be seen to project $\eta(-\log x)$ onto the monomials $x^{\lambda_2 - 1}, x^{\lambda_3 - 1}, \dots$, and since $\sum \frac{1}{\lambda_i - 1} < \infty$ the theorem implies that E has a non-trivial kernel even if $n = \infty$.

4.3 Theory of linear inverse problems

To study the operator described above we will use some basic notions from the spectral theory of linear inverse problems. The results in this section are standard; we follow the exposition of Cavalier (2011).

Let $A : H \to G$ be a linear map between Hilbert spaces H and G. In the standard linear inverse problem, we aim to recover $f \in H$ given linearly-related measurements

$$Y = Af + \varepsilon\xi \tag{4.14}$$

where $\xi \in G$ is an error and $\varepsilon \in \mathbb{R}$ is the noise level. For deterministic ξ, the inverse problem (4.14) is very well studied and occurs throughout science (Denisov, 1999, e.g.).

The distinguishing feature in the statistical setting is that ξ is treated as random. In full generality ξ can be considered to be a Hilbert space process, but we will simply assume that it is a random variable, i.e. a measurable map from some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ into G. Our general aim is to give guarantees on the quality of the inferred $\hat{f}$ as a function of ε, in particular as $\varepsilon \to 0$.

The main theorem for studying linear operators is the following functional generalization of the singular value decomposition.

Theorem 4.1 (Halmos, 1963)**.** *For any self-adjoint operator $A : D(A) \to G$ defined on a dense subset of separable Hilbert space H there exists a σ-compact space S, a Borel measure Σ on S, a unitary operator $U : H \to L^2(\Sigma)$, and a positive measurable function $\rho : S \to \mathbb{R}_+$ such that for all $f \in D(A)$,*

$$UAf = \rho \cdot Uf, \quad \Sigma \text{ a.e.} \tag{4.15}$$

The assumption that A is self-adjoint is not restrictive since we may consider AA^*. Defining $M_\rho : L^2(\Sigma) \to L^2(\Sigma)$ to be the multiplication operator, $M_\rho \psi = \rho \cdot \psi$, equation (4.15) can be rewritten as

$$A = U^{-1} M_\rho U.$$

In other words, every operator acts as multiplication in some basis.

If A is compact, then the spectral theorem implies that A has a complete orthonormal system of eigenfunctions $\{\psi_k\}$ with eigenvalues b_k^2, such that

$$A^*Af = \sum_{k=1}^{\infty} b_k^2 \langle f, \phi_k \rangle \phi_k = \sum_{k=1}^{\infty} b_k^2 \rho_k \phi_k$$

for $\rho_k = \langle f, \phi_k \rangle$. Defining $\psi_k = b_k^{-1} A\phi_k$, we have

$$\begin{aligned} y_k &\stackrel{\text{def}}{=} \langle Y, \psi_k \rangle \\ &= \langle Af + \varepsilon\xi, \psi_k \rangle \\ &= \left\langle Af, b_k^{-1} A\phi_k \right\rangle + \varepsilon\xi_k \\ &= b_k^{-1} \langle A^*Af, \phi_k \rangle + \varepsilon\xi_k \\ &= b_k \rho_k + \varepsilon\xi_k \end{aligned} \tag{4.16}$$

for $\xi_k \stackrel{\text{def}}{=} \langle \xi, \phi_k \rangle$. This is the so-called *sequence space model* which reduces the problem to estimating the sequence $\rho_1, \rho_2, \ldots$ given $y_1, y_2, \ldots$.

Since A is compact the singular values $b_k \to 0$ and the difficulty of the estimation problem is governed by rate at which they do so. Another way to see this is to define $z_k = b_k^{-1}$ and $x_k = z_k y_k$ so that

$$x_k = \rho_k + \varepsilon z_k \xi_k. \tag{4.17}$$

Rapid spectral decay is equivalent to rapidly increasing variance, which limits the amount of signal (i.e., the number of ρ_k) that can be recovered for a given noise level ε.

The singular decomposition of F

Let us now study the singular system of F (4.13). $FF^* : \mathbb{R}^{(n-1)} \to \mathbb{R}^{(n-1)}$ is represented by the matrix product $\mathbf{M} \stackrel{\text{def}}{=} \mathbf{L}^{-1/2}\mathbf{W}\mathbf{G}\mathbf{W}^*\mathbf{L}^{-1/2}$ where $\mathbf{G}$ represents EE^*,

$$G_{ij} = \int_0^{\infty} e^{-(\lambda_i + \lambda_j)t} \, \mathrm{d}t = \frac{2}{\lambda_i + \lambda_j}.$$

The eigensystem of the matrix $\mathbf{M}$ can be computed numerically and from it the singular functions $\phi_1, \ldots, \phi_{n-1}$ of F obtained. Figure 4.1 plots a few of them for $n = 50$. The functions have increasing oscillation towards the present, which suggests that the resolution of SFS estimates will be inherently better in the recent past. In the more distant past, the functions are much smoother, such that high frequency details of the size history function will be lost when projecting onto the singular functions.

It is interesting to visualize these functions because they are intrinsic to the operator under study and hence universal in a sense. This is often mentioned as a criticism of the SVD approach to linear estimation, since the basis functions are determined solely by the operator and

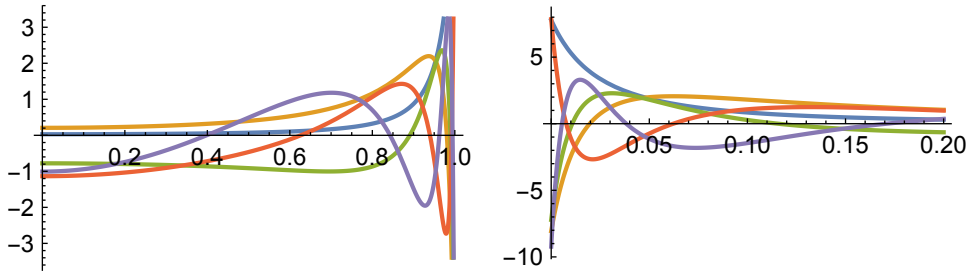


Figure 4.1: Top five singular functions for $n = 50$. The left panel plots the functions for $u \in (0, 1]$ under the transformation $t = -\log u$, so that the entire positive real line is depicted and the present corresponds to $t = 1$. The right panel plots the actual singular functions over the range $t \in [0, .02]$.

might bear little resemblance to the solutions f actually encountered in practice. On the other hand, in cases where there is not an *a priori* obvious basis for solutions (including, arguably, the demographic inference problem), the singular functions furnish one. Additionally, they have an appealing parsimony property: the solution to (4.14) of minimum norm is obtained by projecting onto their span.

Nonparametric estimation

The statistical theory of nonparametric estimation is particularly elegant in the spectral domain. We again follow Cavalier (2011). Let $\alpha = \alpha_1, \alpha_2, \ldots$ be a sequence of weights, let

$$\hat{\rho}_k = \alpha_k x_k \tag{4.18}$$

where x_k is defined in (4.17), and define the linear estimator

$$\hat{f}(\alpha) = \sum_{k=1}^{\infty} \hat{\rho}_k \phi_k.$$

For any estimator $\hat{f}$, the *risk* of $\hat{f}$ with respect to f, $\mathcal{R}(f, \hat{f})$ is defined to be the mean-squared error

$$\mathcal{R}(\hat{f}, f) = \mathbb{E}_f \|f - \hat{f}\|^2 \stackrel{\text{def}}{=} \mathbb{E}_\rho \sum_{k=1}^{\infty} (\hat{\rho}_k - \rho_k)^2 \stackrel{\text{def}}{=} \mathbb{E}_\rho \|\hat{\rho} - \rho\|^2.$$

Thus, by (4.17) and (4.18), and assuming that the ξ_k are white noise,

$$\mathcal{R}(\hat{f}, f) = \sum_{k=1}^{\infty} (1 - \alpha_k)^2 \rho_k^2 + \varepsilon^2 \sum_{k=1}^{\infty} \alpha_k^2 z_k^2.$$

For a given smoothness class of functions $\mathcal{F}$ and noise level ε, the *minimax risk* as defined as

$$r_\varepsilon(\mathcal{F}) = \inf_{\hat{f}} \sup_{f \in \mathcal{F}} \mathcal{R}(f, \hat{f}(\alpha)) \tag{4.19}$$

A weight sequence which is $\Theta(r_\varepsilon(\mathcal{F}))$ as $\varepsilon \to 0$ is said to be *asymptotically minimax*.

In this chapter we will consider only the most basic linear estimator:

$$\alpha_k = 1 \text{ if } k \leq T, 0 \text{ otherwise.} \tag{4.20}$$

That, we project to the first T singular components and discard the rest. Despite its simplicity, the spectral cutoff estimator is asymptotically minimax over a broad range of smoothness classes (Bissantz et al., 2007).

The projection estimator has risk

$$\mathcal{R}(T, \mathcal{F}) = \varepsilon^2 \sum_{k=1}^{T} z_k^2 + \sum_{k=T+1}^{\infty} \rho_k^2 \tag{4.21}$$

where the first summation is a variance term and the second represents bias. We seek the optimal N^* that balances the two. Clearly this depends on the coordinates ρ_k of $f = \sum \rho_k \phi_k$ in the spectral basis, i.e. on the function class $\mathcal{F}$. In special cases, this function class is interpretable. For example, differentiation and deconvolution are diagonalized by the Fourier basis, and elliptical function classes of the form

$$\mathcal{F}(a, L) = \left\{ f = \sum \rho_k \phi_k \,:\, \sum a_k^2 \rho_k^2 \leq L \right\}$$

have a direct interpretation in terms of the smoothness of f.

In our problem setting, it is not yet clear how the spectral coordinates ρ_k relate to properties of the functions f in the singular span (but see the remark immediately following.)

Remark (Müntz polynomials). For a sequence $c_1, c_2, \cdots \in \mathbb{N}$, polynomials of the form $p(x) \in \text{span}(x^{c_1}, x^{c_2}, \ldots)$ have been studied in the approximation theory literature under the name *Müntz polynomials*. In particular, applying the Graham-Schmidt procedure to the monomials $x^{c_1}, x^{c_2}, \ldots$ one obtains the so-called Müntz-Legendre polynomials (Borwein, Erdélyi, and Zhang, 1994)

$$L(c_1, \ldots, c_n; x) \in \langle x^{c_1}, \ldots, x^{c_n} \rangle,$$

which are pairwise orthonormal with respect to Lebesgue measure on the unit interval. The polynomials can be written down by an explicit formula, and there are also a number of connected results on the ability of these polynomials to approximate arbitrary powers of x and other functions (e.g. Newman, 1965).

As mentioned above, the operator E projects $\tilde{\eta}(-\log x)$ onto the Müntz space

$$\mathcal{M}_n \stackrel{\text{def}}{=} \left\langle x^{\lambda_2 - 1}, \ldots, x^{\lambda_n - 1} \right\rangle;$$

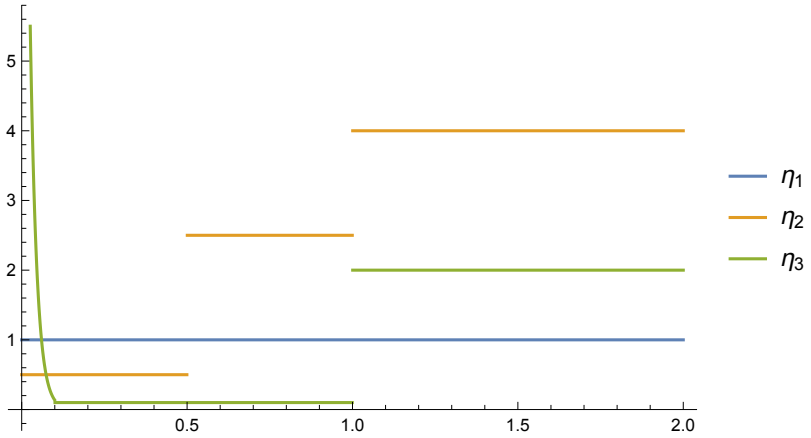


Figure 4.2: Piecewise size history functions used to test approximability.

this subspace is the "identifiable component" of $L^2[0,1]$ when using the SFS to estimate size history. The Müntz-Legendre polynomials $L(\lambda_2 - 1, \dots, \lambda_n - 1; x)$ form an orthogonal basis for this subspace; in particular, the first $(n-1)$ such polynomials have the same span as the singular functions $\phi_1, \dots, \phi_{n-1}$. Hence, they are intimately related to the types of size history functions that one might hope to recover from the frequency spectrum.

It is also interesting to note the connection with the identifiability result of Bhaskar and Song (2014). That result establishes injectivity of the operator WER (4.10) for piecewise constant, exponential, and generalized exponential size histories for sufficiently large n. This implies that these functions project uniquely onto the corresponding $\mathcal{M}_n$, though it is unclear that they live inside the space. For such functions, recovering the original model in a nonparametric setting is at least as difficult as accurately recovering its projection onto $\mathcal{M}_n$.

In short, the relationship between Müntz spaces and demographic inference using the SFS appears to be very interesting. We do not pursue these questions here, but aim to take them up in future research.

4.4 Numerical results

In this section we perform some numerical experiments based on the theoretical model derived above. First, we study the approximability of piecewise size functions (the type most commonly found in the literature) by the singular system of the SFS operator. The three functions η_1, η_2, η_3 are plotted in Figure 4.2. $\eta_1 = 1$ is constant. η_2 consists of three pieces which increase as time runs backwards. η_3 has a recent bottleneck followed by sharp exponential growth in the recent past. The functions are constant beyond $t = 1$.

In Figure 4.3 we have plotted the projections of these functions onto the singular spaces for the SFS operators of samples sizes $n = 5, 10, 25$. These plots give some idea of the increase

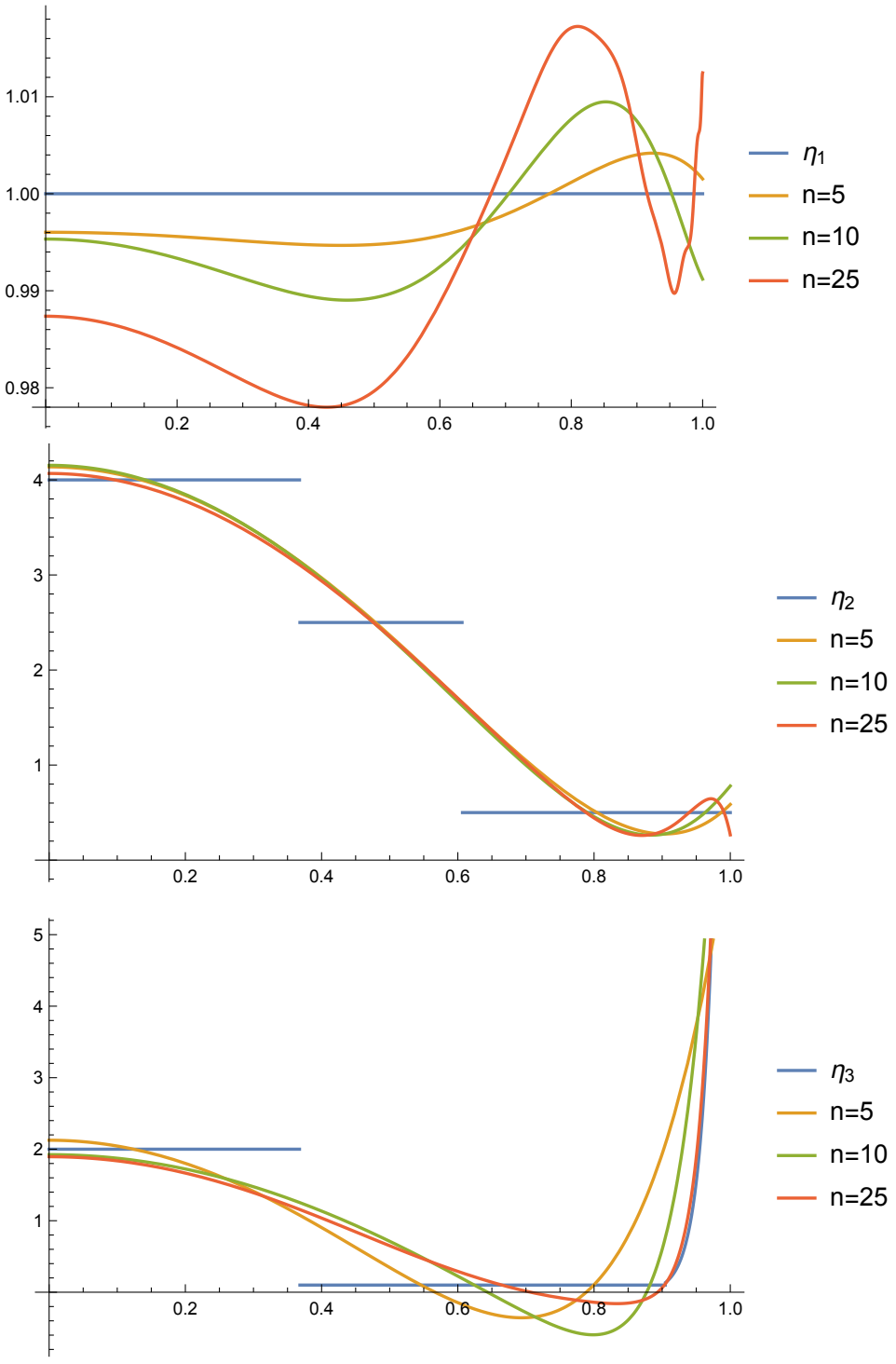


Figure 4.3: Projection of η_1, η_2, η_3 onto the singular components of the SFS operator for $n = 5, 10, 25$. To show the entire time range we have plotted $\eta_i(-\log t)$ for $t \in [0, 1]$.

approximation degree realized by sampling more individuals. For η_1 all three methods are within a few percent of the truth (note the scale of the y-axis.) Visually, the best fit in this case is obtained by using $n = 5$ samples, suggesting the need for regularization.

For η_2 all three methods are comparable and do a good job of interpolating the piecewise constant function. The is again a slight increase in variance towards the recent past observed when we increase the sample size to $n = 25$. It is interesting to note that when simulating data under a piecewise constant model, we have commonly observed that demographic inference methods tend to smooth over the boundaries of the disjoint pieces exactly as is depicted in this plot. To the extent that they are implicitly projecting onto this singular space, this offers a potential explanation for that behavior.

The final plot in the figure shows the η_3, the size history with recent exponential growth. The results here are interesting in several ways. First, the increase in estimation accuracy is more pronounced, which serves to confirm the empirical observation that larger sample sizes are critical for accurately estimating size history in the recent past. Second, we note that all three projections go negative for a portion of time. This highlights an import shortcoming of the SVD-based estimator: we do not have a way to encode prior knowledge about the size history functions, in particular non-negativity, when performing estimation. (One could simply threshold the estimated size history, but this would destroy any theoretical guarantees concerning statistical optimality). On the other hand, if the size history functions are known to be bounded below by some constant, then it should be possible to prove a result that the estimated function $\hat{\eta}$ is, with high probability or in expectation, non-negative by virtue of it being close to the true function η. We leave this for future work.

In the next set of experiments, we study the performance of the spectral cutoff estimator (4.20). We took the expected SFS η_3 and perturbed it under the noise model (4.11). Then, we computed the spectral decomposition of the operator FF^* (4.13) and used it to derive the projection estimator. Since we do not yet have a coherent theory relating the spectral coordinates ρ_k to the types of functions we hope to infer (i.e. we lack a way to choose the optimal cutoff T^* in (4.21)), we simply plot the inferred function for a few values of T.

We focus on η_3 as being the most realistic of the three contrived examples, and a sample of size $n = 10$. With $\theta = 1$, we investigated two noise levels, $\varepsilon = 10^{-2}$ and $\varepsilon = 1$. Results are shown in Figures 4.4 (low noise) and 4.5 (high noise). In the low noise regime performance is comparable to the noise-free setting examined above up the first 3 singular components. Including additional components starts to degrade the fits significantly. In the high noise setting we see quite bad performance with $T = 5$ singular components, producing an extremely oscillatory and erroneous fit. Performance at $T = 3$ is acceptable. Interestingly, the projecting onto the first singular component ($T = 1$) produces an over-smoothed but acceptable fit in both cases; it is able to detect the recent increase in population growth even with a relatively small sample size of $n = 10$. For the other cases, the estimated fits are negative in places, and the problem worsens as the value of the cutoff increases to its maximum value of $n - 1$.

These results hint at the possibility that dimension of model space which can actually be estimated using frequency spectrum data is perhaps on the order of 3–5, and in any case is probably much lower than the dimension of the spectrum itself which is $n-1$. This is because

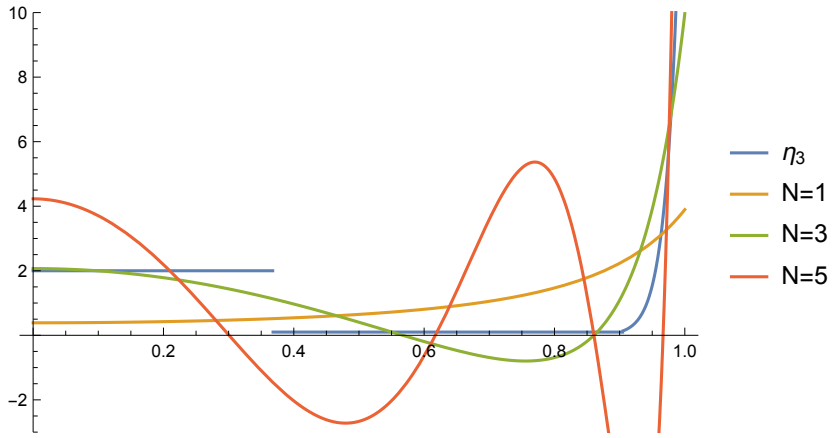


Figure 4.4: Performance of spectral cutoff estimator, low-noise case.

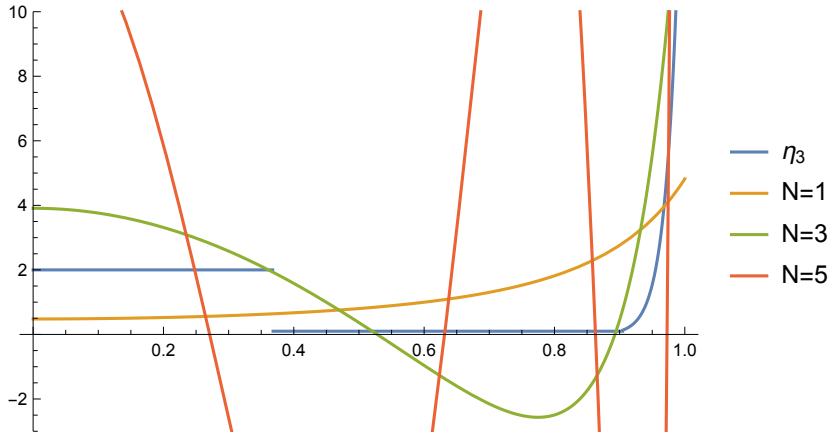


Figure 4.5: Performance of spectral cutoff estimator, high-noise case.

the operator in question is very ill-conditioned, a question which we take up in the next section.

4.5 Spectral decay of the SFS operator

The goal in this section is to provide some theoretical intuition for the computational results observed above. In particular, we aim to estimate the rate of spectral decay of the operator[3] WE which, as we have seen, controls the degree of ill-posedness of the inverse problem.

[3]The whitening transform $L^{-1/2}$ applied to $\mathbf{W}$ only alters each row by a factor of b, and does not change the analysis.

Our strategy will be to estimate the rate of spectral decay for the operators W and E separately. Assuming the rates are comparable, we can then estimate the decay of WE using the following elegant infinite dimensional generalization of the fundamental inequality $2ab \leq a^2 + b^2$. (In the theorem, $s_j(A)$ denotes the j-th singular value of A.)

Theorem 4.2 (Bhatia and Kittaneh, 1990)**.** *Let A, B be compact operators. Then for $j = 1, 2, \ldots$, we have*

$$2s_j(A^*B) \leq s_j(AA^* + BB^*).$$

It is straightforward to show that the spectrum of EE^* decays like e^{-j}. As for the spectrum of WW^*, there is ample numerical evidence that it obviously also decays exponentially, but our argument is delicate and we do not currently claim a proof.

Spectrum of E

To compute the spectrum of $\mathbf{E}$ we note that its adjoint $\mathbf{E}^* : \mathbb{R}^{n-1} \to L_2[0, \infty)$ is given by

$$\mathbf{E}^*\mathbf{v} = \sum_{i=1}^{n-1} v_i e^{-\lambda_i t},$$

and therefore the squared singular values of $\mathbf{E}$ are equal to the eigenvalues of

$$\mathbf{E}\mathbf{E}^* = \left(\frac{2}{\lambda_i + \lambda_j}\right)_{i,j=2}^{n}.$$

The matrix $\mathbf{E}\mathbf{E}^*$ is a Cauchy matrix; this class of matrices is known to exhibit rapid spectral decay.

Theorem 4.3. *Let $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_{n-1}^2 \geq 0$ be the eigenvalues of $\mathbf{E}\mathbf{E}^*$ Then*

$$\frac{\sigma_r^2}{\sigma_1^2} \leq a_n \exp(-b_n r),$$

where $a_n, b_n > 0$ are constants depending on n with $a_n = \Theta(1)$ and $b_n = \Theta(\log^{-1} n)$ as $n \to \infty$.

Proof. Let $\kappa = \lambda_n/\lambda_2 = \Theta(n^2)$. By Theorem 1 of Penzl (2000),

$$\frac{\sigma_{r+1}^2}{\sigma_1^2} \leq \underbrace{\left(\prod_{j=0}^{r-1} \frac{\kappa^{(2j+1)/(2r)} - 1}{\kappa^{(2j+1)/(2r)} + 1}\right)^2}_{(\star)}.$$

To make sense of the bound, we note that

$$\frac{\kappa^{(2j+1)/(2r)} - 1}{\kappa^{(2j+1)/(2r)} + 1} = \frac{1 - \sqrt{\kappa^{-(2j+1)/r}}}{1 + \sqrt{\kappa^{-(2j+1)/r}}}$$

so that by Lemma 4.7,

$$\begin{aligned}\log(\star) &\le -4\sum_{j=0}^{r-1}\kappa^{-(2j+1)/(2r)} \\ &\le -4\kappa^{-1/(2r)} r\int_0^r \kappa^{-x}\,\mathrm{d}x \\ &= \underbrace{-4r\kappa^{-1/(2r)}\frac{1-\kappa^{-1}}{\log\kappa}}_{f_\kappa(r)}.\end{aligned}$$

The function $f_\kappa(r)$ is concave on $r > 0$:

$$f_\kappa''(r) = -\frac{4}{r^3}(\kappa-1)\kappa^{-\frac{r+1}{r}}\log(\kappa) < 0.$$

Therefore, it is upper bounded by any tangent. In particular, taking the first two terms of its asymptotic expansion, we get

$$f_\kappa(r) \le \left(4 - \frac{4}{\kappa}\right) - \frac{4(\kappa-1)r}{\kappa\log\kappa},$$

which completes the proof. □

Figure 4.6 compares the observed spectral decay for the matrix EE^* with the bound given in the theorem. On a log scale, the bound predicts linear decay, whereas the actual rate appears to be faster. In practice there is little difference; already for $r = 10$ the actual singular value is $O(10^{-5})$.

Spectrum of W

In this section we derive some bounds of the spectral decay of the matrix $\mathbf{W}$ defined above. The results of this section suggest that the spectrum of this matrix also exhibits exponential decay. Compared with the preceding section we are only able to obtain somewhat less precise results on the rate of decay. We will focus on the regime where n and j are both large with $j = \Theta(n)$. Owing to this approximation, we currently regard the main result of this section as conjectural; our argument is tedious and some subtleties have been glossed over. We strongly suspect that a simpler, more readily verified proof exists since it is numerically very obvious that the entries W_{bj} decay extremely fast in j.

Conjecture 4.4. *Let $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_{n-1} > 0$ denote the singular values of $\mathbf{W}$. Then, as $n \to \infty$ such that $k/n \to \alpha \in (0,1)$, $\sigma_k = O(e^{-k^2/(2n)})$.*

By symmetry of the integrand, $|W_{bj}| = |W_{n-b,j}|$, so it would suffice to show this for $b \le n/2$, which we will assume for the rest of the section. The result relies on the fact that, for fixed b,

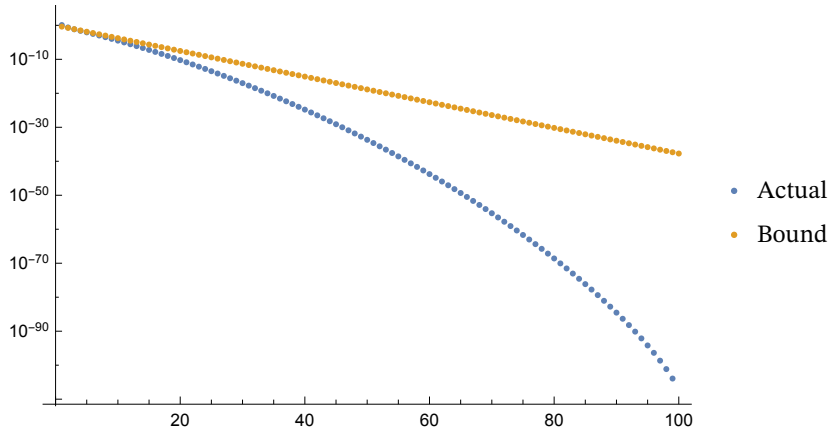


Figure 4.6: Actual vs. predicted spectral decay for the operator EE^*, $n = 50$

.

the norm of the entries W_{bj} is $O(e^{-j^2/n})$. Hence, we may form a rank-k matrix $\mathbf{W}_k$, obtained by zeroing out the last $n-1-k$ columns, such that

$$\|\mathbf{W} - \mathbf{W}_k\|_F^2 \leq n \sum_{j=k+1}^{n} e^{-j^2/n}.$$

If n is large and $k = \alpha n$ is of the same order as n, then the sum is $(n/\alpha)e^{-\alpha^2 n} = O(e^{-k^2/n})$. By the Eckhart-Young theorem, this implies that $\sigma_k = O(e^{-k^2/n})$, which would give the result.

The decay of W_{bj}

Hence, it suffices to show that, in the asymptotic regime described above, $|W_{bj}| = O(e^{-j^2/n})$. Before doing so, let us give some intuition for why this result holds, and also some of the difficulties we encountered when attempting to prove it.

Figure 4.7 plots the polynomial $B_{50,20}(x)$ (rescaled by a factor of 10) and several of the Jacobi polynomials. The sequence $W_{bj}, j = 2, \ldots, n$ consists of integrals of the products of these functions and can be thought of as, up to a scaling factor, the coordinates of the Bernstein polynomial

$$B_{n-2,b-1} = \binom{n-2}{b-1} x^{b-1}(1-x)^{n-b-1}$$

in the Jacobi basis. From the figure, one sees that the Jacobi polynomials oscillate at increasing frequencies as j increases. A smooth function like the one represented by $B_{50,20}(x)$ in the figure is "locally constant" on very small intervals, so the the higher frequency oscillations will tend to cancel out, leading to a rapid decay in the magnitude of the coefficients.

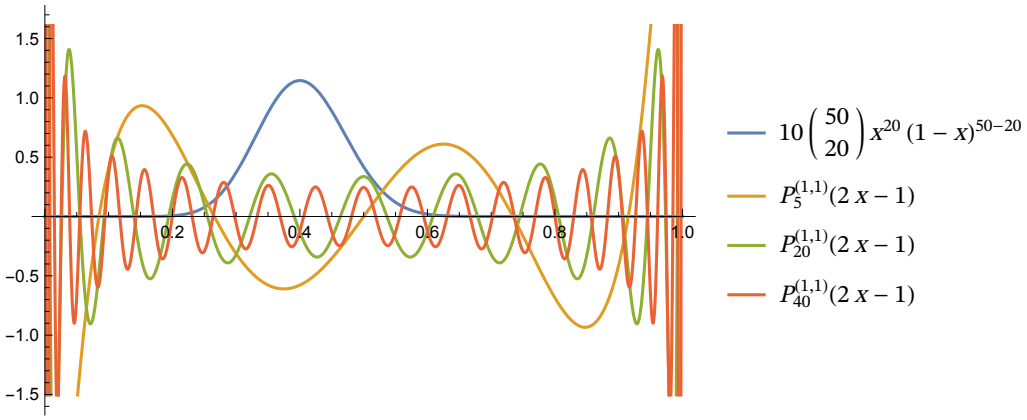


Figure 4.7: A plot of $B_{50,20}(x)$ and the Jacobi-$(1,1)$ polynomials of order 5, 20, and 40. The Bernstein polynomial has been rescaled by a factor of 10 to improve legibility.

To take advantage of the cancellations one typically integrates by parts, an example of the so-called *method of stationary phase* (Olver, 1997). If f has n derivatives then its n-th coefficient in the Jacobi basis,

$$c_n = \int_{-1}^{1} f(x)(1-x)(1+x)P_n^{(1,1)}(x)\,\mathrm{d}x = \frac{(-1)^n}{2^n n!}\int_{-1}^{1} f^{(n)}(x)(1-x^2)^{n+1}\,\mathrm{d}x, \tag{4.22}$$

where we have applied Rodrigues' formula (4.35) and integrated by parts n times. Hence if $|f^{(n)}|/n! \le M$, say, then $c_n \le M/2^n$.

Unfortunately, in the case $f = B_{n,b}$ this method does not produce the desired result. The difficulty is that the higher derivatives are also highly oscillatory (Figure 4.8), resulting in a very weak (essentially constant) bound. Thus, both integrals in (4.22) rely on substantial cancellations in order to form the result. One can also see this in the series definition (4.36) of the Jacobi polynomials: integrating termwise and summing introduces in many cancellations and ultimately, a very small number. It is not obvious how to effectively control these cancellations; naïvely applying the triangle inequality is very wasteful.

The phenomenon of highly oscillatory functions have large alternating coefficients in the polynomial basis indicates that this basis is somehow "inefficient" for representing these functions. This suggests that the problem may become easier if represented in, say, a trigonometric basis.

This turns out to be correct to an extent. That is, we will reduce (4.9) to an integral over the Chebyshev polynomials $P_j^{(-1/2,-1/2)}(x)$, which have the representation (4.41). It appears possible to analyze latter expression and obtain the result. However our current argument is long and tedious, with various subtleties ignored, and we do not yet claim a proof.

Figure 4.9 shows a comparison of the computed singular values for $n = 50, 100$ versus the bound given by the conjecture. No effort was made to obtain the correct constant and we are only interested in the rate of decay for large j. The bound seems to capture the correct

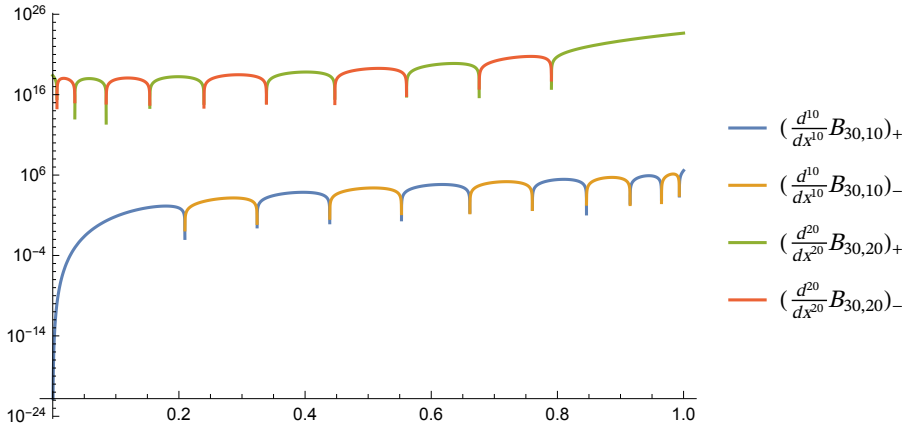


Figure 4.8: Derivatives of $B_{30,10}(x)$ on a log scale. The alternating colors indicate sign changes.

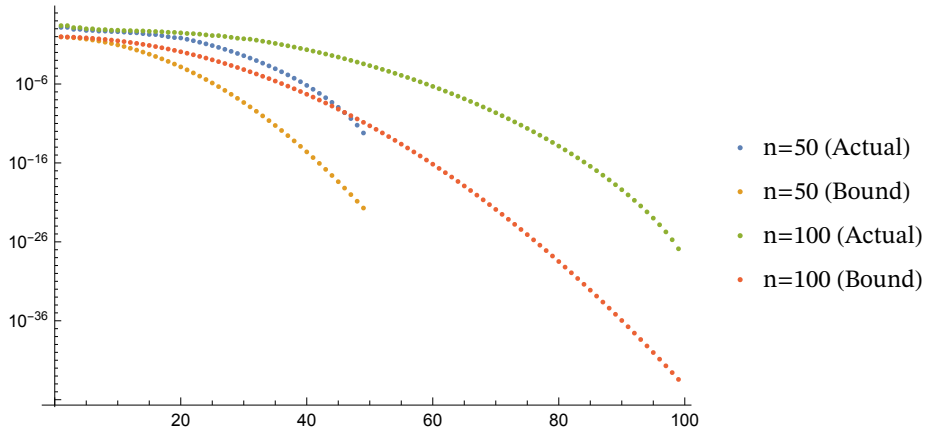


Figure 4.9: Actual vs. bound estimate of singular value decay for the $\mathbf{W}$ $n = 50, 100$.

behavior as j grows, while being perhaps slightly conservative in the sense that it gives exhibits (log-)linear decay for j close to n, whereas the computed values appear to be decreasing at a slightly faster rate.

Evidence for Conjecture 4.4

First, we show that W_{bj} is within polynomial factors of the corresponding Legendre integral.

Lemma 4.5. *Let $P_j(x) = P_j^{(0,0)}(x)$ be the j-th Legendre polynomial. Then*

$$\int_0^1 B_{n,b}(x) P_j^{(1,1)}(2x-1)\,\mathrm{d}x = \int_0^1 \left(\frac{1}{n} B_{n-1,b-1}(x) + \left(\frac{1}{n} - \frac{1}{b}\right) B_{n-1,b}(x)\right) \frac{P_{j+1}(2x-1)}{j+2}\,\mathrm{d}x. \quad (4.23)$$

Proof. By the differentiation identity (4.38),

$$\int P_j^{(1,1)}(2x-1) = \frac{P_{j+1}(2x-1)}{j+2}.$$

Integrate by parts; the boundary terms vanish since $B_{n,k}(0) = B_{n,k}(1) = 0$. □

Next, we invoke a theorem which says that, for functions with fast decaying spectrum, the Legendre and Chebyshev coefficients are equivalent up to a factor which is polynomial in the degree of P_j. In our notation, the theorem reads as follows.

Theorem 4.6 (Boyd and Petschek, 2014)**.** *Let a_j and b_j denote the Chebyshev and Legendre coefficients of a function $f(x)$, where the polynomials are normalized so that $T_j(1) = P_j(1) = 1$ for all j. If $f(x)$ is a function such that its Chebyshev coefficients have the property that for all $j > j_*$,*

$$|a_{j+k}| \leq \exp(-\chi(j)\,k)\,|a_j| \quad \forall k > 0$$

where for some positive constant q

$$\chi(j) > q \log(j) \quad \forall j > j_*$$

then

$$b_j \sim \frac{\sqrt{\pi\, j}}{2} a_j$$

.

(The normalization introduces only additional polynomial factors in j, which we discard.) Thus, if $|a_j| \asymp e^{-j^2/n}$ then there exists a j_* such that

$$\begin{aligned} |a_{j+k}| \leq e^{-(j^2+2jk+k^2)/n} &= e^{-j^2/n} e^{-(2jk+k^2)/n} \\ &\leq e^{-j^2/n} e^{-(2jk+1)/n} \\ &\leq |a_j| e^{-\chi(j)k} \end{aligned}$$

for $\chi(j) = 2j/n$, where j_*, c are such that $e^{-(j^2+c)/n} \leq |a_j|$ for $j > j_*$.

At this point we have reduced the problem to showing that the integrals

$$\int_0^1 B_{n,b}(x) \frac{T_j(2x-1)}{\sqrt{x(1-x)}}\,\mathrm{d}x \quad (4.24)$$

decay exponentially in j. Changing variables $x \to (1+\cos\theta)/2$, we have

$$\begin{aligned}(4.24) &= 2^{-(n+1)} \int_0^\pi (1-\cos\theta)^{b-1/2}(1+\cos\theta)^{n-b-1/2}\cos(j\theta)\,\mathrm{d}\cos\theta \\ &= 2^{-(n+1)} \int_0^\pi (\sin\theta)^{2b}(1+\cos\theta)^{n-2b}\cos(j\theta)\,\mathrm{d}\theta \\ &= 2^{-(n+1)} \int_0^\pi (\sin\theta)^{2b}(2\cos^2(\theta/2))^{n-2b}\cos(j\theta)\,\mathrm{d}\theta \\ &= 2^{-2b+1} \int_0^\pi (\sin\theta)^{2b}(\cos(\theta/2))^{2(n-2b)}\cos(j\theta)\,\mathrm{d}\theta. \end{aligned} \tag{4.25}$$

Using Lemma 4.8, we can decompose the integral as

$$\int_0^\pi (I_1 + I_2 + I_3 + I_4)\cos(j\theta)\,\mathrm{d}\theta$$

where

$$I_1 = C_1 C_2 \tag{4.26}$$

$$I_2 = C_2 \frac{(-1)^b}{2^{2b-1}} \sum_{k=0}^{b-1} (-1)^k \binom{2b}{k} \cos(2(b-k)\theta) \tag{4.27}$$

$$I_3 = C_1 \frac{1}{2^{2(n-2b)-1}} \sum_{k=0}^{n-2b-1} \binom{2(n-2b)}{k} \cos((n-2b-k)\theta) \tag{4.28}$$

$$\begin{aligned} I_4 = \frac{(-1)^b}{2^{2(n-b-1)}} \sum_{\ell=0}^{b-1} \sum_{k=0}^{n-2b-1} (-1)^\ell \binom{2b}{\ell}\binom{2(n-2b)}{k} \times \\ \cos(2(b-\ell)\theta)\cos((n-k-2b)\theta) \end{aligned} \tag{4.29}$$

for combinatorial coefficients

$$C_1 = 2^{-2b}\binom{2b}{b} \qquad\qquad C_2 = 2^{-2(n-2b)}\binom{2(n-2b)}{n-2b}.$$

Upon integrating, most of the terms in the summations vanish owing to Lemma 4.9 in Section 4.6. The terms that remain have an implicit dependence on j which we can use to estimate the magnitude of each $I_1, \ldots, I_4$ as j increases.

To apply Lemma 4.9, we must decompose each of the summations (4.27)–(4.29) into index sets such that at least one of the conditions in the lemma is true. We illustrate this process for the double summation I_4, by forming the decomposition

$$\int_0^\pi I_4 \cos(j\theta)\,\mathrm{d}\theta) = \int_0^\pi \cos(j\theta)(I_{4,1} + I_{4,2} + I_{4,3} + I_{4,4})\,\mathrm{d}\theta,$$

where e.g. $I_{4,1}$ sums over the ℓ, k such that $2(b-\ell) + (n-k-2b) + j = 0$. Setting $u = j, v = n-2b-k, w = 2(b-\ell)$, we proceed by cases:

- $I_{4,1}$: $u+v+w=0$. Since

$$\begin{aligned}(n-2b-k)+2(b-\ell)+j &= n+j-(k+2\ell)\\ &\geq n+j-((n-2b-1)+(b-1))\\ &= j+b-1 \geq b+1>0,\end{aligned}$$

  there are no terms which satisfy this condition.

- $I_{4,2}$: $u-v-w=0$. This is equivalent to $k=n-j-2\ell$. Hence, $0\leq n-j-2\ell\leq n-2b-1$ which implies

$$\max\left(0,\frac{2b+1-j}{2}\right)\leq \ell \leq \min\left(\frac{n-j}{2},b-1\right),$$

  so that $I_{4,1}$ reduces to

$$I_{4,1}=\frac{\pi}{4}\sum_{\ell=\left\lceil 0\vee\frac{2b-j+1}{2}\right\rceil}^{\left\lfloor\frac{n-j}{2}\wedge b-1\right\rfloor}(-1)^{\ell}\binom{2b}{\ell}\binom{2(n-2b)}{n-j-2\ell}.$$

  We look at the leading order behavior in j. First, we note that each term has norm bounded by a constant which does not depend on n or j. Indeed, the binomial terms are upper bounded by the central binomial coefficients, $\binom{2b}{\ell}\leq\binom{2b}{b}\leq 2^{2b}$ and similarly $\binom{2(n-2b)}{n-j-2\ell}\leq 2^{2(n-2b)}$, and these are, up to a spare factor of four, precisely the (inverses of the) leading-order factors in (4.29).

  Applying the basic bound

$$\binom{n}{k}\leq\left(\frac{en}{k}\right)^{k},$$

  we have that the binomial terms are upper bounded by

$$\binom{2b}{\ell}\binom{2(n-2b)}{n-j-2\ell}\leq 2^{2b}e^{n-j-2\ell}\left(\frac{2n}{n-j-2\ell}\right)^{n-j} \tag{4.30}$$

  Taking logs of the term containing j, we obtain

$$\begin{aligned}-j+(n-j)\log\left(\frac{2n}{n-j-2\ell}\right) &\leq -j+(n-j)\left[\log 2-\log\left(1-\frac{j}{n}-\frac{2\ell}{n}\right)\right]\\ &\leq -j+(n-j)\left[\log 2-\log\left(1-\frac{j}{n}-\frac{n-j}{2n}\right)\right]\\ &= -j+(n-j)\left[-\log\left(1-\frac{j}{n}\right)\right]\\ &\leq -j+(n-j)\left(\frac{j}{n}+O\left(\frac{j}{n}\right)^2\right)\\ &= -\frac{j^2}{n}+nO\left(\frac{j}{n}\right)^2.\end{aligned}$$

- $I_{4,3}$: $u + v - w = 0$. Equivalently, $n + j - 4b + 2\ell = k$. This implies that

$$\max\left(0, \left\lceil \frac{4b - (n+j)}{2} \right\rceil\right) \le \ell \le \min\left(b - 1, \frac{2b - j + 1}{2}\right). \tag{4.31}$$

We proceed as in the previous case, starting with the upper bound

$$\binom{2b}{\ell}\binom{2(n-2b)}{n+j-4b+2\ell} \le 2^{2b} e^{n+j-4b+2\ell} \left(\frac{2n}{n+j-4b+2\ell}\right)^{n+j-4b+2\ell}$$

and taking logs of the terms containing j:

$$\begin{aligned} -j + (n+j-4b+2\ell)\log\left(\frac{2n}{n+j-4b+2\ell}\right) \le \\ -j + (n+j+2\ell)\left[\log 2 - \log\left(1 + \frac{j}{n} - \frac{4b-2\ell}{n}\right)\right] \le \\ -j(1-\log 2) + 2n\log 2 \end{aligned} \tag{4.32}$$

where in the second inequality we used (4.31) to lower bound $-4b + 2\ell$.

- $I_{4,4}$: $u - v + w = 0$. This case is the same as $I_{4,3}$ up to a sign change in j: $k = n - j - 4b + 2\ell$. Starting from the first line of equation (4.32), we have

$$\begin{aligned} -j + (n-j-4b+2\ell)\log\left(\frac{2n}{n-j-4b+2\ell}\right) \le \\ -j + (n-j+2\ell)\left[\log 2 - \log\left(1 - \frac{j}{n} - \frac{4b-2\ell}{n}\right)\right] \le \\ -j + (n-j+2\ell)\left[\log 2 - \log\left(\frac{1-2j}{2n}\right)\right] = \\ -j + (n-j+2\ell)\left(\log n + \frac{2j}{n} + O\left(\frac{j}{n}\right)^2\right) \\ = -\frac{j^2}{n} - j(1+\log n) + O(n\log n). \end{aligned} \tag{4.33}$$

where in the second inequality we again used the bound (4.31) (with $-j$ substituted for $+j$) to lower bound the quantity $-4b + 2\ell$.

Putting everything together, we can upper bound the rate of decay in j for I_4 by the slowest computed rate from among $I_{4,i}$: $O(e^{-j^2/n})$.

The remaining cases I_2 (4.27) and I_3 (4.28) correspond to $u = j, v = 0, w = 2(b-k)$ and $u = j, v = n - 2b - k, w = 0$, respectively. The proofs use the same ideas as above and are omitted.

4.6 Appendix

In this section we collect some standard results and miscellaneous technical lemmas which are needed in the main text.

Orthogonal polynomials

Orthogonal polynomials play a key role in this theory. For the reader's convenience we record some standard results which will are used above; for proofs see e.g. Szegö (1939).

The *Jacobi-*(α, β) *polynomials* $P_n^{(\alpha,\beta)}(x), n = 0, 1, 2, ...$ are a sequence of polynomials with $\deg P_n = n$ which are orthogonal on the interval $[-1, 1]$ with respect to the measure $(1 - x)^\alpha (1 + x)^\beta \, \mathrm{d}x$:

$$\int_{-1}^{1} (1-x)^\alpha (1+x)^\beta P_m^{(\alpha,\beta)}(x) P_n^{(\alpha,\beta)}(x) \, \mathrm{d}x = a_n^{(\alpha,\beta)} \delta_{nm},$$

where

$$a_n^{(\alpha,\beta)} \stackrel{\text{def}}{=} \frac{2^{\alpha+\beta+1}}{2n+\alpha+\beta+1} \frac{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{\Gamma(n+\alpha+\beta+1)n!} \tag{4.34}$$

is the normalization coefficient. (Hence, the Jacobi polynomials are not customarily normalized.)

Standard results from the theory of orthogonal polynomials state that the Jacobi polynomials form a basis for $L^2([-1, 1])$: every function $f \in L^2([-1, 1])$ may be represented as $f(x) = \sum_{n=0}^{\infty} c_n P_n^{(\alpha,\beta)}(x)$ for the coefficient sequence

$$c_n = \int_{-1}^{1} (1-x)^\alpha (1+x)^\beta P_n^{(\alpha,\beta)}(x) f(x) \, \mathrm{d}x.$$

An equivalent definition of the Jacobi polynomials is in terms of the so-called *Rodrigues formula*:

$$P_n^{(\alpha,\beta)}(x) \stackrel{\text{def}}{=} \frac{(-1)^n}{2^n n!} (1-x)^{-\alpha} (1+x)^{-\beta} \frac{\mathrm{d}^n}{\mathrm{d}x^n} \left\{ (1-x)^\alpha (1+x)^\beta \left(1-x^2\right)^n \right\}. \tag{4.35}$$

Using this relation and integration by parts, the Jacobi polynomials can be given in closed form:

$$P_n^{(\alpha,\beta)}(x) = \frac{\Gamma(\alpha+n+1)}{n! \, \Gamma(\alpha+\beta+n+1)} \sum_{m=0}^{n} \binom{n}{m} \frac{\Gamma(\alpha+\beta+n+m+1)}{\Gamma(\alpha+m+1)} \left(\frac{x-1}{2}\right)^m. \tag{4.36}$$

The Jacobi polynomials satisfy a number of useful relations. In particular, the following three-term recurrence holds for fixed $\alpha, \beta > -1$ and $n \geq 2$:

$$\begin{aligned} 2n(n+\alpha+\beta)(2n+\alpha+\beta-2) P_n^{(\alpha,\beta)}(x) = \\ (2n+\alpha+\beta-1)\left\{(2n+\alpha+\beta)(2n+\alpha+\beta-2)x + \alpha^2 - \beta^2\right\} P_{n-1}^{(\alpha,\beta)}(x) - \\ 2(n+\alpha-1)(n+\beta-1)(2n+\alpha+\beta) P_{n-2}^{(\alpha,\beta)}(x). \end{aligned} \tag{4.37}$$

We also have the differentiation identity

$$\frac{d^k}{dx^k}P_n^{(\alpha,\beta)}(x) = \frac{\Gamma(\alpha+\beta+n+1+k)}{2^k\Gamma(\alpha+\beta+n+1)}P_{n-k}^{(\alpha+k,\beta+k)}(x). \tag{4.38}$$

The Jacobi polynomials generalize many other families of orthogonal polynomials. Two of them will be important for our analysis:

$$T_n(x) \stackrel{\text{def}}{=} P_n^{(-1/2,-1/2)}(x) \qquad \text{(Chebyshev)} \tag{4.39}$$

$$P_n(x) \stackrel{\text{def}}{=} P_n^{(0,0)} \qquad \text{(Legendre).} \tag{4.40}$$

Finally, the Jacobi polynomials have interesting properties in the trigonometric basis. For the Chebyshev polynomials (4.39) we have the useful relation

$$T_n(\cos\theta) = \cos(n\theta). \tag{4.41}$$

More generally, the *Darboux formula* gives the following asymptotic characterization of $P_n^{(\alpha,\beta)}(\cos\theta)$:

$$P_n^{(\alpha,\beta)}(\cos\theta) = n^{-\frac{1}{2}}k(\theta)\cos(N\theta+\gamma) + O\left(n^{-\frac{3}{2}}\right) \tag{4.42}$$

for x bounded away from ± 1 and

$$\begin{aligned} k(\theta) &= \pi^{-\frac{1}{2}}\sin^{-\alpha-\frac{1}{2}}\tfrac{\theta}{2}\cos^{-\beta-\frac{1}{2}}\tfrac{\theta}{2} \\ N &= n + \tfrac{1}{2}(\alpha+\beta+1) \\ \gamma &= -\tfrac{\pi}{2}\left(\alpha+\tfrac{1}{2}\right). \end{aligned}$$

Miscellaneous results

Lemma 4.7. *For* $0 \le x \le 1$,

$$\log\left(\frac{1-\sqrt{x}}{1+\sqrt{x}}\right) \le -2\sqrt{x}.$$

Proof. There is equality at $x = 0$ and

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[-2\sqrt{x} - \log\left(\frac{1-\sqrt{x}}{1+\sqrt{x}}\right)\right] = \frac{\sqrt{x}}{1-x} \ge 0.$$

□

Lemma 4.8 (Trigonometric power formulas)**.** *Let n be an integer. For $f \in \{\sin, \cos\}$,*

$$f(\theta)^{2n} = \frac{1}{2^n}\binom{n}{n/2} + \frac{2}{2^n}\sum_{k=0}^{\frac{n}{2}-1}(-1)^{\mathbf{s}\cdot k}\binom{n}{k}\cos\big((n-2k)\theta\big),$$

where $\mathbf{s} = 0$ *for* $f = \cos$ *and 1 for* $f = \sin$.

Proof. See Weisstein (2017). □

Lemma 4.9. *Let $u, v, w \in \mathbb{Z}$. Then*

$$\int_0^\pi \cos(u\theta)\cos(v\theta)\cos(w\theta)\,\mathrm{d}\theta = \frac{\pi}{4}\left(\delta_{u+v+w,0} + \delta_{u+v-w,0} + \delta_{u-v+w,0} + \delta_{u-v-w,0}\right).$$

Proof. For $u, v, w \in \mathbb{R}$ in general position, the integral equals

$$\frac{1}{4}\left(\frac{\sin(\pi(u-v-w))}{u-v-w} + \frac{\sin(\pi(u+v-w))}{u+v-w} + \frac{\sin(\pi(u-v+w))}{u-v+w} + \frac{\sin(\pi(u+v+w))}{u+v+w}\right).$$

If $u \pm v \pm w \in \mathbb{Z}\backslash\{0\}$ then the corresponding term is zero. Taking limits $\to 0$ gives the result. □

Bibliography

1000 Genomes Project Consortium (2010). "A map of human genome variation from population-scale sequencing". In: *Nature* 467, pp. 1061–1073.

Bhaskar, A. and Y. S. Song (2014). "Descartes' rule of signs and the identifiability of population demographic models from genomic variation data". In: *Annals of Statistics* 42.6, pp. 2469–2493.

Bhaskar, A., Y. X. R. Wang, and Y. S. Song (2015). "Efficient inference of population size histories and locus-specific mutation rates from large-sample genomic variation data". In: *Genome Research* 25.2, pp. 268–279.

Bhatia, R. and F. Kittaneh (1990). "On the Singular Values of a Product of Operators". In: *SIAM Journal on Matrix Analysis and Applications* 11.2, pp. 272–277. DOI: 10.1137/0611018.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer. ISBN: 0387310738.

Bissantz, N. et al. (2007). "Convergence Rates of General Regularization Methods for Statistical Inverse Problems and Applications". In: *SIAM Journal on Numerical Analysis* 45.6, pp. 2610–2636. DOI: 10.1137/060651884.

Borwein, P., T. Erdélyi, and J. Zhang (1994). "Müntz Systems and Orthogonal Müntz-Legendre Polynomials". In: *Transactions of the American Mathematical Society* 342.2, pp. 523–542.

Boyd, J. P. and R. Petschek (2014). "The Relationships Between Chebyshev, Legendre and Jacobi Polynomials: The Generic Superiority of Chebyshev Polynomials and Three Important Exceptions". In: *Journal of Scientific Computing* 59.1, pp. 1–27.

Browning, S. R. and B. L. Browning (2011). "Haplotype phasing: existing methods and new developments". In: *Nature Reviews Genetics* 12.10, pp. 703–714.

Cavalier, L. (2011). "Inverse problems in statistics". In: *Inverse problems and high-dimensional estimation*. Springer, pp. 3–96.

Chen, H. (2012). "The joint allele frequency spectrum of multiple populations: A coalescent theory approach". In: *Theoretical Population Biology* 81.2, pp. 179–195.

Coventry, A. et al. (2010). "Deep resequencing reveals excess rare recent variants consistent with explosive population growth". In: *Nature Communications* 1, p. 131.

Delaneau, O., J.-F. Zagury, and J. Marchini (2013). "Improved whole-chromosome phasing for disease and population genetic studies". In: *Nat Meth* 10.1, pp. 5–6.

Denisov, A. (1999). *Elements of the Theory of Inverse Problems*. Inverse and ill-posed problems series. VSP.

Drmanac, R. et al. (2009). "Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays". In: *Science* 327.5961, pp. 78–81. eprint: http://science.sciencemag.org/content/327/5961/78.full.pdf.

Dutheil, J. et al. (2009). "Ancestral Population Genomics: The Coalescent Hidden Markov Model Approach". In: *Genetics* 183, pp. 259–274.

Excoffier, L. et al. (2013). "Robust Demographic Inference from Genomic and SNP Data". In: *PLoS Genetics* 9.10, e1003905.

Fu, Q. et al. (2014). "Genome sequence of a 45,000-year-old modern human from western Siberia". In: *Nature* 514.7523, pp. 445–449.

Gazave, E. et al. (2014). "Neutral genomic regions refine models of recent rapid human population growth". In: *Proceedings of the National Academy of Sciences* 111.2, pp. 757–762.

Gladstien, K. (1978). "The Characteristic Values and Vectors for a Class of Stochastic Matrices Arising in Genetics". In: *SIAM Journal on Applied Mathematics* 34.4, pp. 630–642.

Graham, R. L., D. E. Knuth, and O. Patashnik (1994). *Concrete Mathematics: A Foundation for Computer Science*. 2nd ed. Reading, Mass.: Addison-Wesley. ISBN: 0201558025.

Gravel, S. et al. (2011). "Demographic history and rare allele sharing among human populations". In: *Proceedings of the National Academy of Sciences* 108.29, pp. 11983–11988.

Green, R. E. et al. (2010). "A draft sequence of the Neandertal genome". In: *Science* 328.5979, pp. 710–722.

Griffiths, R. C. and P. Marjoram (1997). "An ancestral recombination graph". In: *Institute for Mathematics and Its Applications* 87, p. 257.

Griffiths, R. C. and S. Tavaré (1994). "Sampling theory for neutral alleles in a varying environment". In: *Proc. R. Soc. London B.* 344, pp. 403–410.

Griffiths, R. and S. Tavaré (1998). "The age of a mutation in a general coalescent tree". In: *Communications in Statistics. Stochastic Models* 14.1-2, pp. 273–295.

Gutenkunst, R. N. et al. (2009). "Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data". In: *PLoS Genetics* 5.10, e1000695.

Halmos, P. R. (1963). "What Does the Spectral Theorem Say?" In: *The American Mathematical Monthly* 70.3, pp. 241–247.

Hobolth, A. and J. L. Jensen (2014). "Markovian approximation to the finite loci coalescent with recombination along multiple sequences". In: *Theoretical Population Biology* 98, pp. 48–58. ISSN: 0040-5809. DOI: http://dx.doi.org/10.1016/j.tpb.2014.01.002. URL: http://www.sciencedirect.com/science/article/pii/S0040580914000033.

Huerta-Sanchez, E. et al. (2014). "Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA". In: *Nature* 512.7513, pp. 194–197.

Kamm, J. A., J. Terhorst, and Y. S. Song (2016). "Efficient computation of the joint sample frequency spectra for multiple populations". In: *Journal of Computational and Graphical Statistics*.

Keightley, P. D. et al. (2014). "Estimation of the Spontaneous Mutation Rate per Nucleotide Site in a Drosophila melanogaster Full-Sib Family". In: *Genetics* 196.1, pp. 313–320.

Kim, J. et al. (2015). "Can one hear the shape of a population history?" In: *Theoretical Population Biology* 100, pp. 26–38.

Kimura, M. (1955). "Solution of a process of random genetic drift with a continuous model". In: *Proceedings of the National Academy of Sciences* 41.3, pp. 144–150.

— (1969). "The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations". In: *Genetics* 61.4, p. 893.

Kingman, J. F. C. (1982a). "Exchangeability and the evolution of large populations". In: *Exchangeability in Probability and Statistics*. Ed. by G. Koch and F. Spizzichino. North-Holland Publishing Company, pp. 97–112.

— (1982b). "On the genealogy of large populations". In: *J. Appl. Prob.* 19A, pp. 27–43.

— (1982c). "The coalescent". In: *Stoch. Process. Appl.* 13, pp. 235–248.

Lack, J. B. et al. (2015). "The Drosophila Genome Nexus: A Population Genomic Resource of 623 Drosophila melanogaster Genomes, Including 197 from a Single Ancestral Range Population". In: *Genetics* 199.4, pp. 1229–1241. ISSN: 0016-6731. DOI: 10.1534/genetics.115.174664. eprint: http://www.genetics.org/content/199/4/1229.full.pdf. URL: http://www.genetics.org/content/199/4/1229.

Langergraber, K. E. et al. (2012). "Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution". In: *Proceedings of the National Academy of Sciences* 109.39, pp. 15716–15721.

Li, H. and R. Durbin (2011). "Inference of human population history from individual whole-genome sequences". In: *Nature* 475, pp. 493–496.

Marjoram, P. and J. D. Wall (2006). "Fast "coalescent" simulation". In: *BMC Genetics* 7.1, p. 16.

Massart, P. (2007). *Concentration inequalities and model selection*. Vol. 1896. Berlin: Springer.

McVean, G. A. and N. J. Cardin (2005). "Approximating the coalescent with recombination". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 360.1459, pp. 1387–1393.

Miller, W. et al. (2012). "Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change". In: *Proceedings of the National Academy of Sciences* 109.36, E2382–E2390.

Myers, S., C. Fefferman, and N. Patterson (2008). "Can one learn history from the allelic spectrum?" In: *Theoretical Population Biology* 73.3, pp. 342–348.

Nelson, M. R. et al. (2012). "An Abundance of Rare Functional Variants in 202 Drug Target Genes Sequenced in 14,002 People". In: *Science* 337.6090, pp. 100–104.

Newman, D. J. (1965). "A Müntz-Jackson Theorem". In: *American Journal of Mathematics* 87.4, pp. 940–944.

Nielsen, R. (2000). "Estimation of population parameters and recombination rates from single nucleotide polymorphisms". In: *Genetics* 154.2, pp. 931–942.

Olver, F. (1997). *Asymptotics and Special Functions*. AKP classics. Taylor & Francis.

Paul, J. S. and Y. S. Song (2012). "Blockwise HMM computation for large-scale population genomic inference". In: *Bioinformatics* 28, pp. 2008–2015.

Paul, J. S., M. Steinrücken, and Y. S. Song (2011). "An accurate sequentially Markov conditional sampling distribution for the coalescent with recombination". In: *Genetics* 187. (PMC3070520), pp. 1115–1128.

Penzl, T. (2000). "Eigenvalue decay bounds for solutions of Lyapunov equations: the symmetric case". In: *Systems & Control Letters* 40.2, pp. 139–144.

Polanski, A., A. Bobrowski, and M. Kimmel (2003). "A note on distributions of times to coalescence, under time-dependent population size". In: *Theoretical Population Biology* 63.1, pp. 33–40.

Polanski, A. and M. Kimmel (2003). "New Explicit Expressions for Relative Frequencies of Single-Nucleotide Polymorphisms With Application to Statistical Inference on Population Growth". In: *Genetics* 165.1, pp. 427–436.

Prüfer, K. et al. (2014). "The complete genome sequence of a Neanderthal from the Altai Mountains". In: *Nature* 505.7481, pp. 43–49.

Racimo, F. et al. (2015). "Evidence for archaic adaptive introgression in humans". In: *Nat Rev Genet* 16.6, pp. 359–371.

Raghavan, M. et al. (2015). "Genomic evidence for the Pleistocene and recent population history of Native Americans". In: *Science* 349.6250, aab3884.

Rasmussen, M. D. et al. (2014). "Genome-wide inference of ancestral recombination graphs". In: *PLoS Genetics* 10.5, e1004342.

Sankararaman, S. et al. (2014). "The genomic landscape of Neanderthal ancestry in present-day humans". In: *Nature* 507.7492, pp. 354–357.

Sawyer, S. A. and D. L. Hartl (1992). "Population genetics of polymorphism and divergence." In: *Genetics* 132.4, pp. 1161–76.

Schiffels, S. and R. Durbin (2014). "Inferring human population size and separation history from multiple genome sequences". In: *Nature Genetics* 46, pp. 919–925.

Sheehan, S., K. Harris, and Y. S. Song (2013). "Estimating variable effective population sizes from multiple genomes: A sequentially Markov conditional sampling distribution approach". In: *Genetics* 194.3. (PMC3697970), pp. 647–662.

Simonsen, K. L. and G. A. Churchill (1997). "A Markov Chain Model of Coalescence with Recombination". In: *Theoretical Population Biology* 52, pp. 43–59.

Singhal, S. et al. (2015). "Stable recombination hotspots in birds". In: *Science* 350.6263, pp. 928–932. ISSN: 0036-8075. DOI: 10.1126/science.aad0843. eprint: http://science.sciencemag.org/content/350/6263/928.full.pdf. URL: http://science.sciencemag.org/content/350/6263/928.

Skoglund, P. et al. (2015). "Genetic evidence for two founding populations of the Americas". In: *Nature* 525.7567, pp. 104–108.

Staab, P. R. et al. (2015). "scrm: efficiently simulating long sequences using the approximated coalescent with recombination". In: *Bioinformatics* 31.10, pp. 1680–1682.

Steinrücken, M., J. A. Kamm, and Y. S. Song (2015). "Inference of complex population histories using whole-genome sequences from multiple populations". In: *bioRxiv*.

Steinrücken, M., J. S. Paul, and Y. S. Song (2013). "A sequentially Markov conditional sampling distribution for structured populations with migration and recombination". In: *Theoretical Population Biology* 87. (PMC3532580), pp. 51–61.

Stewart, J. R. and C. B. Stringer (2012). "Human Evolution Out of Africa: The Role of Refugia and Climate Change". In: *Science* 335.6074, pp. 1317–1321.

Szegö, G. (1939). *Orthogonal Polynomials*. 4th. American Mathematical Society.

Tataru, P., J. A. Nirody, and Y. S. Song (2014). "diCal-IBD: demography-aware inference of identity-by-descent tracts in unrelated individuals". In: *Bioinformatics* 30.23, pp. 3430–3431.

Tennessen, J. A. et al. (2012). "Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes". In: *Science* 337.6090, pp. 64–69.

Terhorst, J., J. A. Kamm, and Y. S. Song (2017). "Robust and scalable inference of population history from hundreds of unphased whole genomes". In: *Nature Genetics* 49.2, pp. 303–309. DOI: 10.1038/ng.3748.

Terhorst, J. and Y. S. Song (2015). "Fundamental limits on the accuracy of demographic inference based on the sample frequency spectrum". In: *Proceedings of the National Academy of Sciences* 112.25, pp. 7677–7682. DOI: 10.1073/pnas.1503717112.

Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer Science+Business Media, New York.

Vernot, B. and J. M. Akey (2014). "Resurrecting surviving neandertal lineages from modern human genomes". In: *Science* 343.6174, pp. 1017–1021.

Weisstein, E. W. (2017). *Trigonometric Power Formulas*. URL: http://mathworld.wolfram.com/TrigonometricPowerFormulas.html (visited on 04/15/2017).

Wilton, P. R., S. Carmi, and A. Hobolth (2015). "The SMC' Is a Highly Accurate Approximation to the Ancestral Recombination Graph". In: *Genetics* 200.1, pp. 343–355.

Wiuf, C. and J. Hein (1999). "Recombination as a point process along sequences". In: *Theoretical Population Biology* 55, pp. 248–259.

Yu, B. (1997). "Assouad, Fano, and Le Cam". In: *Festschrift for Lucien Le Cam*. Ed. by G. L. Y. David Pollard Erik Torgersen. Springer, pp. 423–435.