On the decidability of population size histories from finite allele frequency spectra

Soheil Baharian[1, 2] and Simon Gravel[1, 2]

[1]Department of Human Genetics, McGill University, Montreal, QC, Canada
[2]McGill University and Genome Quebec Innovation Centre, Montreal, QC, Canada

December 15, 2017

Abstract

Understanding the historical events that shaped current genomic diversity has applications in historical, biological, and medical research. However, the amount of historical information that can be inferred from genetic data is finite, which leads to an identifiability problem. For example, different historical processes can lead to identical distribution of allele frequencies. This identifiability issue casts a shadow of uncertainty over the results of any study which uses the frequency spectrum to infer past demography. It has been argued that imposing mild 'reasonableness' constraints on demographic histories can enable unique reconstruction, at least in an idealized setting where the length of the genome is nearly infinite. Here, we discuss this problem for finite sample size and genome length. Using the diffusion approximation, we obtain bounds on likelihood differences between similar demographic histories, and use them to construct pairs of very different reasonable histories that produce almost-identical frequency distributions. The finite-genome problem therefore remains poorly determined even among reasonable histories. Where fits to few-parameter models produce narrow parameter confidence intervals, large uncertainties lurk hidden by model assumption.

1 Introduction

Genetic variation across individuals contains information about the evolutionary and demographic history of populations. A simple and efficient summary statistic of genomic variation commonly used in inference studies of population demography is the allele frequency spectrum, describing the proportion of segregating sites as a function of the population frequency of the derived allele [1, 2, 3, 4, 5, 6]. Several computational models have been proposed to reconstruct historical population sizes that are consistent with observed allele frequency spectra [7, 8, 9, 10, 11, 12, 13]. Under the assumption of a neutral Wright-Fisher model, these inferred histories are often taken to be representative of the effective historical population sizes [14, 15, 16, 11, 17]. They are also used as baseline models to identify regions under selection [18, 19, 20] and to predict patterns of deleterious variation in human genomes [21, 22, 23].

However, Myers *et al.* [24] showed that the solution to this inference problem is not unique. To illustrate this, they constructed a family of distinct demographic histories whose frequency spectra under neutral Wright-Fisher evolution are identical for any sample size. This poses a serious practical challenge, since a demographic model that fits well the observed neutral diversity is not guaranteed to be historically accurate or to provide an appropriate model for deleterious variation.

On the other hand, Bhaskar and Song [25] have argued that the families of demographic models constructed by Myers *et al.* are not biologically realistic, because they require historical population sizes that oscillate on arbitrarily short time-scales. They proved that we can uniquely reconstruct the underlying demography from the allele frequency spectrum if (i) we limit our search to historical population sizes

that are piecewise-continuous functions of time with a given maximum number of oscillations, (ii) we have enough samples, and (iii) we assume an infinitely long genome [25].

Refs [24, 25] are both correct, but they send different messages regarding the reliability of inferred histories. Does the problem of identifiability raised by [24] bear on applied inference, or should it be considered a purely theoretical result about a class of pathological functions with little biological relevance?

In this article, we seek to resolve this question by addressing the identifiability problem in more realistic scenarios where both sample size and genome length are finite. Recent work by Terhorst and Song [6] has started to address this question by providing strict bounds on the accuracy of demographic inference based on the allele frequency spectrum. In particular, they have focused on the possibility of reconstructing history prior to a bottleneck, which is challenging because of the lost diversity (and thus lost information) during the bottleneck.

Extending this work to situations with arbitrary demographies, we argue that the problem remains poorly determined, even without bottlenecks, in the sense that vastly different population histories can produce statistically indistinguishable allele frequency spectra. Ancient history differences are most difficult to detect, as expected, but we also explain how the approach of Myers *et al.* can be modified to construct well-behaved, practically indistinguishable histories with somewhat more recent differences.

Our arguments are based on two simple observations. First, similar histories should produce similar frequency spectra. Second, the Myers *et al.* family of functions may exhibit infinitely fast oscillations, but, given the extremely small amplitude of these oscillations, they can be replaced by smooth, non-oscillating functions with tiny effect on the frequency spectrum. Thus macroscopically different histories can produce microscopic differences in the frequency distribution. These small differences could in principle be detected given an infinitely long genome, as per the Bhaskar and Song result, but they could not be detected given a finite genome of realistic length. To prove this, we first produce upper bounds on the differences between frequency spectra produced by two similar demographic models, and on the likelihood ratio between the two models given an observed frequency spectrum. Using these bounds and the family of functions given by Myers *et al.*, we construct a family of plausible demographic histories that are very distinct but practically indistinguishable.

The main practical message from this study is that any demographic inference study based on the frequency spectrum *must* have large zones of uncertainty. These can be detected, in principle, by exploring the likelihood surface over the space of all possible functions. However, most inference studies use few-parameter demographic models and estimate parameter uncertainties assuming that these models are correct. In such studies, an excellent fit with small parameter uncertainties can still mask a model that is completely wrong.

This paper is organized as follows. We present an intuitive discussion regarding identifiable demographies in Sec. 2, discussing the properties of the construction of Myers *et al.* [24]. In Sec. 3, we provide the preliminary theory necessary for our analysis. We formally derive a bound on the change in the allele frequency spectrum due to a change in the population size history in Sec. 4, both for infinite and finite genome length, and discuss the relationship with Terhorst and Song [6].

2 The diffusion approximation and the identifiability problem

The evolution of the allele frequency spectrum $P(y,t)$ for a large randomly mating population with Wright-Fisher reproduction can be modeled through a diffusion process [26], which we write as

$$\frac{\partial}{\partial t}P(y,t) = \frac{1}{2}\frac{\partial^2}{\partial y^2}\left[\frac{y(1-y)}{2N(t)}\,P(y,t)\right] + 2N(t)\mu\,\delta\left(y - \frac{1}{2N(t)}\right) \tag{1}$$

where $0 < y < 1$ is the allele frequency in the population, μ is the mutation rate per generation, $N(t)$ is the size of the population at time t (measured in generations), and $\delta(\cdot)$ is the Dirac delta function. We assume that the population size is large enough that the frequency y can be approximated by a continuous variable. In this formulation, the first term on the right-hand side describes the effect of genetic drift and

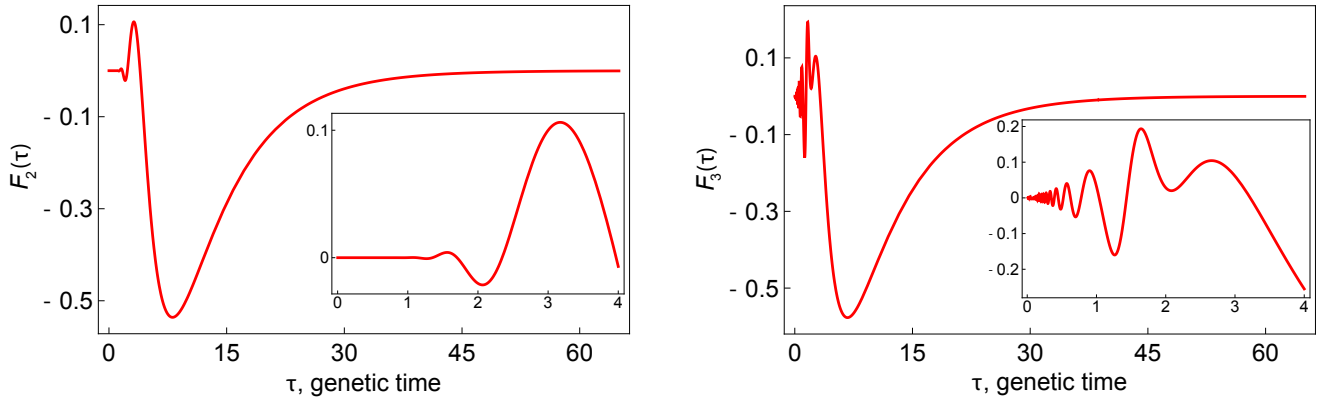

Figure 1: Two examples of a function which can be used to construct histories with identical frequency spectra. The present time is at $\tau = 0$, whereas $\tau > 0$ denotes the past. Left panel: The function $F_2(\tau)$ constructed by Myers *et al.* to illustrate the non-identifiability problem. Right panel: The function $F_3(\tau)$ discussed here. The insets show close-ups of the behavior of the corresponding functions near the present time.

the second describes that of new mutations entering the population with initial frequency $1/2N(t)$. The diffusion equation can also be written in "genetic" or "diffusion" time, $\tau(t) = \int_0^t dt'/2N(t')$, in which drift occurs at a constant rate, i.e.,

$$\frac{\partial}{\partial \tau} P(y,\tau) = \frac{1}{2}\frac{\partial^2}{\partial y^2}\big[y(1-y)P(y,\tau)\big] + \big[2\tilde{N}(\tau)\big]^2 \mu\, \delta\left(y - \frac{1}{2\tilde{N}(\tau)}\right) \tag{2}$$

where $\tilde{N}(\tau) = N(t(\tau))$.

The frequency spectrum $P(y,\tau)$ depends on $\tilde{N}(\tau)$ and therefore contains information about the population size history. The problem of identifiability of demographic histories, raised by [24], is that two different histories $\tilde{N}_1(\tau)$ and $\tilde{N}_2(\tau)$ can lead to identical frequency spectra. Concretely, Myers *et al.* [24] considered functions $\tilde{N}_2(\tau)$ of the form

$$\tilde{N}_2(\tau) = \tilde{N}_1(\tau) + \alpha N_0\, F(\tau) \tag{3}$$

where α is a constant and N_0 denotes the current population size. They showed that histories $\tilde{N}_1(\tau)$ and $\tilde{N}_2(\tau)$ would lead to identical allele frequency spectra if the function $F(\tau)$ obeys

$$\int_0^\infty F(\tau)\, e^{-\lambda_i \tau} d\tau = 0 \quad \text{with} \quad \lambda_i = (i+1)(i+2)/2 \quad \text{for} \quad i \in \{0,1,2,\dots\}. \tag{4}$$

They also showed that such functions exist and constructed an example, $F_2(\tau) = \int_0^\tau f_0(\tau - x) f_1(x) dx$ where $f_0(\tau) = \exp(-1/\tau^2)$ and $f_1(\tau) = \big[\cos(\pi^2/\tau)\exp(-\tau/8)\big]/\sqrt{\tau}$, which is displayed in the left panel of Fig. 1 (they set $\alpha = -9$ to ensure that $\tilde{N}_2(\tau)$ is strictly positive). Since $\tilde{N}_2(\tau)$ cannot be distinguished from $\tilde{N}_1(\tau)$ based on the allele frequency spectrum, the inference problem is poorly determined.

However, Bhaskar and Song [25] pointed out that adding multiples of $F_2(\tau)$ to any smooth history $\tilde{N}_1(\tau)$ leads to unrealistic population histories that oscillate increasingly rapidly as $\tau \to 0^+$. In fact, they showed that any function satisfying Eq. (4) must exhibit an infinite number of sign changes. As such, any population size history function constructed by linear combinations of $F(\tau)$ is biologically unrealistic. They further proved that there is a unique solution to the inference problem in a very general class of realistic model functions. Thus, their argument offers the reassuring message that histories can, in fact, be uniquely reconstructed if we limit ourselves to biologically plausible histories and have sufficient data.

But how much data do we need? The oscillations near $\tau = 0$ in $F_2(\tau)$ are so small that they are barely noticeable in the inset within the left panel of Fig. 1 [for example, $F_2(\tau = 0.5) \sim 10^{-12}$]. It

seems unlikely that these minuscule oscillations are relevant to our ability to reconstruct demographic histories. We expect (and show below) that replacing unrealistic small-amplitude oscillations by a realistic constant value would have an insignificant effect on the resulting spectrum. The resulting history is very different from a constant-sized history but would produce a nearly identical spectrum. If we formulate the inference problem in terms of model likelihoods, the Myers *et al.* construction shows that the likelihood surface is exactly flat along some directions in the space of all histories. Bhaskar and Song show that such flat directions are not present in the space of "reasonable" histories. We show below that the Myers *et al.* construction indicates the existence of almost-flat directions in the space of reasonable histories. For this reason, we cannot practically reconstruct history from the finite frequency spectrum alone.

The existence of almost-flat likelihood directions is not particularly surprising. Very ancient events are expected to leave few traces in the present-day allele frequency spectrum, and we cannot hope to reconstruct them from finite genetic data. In fact, the expansions and bottlenecks in the construction of Myers *et al.* (which are visible in the left panel of Fig. 1 and on Fig. 3) are ancient enough (tens of thousands of generations ago in a population of size 10,000) that they would have no statistically discernible individual effect on the present spectrum (see Appendix C).

To investigate whether the Myers *et al.* construction can provide us with 'unexpected' flat directions in the likelihood surface, we consider another instance of $F(\tau)$ with larger-amplitude oscillations near $\tau = 0$ and a more recent strong bottleneck, while still satisfying the orthogonality condition given in Eq. (4). We construct this function by following the prescription of Myers *et al.* [24] and performing the convolution of $f_1(\tau)$ given above with $f_0(\tau) = 1 - \exp\big[-25(\tau - \frac{1}{2})^2\big]$, i.e.,

$$F_3(\tau) = \int_0^\tau \left(1 - \exp\left\{-25\big[(\tau - x) - \tfrac{1}{2}\big]^2\right\}\right) \frac{\cos(\pi^2/x)\exp(-x/8)}{\sqrt{x}}\, dx\,. \tag{5}$$

The function F_3, depicted in the right panel of Fig. 1, has more prominent oscillations closer to the present time ($\tau = 0$) compared to $F_2(\tau)$ constructed by Myers *et al.*, as can be seen in the two insets. For example, $F_3(\tau = 0.5) \sim 10^{-2}$ whereas $F_2(\tau = 0.5) \sim 10^{-12}$.

To investigate the measurable effect of the fine-scale oscillations on the expected frequency spectrum, we first consider the piecewise-continuous function

$$F_{3c}(\tau) = \begin{cases} 0 & \text{for } \tau \le c \\ F_3(\tau) & \text{for } \tau > c \end{cases} \tag{6}$$

where we have introduced a cutoff $c < 1$ to remove the high-frequency oscillations near present in $F(\tau)$ given by Eq. (5). We numerically study the effect of this cutoff on the allele frequency spectrum. We use ∂a∂i [3] to compare the simulated expected allele frequency spectra of two demographies, one given by a constant population size

$$\tilde{N}_{\text{const}}(\tau) = N_0 \tag{7}$$

and one constructed using the truncated function defined in Eq. (6),

$$\tilde{N}_{3c}(\tau) = N_0\,[1 - 5F_{3c}(\tau)]. \tag{8}$$

By construction, histories $\tilde{N}_{\text{const}}(\tau)$ and

$$\tilde{N}_3(\tau) = N_0\,[1 - 5F_3(\tau)] \tag{9}$$

produce identical expected frequency spectra. Moreover, histories $\tilde{N}_3(\tau)$ and $\tilde{N}_{3c}(\tau)$ differ only for $0 \le \tau \le c$, the region dominated by the infinitely fast but small-amplitude oscillations. Given the small difference between these two population size histories, we expect the difference between the frequency spectra to be small as well. Fig. 2 shows simulated differences for different values of cutoff c. The left panel shows the expected allele frequency spectrum, highlighting the fact that different values of c produce practically indistinguishable frequency spectra, whereas the right panel depicts the relative change in the expected

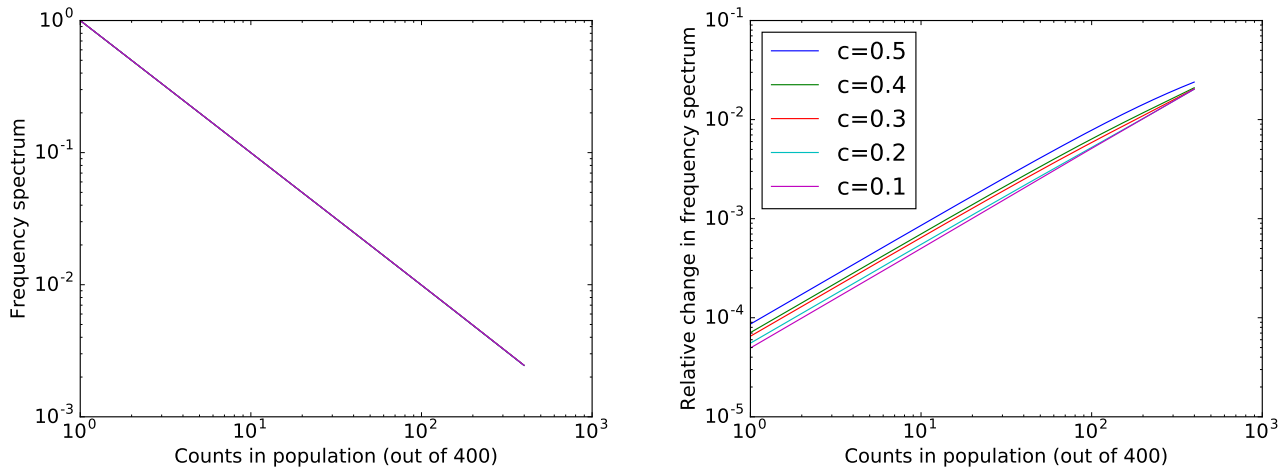

Figure 2: The effect of the cutoff c on the allele frequency spectrum, for $c \in \{0.5, 0.4, 0.3, 0.2, 0.1\}$. Simulations are done with ∂a∂i for a sampled population of size 400, and the plots are in log-log scale. Left panel: Practically identical allele frequency spectra of population histories $\tilde{N}_c(\tau)$ given by Eq. (8) for different values of c shown above. Right panel: Relative difference in the frequency spectra of $\tilde{N}_c(\tau)$ and $\tilde{N}_{\text{const}}(\tau)$ (that is, $|P_c - P_{\text{const}}| / P_{\text{const}}$) for different values of c.

frequency spectrum as a result of the introduction of the cutoff threshold. As expected, decreasing the value of c lead to smaller relative errors.

By eliminating the highly oscillatory part of the demographic model introduced in Ref. [24], one therefore arrives at a history which is biologically plausible and which leads to an allele frequency spectrum that is almost indistinguishable from that of the original history. To investigate whether such small differences could nevertheless allow for model identification given a large enough sample size, we next obtain analytical upper bounds on the difference in likelihoods between histories that are close enough to each other. Using this bound, we generalize the process described above to construct biologically realistic but very different histories that produce practically indistinguishable histories.

Given a realistic demographic history $\tilde{N}_1(\tau)$, we use the prescription of Myers *et al.* to generate a new history $\tilde{N}_3(\tau)$ whose allele frequency spectrum is identical to that of $\tilde{N}_1(\tau)$. We then construct a biologically realistic history $\tilde{N}_{3c}(\tau)$ such that $|\tilde{N}_{3c}(\tau) - \tilde{N}_3(\tau)| < \epsilon\, \tilde{N}_3(\tau)$ for all τ and a small ϵ. Finally, we use the bounds derived below to guarantee that $\tilde{N}_{3c}(\tau)$ is practically indistinguishable from $\tilde{N}_1(\tau)$ for arbitrary sample sizes.

3 Solving the diffusion equation

Before deriving the bounds discussed above, we first provide theoretical background and an overview of the derivation of Myers *et al.* in Ref. [24] which will form the basis of our calculations.

To solve the diffusion equation, we first consider $K(y, \tau_1|x, \tau_0)$, the probability density that an allele whose frequency was x at a time τ_0 has frequency y at a later time τ_1. In other words, $K(y, \tau_1|x, \tau_0)$ is the Green's function for Eq. (2) and, therefore, satisfies

$$\frac{\partial}{\partial \tau_1} K(y, \tau_1|x, \tau_0) = \frac{1}{2} \frac{\partial^2}{\partial y^2} \left[y(1-y) K(y, \tau_1|x, \tau_0) \right] + \delta(y-x)\delta(\tau_1 - \tau_0) \tag{10}$$

for $0 < x, y < 1$. The solution to this equation is given by [26, 24]

$$K(y, \tau_1|x, \tau_0) = \Theta(\tau_1 - \tau_0)\, x(1-x) \sum_{i=0}^{\infty} \frac{J_i(x) J_i(y)}{X_i} \exp[-\lambda_i(\tau_1 - \tau_0)] \tag{11}$$

where $\Theta(\cdot)$ is the Heaviside step function, $\lambda_i = (i+1)(i+2)/2$, and $J_i(\cdot)$ are expressed in terms of the Jacobi polynomials as $J_i(x) = P_i^{(1,1)}(2x-1)$ and are orthogonal according to $\int_0^1 x(1-x)J_i(x)J_j(x)dx = X_i\,\delta_{ij}$ with $X_i = (i+1)/[(i+2)(2i+3)]$. Because $K(y,\tau_1|x,\tau_0)$ is the Green's function for Eq. (2), one can write the solution to Eq. (2) as

$$\begin{aligned} P(y,\tau_1) &= \int_0^1 dx \int_{-\infty}^{\infty} d\tau_0 \left\{ \mu\big[2\tilde{N}(\tau_0)\big]^2 \delta\left(x - \frac{1}{2\tilde{N}(\tau_0)}\right)\right\} K(y,\tau_1|x,\tau_0) \\ &= \int_{-\infty}^{\infty} d\tau_0\, \mu\big[2\tilde{N}(\tau_0)\big]^2 K\big(y,\tau_1\big|1/[2\tilde{N}(\tau_0)],\tau_0\big). \end{aligned} \tag{12}$$

In other words, the present-day frequency spectrum can be obtained by integrating contributions from mutations that appeared at time τ_0.

To proceed further, following Myers *et al.* [24], one can Taylor expand $K(y,\tau_1|x,\tau_0)$ around $x=0$ to find $K(y,\tau_1|x,\tau_0) = x\,Q(y,\tau_1-\tau_0)\,\Theta(\tau_1-\tau_0) + \mathcal{O}(x^2)$ for small x where $Q(y,\tau) = \sum_{i=0}^{\infty}[J_i(0)/X_i]J_i(y)\,e^{-\lambda_i\tau}$. Therefore, assuming a large population size (i.e., $\min_{\tau_0}[\tilde{N}(\tau_0)] \gg 1$) leads to

$$P(y,\tau_1) \simeq 2\mu \int_{-\infty}^{\tau_1} d\tau_0\, \tilde{N}(\tau_0) Q(y,\tau_1-\tau_0). \tag{13}$$

To simplify notation, we now measure time backwards from present (that is, $\tau_1 = 0$ and $\tau_0 = -\tau$), define $P(y) = P(y,\tau_1=0)$, consider the rescaled population size $\tilde{n}(\tau) = \tilde{N}(-\tau)/N_0$, and, without loss of generality, measure the frequency spectrum in units of $2\mu N_0$. Hence, Eq. (13) becomes

$$P(y) = \int_0^{\infty} d\tau\, \tilde{n}(\tau) Q(y,\tau). \tag{14}$$

This is equivalent to Eq. (13) of Ref. [24], which defined the frequency spectrum up to an arbitrary constant. Using the expansion of $Q(y,\tau)$, one finds $P(y) = \sum_{i=0}^{\infty} d_i J_i(y)$ where

$$d_i = \frac{J_i(0)}{X_i} \int_0^{\infty} \tilde{n}(\tau)\, e^{-\lambda_i\tau} d\tau. \tag{15}$$

For example, a population with constant size N_0 has $P(y) = 2/y$ and $d_i = 2J_i(0)(2i+3)/(i+1)^2$.

The function $F_2(\tau)$ defined in Ref. [24] was designed to ensure that the coefficients $\{d_i\}$, and therefore the frequency spectrum, are unaffected by the change in population history shown in (3). This condition leads to orthogonality condition (4) and the identifiability problem. In the next section, we investigate what happens when condition (4) is met only approximately, as happens when we construct a smoothed version of function $F(\tau)$.

4 Bounds on the change in the frequency spectrum and likelihood

We now bound the difference between two allele frequency spectra resulting from two different histories. Consider two histories, given by $\tilde{n}'(\tau)$ and $\tilde{n}(\tau)$, and their resulting frequency spectra, $P'(y)$ and $P(y)$ respectively. Using Eq. (14), we can write

$$P'(y) - P(y) = \int_0^{\infty} d\tau \big[\tilde{n}'(\tau) - \tilde{n}(\tau)\big] Q(y,\tau) \tag{16}$$

for the difference between the two allele frequency spectra. Let $\delta\tilde{n}(\tau) = \tilde{n}'(\tau) - \tilde{n}(\tau)$ denote the difference between the two histories. Because $Q(y,\tau) \geq 0$, we find

$$\big|P'(y) - P(y)\big| = \left|\int_0^{\infty} d\tau\, \delta\tilde{n}(\tau) Q(y,\tau)\right| \leq \int_0^{\infty} |\delta\tilde{n}(\tau)|\, Q(y,\tau)\, d\tau. \tag{17}$$

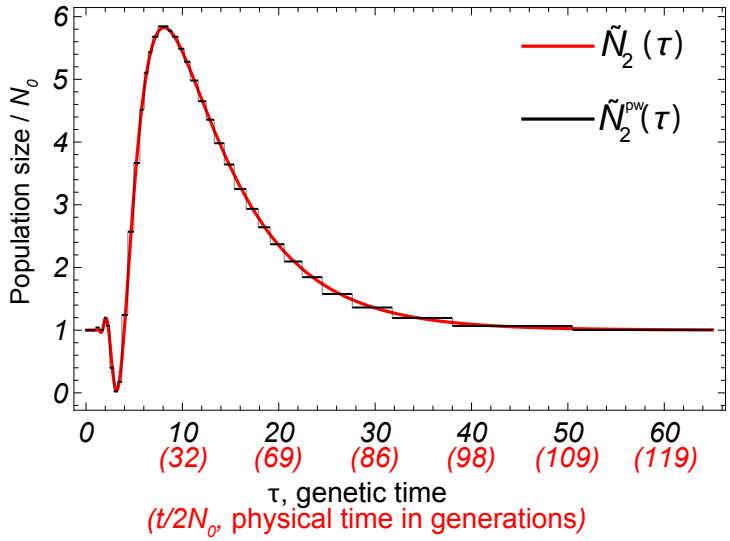

Figure 3: Population history $\tilde{N}_2(\tau)$ derived by Myers *et al.* (red, smooth) and its piecewise-constant approximation $\tilde{N}_2^{\rm pw}(\tau)$ (black, piecewise-defined) with $p = 35$ pieces. Physical times corresponding to the indicated genetic times are presented in units of $2N_0$ generations. The time axis is not linear in physical time.

If there exists a small ϵ for which

$$|\tilde{n}'(\tau) - \tilde{n}(\tau)| < \epsilon\, \tilde{n}(\tau) \tag{18}$$

for all $\tau \geq 0$, then

$$\big|P'(y) - P(y)\big| < \epsilon\, P(y). \tag{19}$$

Therefore, ϵ also determines the upper bound on the difference between the two frequency spectra.

As an example, define $F_3^{\rm pw}(\tau)$ to be an approximation of the function $F_3(\tau)$ by p piecewise-constant segments for $\tau > c$ and set $F_3^{\rm pw}(\tau) = 0$ for $\tau \leq c$ where c is an arbitrary, small number. Then consider the population size history given by the function $\tilde{N}_3^{\rm pw}(\tau) = N_0\,[1 - 9\tilde{F}_3^{\rm pw}(\tau)]$. This approximates the original model of Myers, and the approximation can be made arbitrarily accurate by choosing a large but finite number of pieces *et al.* (see Fig. 3). Equation (19) tells us that the allele frequency spectrum from the approximate, piecewise-constant history can be made arbitrarily close to the history constructed by Myers *et al.* and, therefore, and thus also close to that of a history with constant population size. In short, we have shown that two biologically reasonable but *very distinct* histories $\tilde{N}_{\rm const}(\tau) = N_0$ and $\tilde{N}_3^{\rm pw}(\tau) = N_0\,[1 - 9\tilde{F}_3^{\rm pw}(\tau)]$, which belong to the same piecewise-constant family of functions and whose difference has a finite number of sign changes, lead to very similar allele frequency spectra. Our ability to detect the differences in the two spectra will therefore depend on the sample size and the length of the genome.

For a finite sample of size M and an infinite genome, the expected allele frequency spectrum becomes

$$\phi_{M,i} = \int_0^1 \binom{M}{i} y^i (1-y)^{M-i} P(y)\, dy \quad \text{for} \quad i \in \{1, 2, \ldots, M-1\} \tag{20}$$

with $\phi_{M,i}$ representing the number of loci where i out of M chromosomes in the sample have a derived allele. For a finite genome with S unlinked polymorphic loci in the sample of size M, we can compute a likelihood by assuming that allele frequencies are distributed according to the Poisson distribution [27, 3]. The likelihood $\mathcal{L}(\{\phi_{M,i}\})$ of the frequency spectrum (and of the history that generated it) equals the probability distribution for the observed allele frequencies spectrum $\{m_i\}$:

$$\mathcal{L}(\{\phi_{M,i}\}) = p_{M,S}(\{m_i\}|\{\phi_{M,i}\}) = \prod_{i=1}^{M-1} \frac{e^{-\phi_{M,i}}\,(\phi_{M,i})^{m_i}}{m_i!} \tag{21}$$

where $S = \sum_{i=1}^{M-1} m_i$ is the number of segregating sites.

Armed with this result, we can compare the likelihoods of two models originating from two very close population size histories.

First, we calculate a bound on the change in the expected finite–sample allele frequency spectrum for any two histories that are similar according to condition (18). We write the change in the allele frequency spectrum as $\delta\phi_{M,i} = \phi'_{M,i} - \phi_{M,i} = \int_0^1 \binom{M}{i} y^i (1-y)^{M-i} \big[P'(y) - P(y)\big]\, dy$. Using Eq. (19), we find

$$|\delta\phi_{M,i}| < \epsilon\, \phi_{M,i}. \tag{22}$$

Equations (22) and (19) provide tight bounds, since there exist two population size histories that fulfil these bounds, namely $\tilde{n}'(\tau) = (1+\epsilon)\,\tilde{n}(\tau)$ and $\tilde{n}'(\tau) = (1-\epsilon)\,\tilde{n}(\tau)$. These bounds follow from the fact that finite and infinite-sample frequency spectra can be computed as integrals of the population history function with a positive kernel.

To compute the bound in likelihood differences between two models, we want to use the fact that one of the spectra $\phi_{M,i}$ is optimal, in the sense that it is expected according to the maximum-likelihood history $\tilde{n}(\tau)$. Technically, we further require that $\tilde{n}(\tau)$ is not on the boundary of allowed histories, in the sense that, for γ small enough, $\tilde{n}(\tau) \pm \gamma(\tilde{n}'(\tau) - \tilde{n}(\tau))$ are acceptable histories, and thus with a likelihood worse than that of $\tilde{n}(\tau)$.

Under these conditions, Appendix A shows that the difference in likelihood obeys the tight bound

$$|\delta \ln \mathcal{L}| < \frac{1}{2} S \epsilon^2 + \mathcal{O}\big(\epsilon^3\big). \tag{23}$$

The same result holds if we use a multinomial distribution for allele frequencies rather than a Poisson random field (see Appendix B).

Because $\tilde{n}(\tau)$ maximizes the likelihood, the quadratic form of the bound was expected. However, the prefactor allows for specific predictions. Suppose that the maximum-likelihood estimate is the constant population size $\tilde{N}_{\text{const}}(\tau) = N_0$ (i.e., $\tilde{n}(\tau) = 1$). We ask whether this can be distinguished from the history $\tilde{N}_{3c}(\tau)$ obtained by removing the recent fast oscillations from $\tilde{N}_3(\tau)$ up to $c = 0.028556$. The expected number of segregating sites for a sample of size M according to the constant history is $S \simeq 4\mu N_0 \ln M$.

The difference in population size history provided induced by the flattening corresponds to $\epsilon = 10^{-4}$.

To determine whether $\tilde{N}_{const}(\tau)$ and $\tilde{N}_{3c}(\tau)$ are distinguishable, we consider $|\delta \ln \mathcal{L}| \sim 2$ as the minimal difference in likelihood to enable reliable distinction between the two models. Assuming a uniform prior on the constant and variable models, $P(\tilde{N}_{const}(\tau)) = P(\tilde{N}_{3c}(\tau)) = \frac{1}{2}$, a Bayesian estimate for model posteriors is bounded as $\frac{1}{1+e^{|\delta \ln \mathcal{L}|}} \leq P(\tilde{N}_{3c}(\tau)|\mathbf{m}) \leq \frac{1}{1+e^{-|\delta \ln \mathcal{L}|}}$. In other words, $|\delta \ln \mathcal{L}| < 2$ means that $0.12 \leq P(\tilde{N}_{3c}(\tau)|\mathbf{m}) \leq 0.88$: If $\tilde{N}(\tau) = N_0$ is the maximum likelihood estimate, we cannot rule out $\tilde{N}(\tau) = \tilde{N}_{3c}(\tau)$.

Taking $|\delta \ln \mathcal{L}| \sim 2$ as the measure of a detectable change in the log-likelihood and using a genome-wide mutation rate of $\mu \simeq (1.4 \times 10^{-8}\,\text{bp}^{-1}\text{gen}^{-1}) \times (3.6 \times 10^9\,\text{bp}) = 50\,\text{gen}^{-1}$ [11], we find that for $N_0 = 10{,}000$, the sample size M needed to distinguish these two models is approximately 10^{43}, which is much larger than the current human population size. Similarly, given a sample of size of the entire population of 10,000 and a per-locus mutation rate of $1.4 \times 10^{-8}\,\text{bp}^{-1}\text{gen}^{-1}$, we would need a genome with length $L > 3.8 \times 10^{10}$ to identify the two spectra. This is longer than the human genome. There simply is not enough data in the human genome to allow us to resolve the difference between the two models, leaving aside the fact that for such large sample sizes and small effects, failures of the diffusion approximation might have a much larger effect than the differences that we are trying to detect between the population histories [28].

4.1 Comparison with the findings of Terhorst and Song

Terhorst and Song also proposes bounds on identifiability with finite genomes [6] and bounds on identifiability that are independent of sample size. Specifically, it shows that *any* estimator of the demographic history that is based on the frequency spectrum has an expected worst-case error proportional to $1/\log(S)$, i.e.,

there exists a demographic history $\tilde{n}(\tau)$ for which the expected τ_c-truncated error $\mathbf{E}_{\tilde{n}}\left[\int_0^{\tau_c}|\hat{n}(\tau)-\tilde{n}(\tau)|d\tau\right]$ between the true history and inferred history $\hat{n}(\tau)$ is proportional to $1/\log S$. In a sense, this appears to be stricter and more pessimistic than our bound: for a given difference in likelihood, our detectable error threshold scales as $1/\sqrt{S}$, which decreases much faster in a long genome.

The bounds have quite different interpretation, however: we are asking about the best we can possibly do, whereas Ref. [6] considers the worst-case scenario. In other words, we show that *any* two histories within an error range are indistinguishable, whereas Ref. [6] shows that *there exist* histories which are indistinguishable for a given error threshold.

This is an important difference. The $1/\log S$ bound relies on constructing a set of histories that change with S by assuming the existence of a historical bottleneck whose size ϵ decreases as $1/\log(S)$. Intuitively, a tighter bottleneck limits our ability to see through the bottleneck into previous history by reducing observable diversity. In fact, we can use the Terhorst and Song formalism to obtain even more conservative bounds by making the bottleneck size ϵ decrease faster than $1/\log(S)$. At that rate, the amount of information lost to the narrower bottleneck will not be compensated for fast enough by the increasing S. Thus differences in histories prior to the bottleneck can be made practically unconstrained for a given S. In the Terhorst and Song notation, we could choose $\delta = (M-\epsilon)/J$ in their Equation 21, then make ϵ arbitrarily small (specifically, for an arbitrarily small constant ν, we can choose ϵ small enough such that $sJ\frac{\delta}{\epsilon\log(2)}e^{-\tau_B/\epsilon} < \nu$), leading to a bound

$$\inf_{\hat{\eta}}\sup_{\eta\in\mathcal{F}_{IJ}}\mathbf{E}_{\eta}||\hat{\eta}^{n,s}-\eta||_{1,T} \geq \frac{\tau_A M(1-\frac{8(1+\nu)}{J})}{16},$$

where τ, M and J are parameters defining the family of functions considered in Ref. [6]. Selecting $J \geq 9$, and $\nu \ll 1$, this bound is positive and independent of S. It is also stricter than the $1/\log(S)$ bound of Terhorst and Song. The $1/\log S$ convergence of Ref. [6] is therefore correct, but a consequence of the particular choice of the scaling of the bottleneck size. In general, worst-case bound are dominated by unconstrained histories with very tight bottlenecks. It is not clear that these are the most relevant for practical inference.

Ref. [6] also presents, in Theorems 4 and 7, bounds given a fixed bottleneck size, which are more readily compared with the bounds presented here. In that case, the bound on the expected error decreases as $1/S$. Thus this minimax bound from [6] is less strict for long genomes than the best-case bound presented here, while being valid for a much wider range of estimators. There are many differences in the specific families of functions and estimators and the bounding strategies presented here and in [6] that could explain this difference in convergence rate.

From a practical perspective, the bound obtained here does provide a simple, clear, and practical bound for likelihood-based estimation which we could use directly to estimate the possible effect of minor fluctuations in population histories.

5 Conclusion

Demographic histories inferred from genetic data help us understand human history, evolution, and the distribution of deleterious variants. For this reason, unacknowledged errors in inference can lead to flawed conclusions in many downstream interpretations. The work of Myers *et al.* highlighted an uncontrolled source of potential error and led to more cautious interpretation of demographic inference studies [25, 5, 6]. However, it did not provide a strategy to quantify or limit this uncertainty. Bhaskar and Song [25] suggests that the uncertainty may not be as large as suggested by Myers *et al.*, once we limit our attention to biologically realistic functions.

Even though the limitation to biologically realistic functions does resolve the identifiability problem in an idealized setting, we showed that it has little bearing on practical inference; because the genome is finite, we still do not have the statistical power to distinguish between vastly different histories. The

work of Bhaskar and Song (and that of Terhorst and Song [6]) provides, in principle, a systematic way of assessing the uncertainty by considering very general families of possible histories. This is computationally demanding, however, and many studies continue to consider histories parameterized by a small number of parameters.

The main practical message from the present work is that is any inference that identifies a demographic history with high precision based on human data must rely on strong implicit or explicit assumptions, and is missing out smooth, biologically realistic alternatives. Low parameter uncertainties should be taken as a warning sign rather than a reassurance that uncertainties have been taken into account. To ensure that downstream analyses do not depend strongly on these uncertainties, validation over multiple histories consistent with the data is necessary.

Acknowledgements

This work was supported by CIHR through the Canada Research Chair program and by a Sloan Research Fellowship (to S.G.).

A Bounds on the change in log-likelihood: Poisson formulation

In this section, we derive the change in log-likelihood of a demographic model as a result of a change in the population size history. Given the allele frequency spectrum $P(y)$, the expected allele frequency spectrum in a sample of M individuals is $\phi_{M,i} = \int_0^1 \binom{M}{i} y^i (1-y)^{M-i} P(y)\, dy$ for $i \in \{1, 2, \ldots, M-1\}$. In a finite genome with $S = \sum_{i=1}^{M-1} m_i$ unlinked polymorphic loci in the sample of size M, we assume that the probability of observing frequency distribution $\{m_i\}$ is a product of Poisson distributions, i.e., $p_{M,S}(\{m_i\}|\{\phi_{M,i}\}) = \prod_{i=1}^{M-1} \frac{e^{-\phi_{M,i}} (\phi_{M,i})^{m_i}}{m_i!}$ where $\sum_{i=1}^{M-1} m_i = S$, the number of segregating sites. We write the difference in log-likelihood between the two models as

$$\begin{aligned}\delta \ln \mathcal{L} &= \ln p_{M,S}(\{m_i\}|\{\phi'_{M,i}\}) - \ln p_{M,S}(\{m_i\}|\{\phi_{M,i}\}) \\ &= \sum_{i=1}^{M-1} \left[\left(-1 + \frac{m_i}{\phi_{M,i}} \right) \delta\phi_{M,i} - \frac{1}{2} \frac{m_i}{(\phi_{M,i})^2} (\delta\phi_{M,i})^2 + \mathcal{O}\Big((\delta\phi_{M,i}/\phi_{M,i})^3 \Big) \right]. \end{aligned} \tag{24}$$

We now assume that $\phi_{M,i}$ is the frequency spectrum obtained by maximizing the likelihood function $\mathcal{L}$ with respect to the model $\tilde{n}(\tau)$; in other words, we compute the change in likelihood for models close to the maximum likelihood model. We wish to use this condition to set the linear term in $\delta\phi_{M,i}$ to zero. To this end, first note that $\phi_{M,i}$ is a linear functional of $\tilde{n}(\tau)$: we can relate finite changes $\delta\tilde{n}(\tau)$ in the demographic model to changes $\{\delta\phi_{M,i}\}_{i=1,\ldots,M}$ in the frequency spectrum as

$$\delta\phi_{M,i} = \int \frac{\delta\phi_{M,i}}{\delta\tilde{n}(\tau)} \delta\tilde{n}(\tau) d\tau \tag{25}$$

where the functional derivative

$$\frac{\delta\phi_{M,i}}{\delta\tilde{n}(\tau)} = \int_0^1 dy \frac{\delta\phi_{M,i}}{\delta P(y)} \frac{\delta P(y)}{\delta\tilde{n}(\tau)} = \int_0^1 dy \binom{M}{i} y^i (1-y)^{M-i} Q(y,\tau) \tag{26}$$

is independent of $\delta\tilde{n}(\tau)$.

If the linear term $\sum_{i=1}^{M-1} \left(-1 + \frac{m_i}{\phi_{M,i}} \right) \delta\phi_{M,i}$ is not zero, we can construct a history $\tilde{n}(\tau) + \gamma\delta\tilde{n}(\tau)$, for a small constant $|\gamma| \leq 1$, with a better likelihood than $\tilde{n}(\tau)$. This is a contradiction, and the linear term must be zero, unless the history $\tilde{n}(\tau) + \gamma\delta\tilde{n}(\tau)$ is not a valid demographic model. This can happen if $\tilde{n}(\tau)$ is constrained at the boundary of allowed histories, for example if $\tilde{n}(\tau_0) = 0$ for some τ_0, or if the space of allowed $\tilde{n}(\tau)$ is nonlinear such that two histories can be close to each other without intermediate

histories being allowed. Our result therefore requires that histories of the form $\tilde{n}(\tau) + \gamma\delta\tilde{n}(\tau)$ are valid for γ sufficiently small.

Assuming that this condition is respected, we have $\sum_{i=1}^{M-1}\left(-1 + \frac{m_i}{\phi_{M,i}}\right)\delta\phi_{M,i} = 0$ and we can write

$$\delta \ln \mathcal{L} = -\frac{1}{2}\sum_{i=1}^{M-1} m_i \left[\left(\frac{\delta\phi_{M,i}}{\phi_{M,i}}\right)^2 + \mathcal{O}\Big((\delta\phi_{M,i}/\phi_{M,i})^3\Big)\right]. \tag{27}$$

Therefore, we find

$$|\delta \ln \mathcal{L}| \leq \frac{1}{2}\sum_{i=1}^{M-1} m_i \left[\left(\frac{|\delta\phi_{M,i}|}{\phi_{M,i}}\right)^2 + \mathcal{O}\Big((|\delta\phi_{M,i}|/\phi_{M,i})^3\Big)\right] \tag{28}$$

which, with Eq. (22) and the definition of S, becomes

$$|\delta \ln \mathcal{L}| < \frac{1}{2}S\epsilon^2 + \mathcal{O}(\epsilon^3). \tag{29}$$

Here $S = \sum_i m_i$ is an observed random variable and the result holds independently of the particular random realization of the m_i. A similar bound can be obtained using Fisher Information:

A.1 Fisher Information estimate of parameter variance

If we parameterize the demographic history as $n_\lambda(\tau) = \tilde{n}(\tau) + \lambda(\tilde{n}'(\tau) - \tilde{n}(\tau))$, we can estimate the variance of parameter λ using Fisher's information matrix

$$I_\lambda = E_\lambda\left[\left(\frac{\partial \ln \mathcal{L}}{\partial \lambda}\right)^2\right],$$

where the expectation is taken at fixed λ over the possible observations.

Computing derivatives as above, and using the assumption that the m_i are independent with mean and variance $\phi_{M,i}$, we find

$$\begin{aligned} I_{\lambda=0} &= E_\lambda\left[\left(\sum_i (-1 + \frac{m_i}{\phi_{M,i}})\frac{\partial\phi_{M,i}}{\partial\lambda}\right)^2\right] \\ &= \sum_i E_\lambda\left[(-1 + \frac{m_i}{\phi_{M,i}})^2\right]\left(\frac{\partial\phi_{M,i}}{\partial\lambda}\right)^2 \\ &= \sum_i E_\lambda\left[(m_i - \phi_{M,i})^2\right]\frac{1}{\phi_{M,i}^2}\left(\frac{\partial\phi_{M,i}}{\partial\lambda}\right)^2 \\ &= \sum_i \frac{1}{\phi_{M,i}}\left(\frac{\partial\phi_{M,i}}{\partial\lambda}\right)^2 \end{aligned} \tag{30}$$

Since ϕ is a linear functional of $\tilde{n}(\tau)$ and a linear function of λ, we find $\frac{\partial\phi_{M,i}}{\partial\lambda} = \delta\phi_{M,i}$, the difference in frequency spectra generated by histories $\tilde{n}'(\tau)$ and $\tilde{n}$. Thus

$$I_{\lambda=0} = \sum_i \phi_{M,i}\left(\frac{\delta\phi_{M,i}}{\phi_{M,i}}\right)^2$$

By virtue of Equation (22), we can bound this by

$$I_{\lambda=0} \leq \sum_i \phi_{M,i}\epsilon^2 = E[S]\epsilon^2.$$

By Fisher's theorem, if $\lambda = 0$, the MLE estimator $\hat{\lambda}$ is asymptotically distributed with variance $I_{\lambda=0}^{-1}$. The bound on the log-likelihood discussed in the text roughly corresponds to asking that the standard deviation $\sigma_{\hat{\lambda}} = \sqrt{I_{\lambda=0}^{-1}} = \frac{1}{E[S]\epsilon^2} < 1/2$, such that $\lambda = 1$ lies just outside a 95% confidence interval.

B Bounds on the change in log-likelihood: Multinomial formulation

In this Appendix, we rederive the change in log-likelihood of the model as a result of a change in the population size history using a different distribution for the allele frequencies.

As discussed above, given the allele frequency spectrum $P(y)$, the expected allele frequency spectrum in a sample of M individuals is $\phi_{M,i} = \int_0^1 \binom{M}{i} y^i (1-y)^{M-i} P(y)\, dy$ for $i \in \{1, 2, \ldots, M-1\}$. We now assume that in a finite genome with $S = \sum_i m_i$ unlinked polymorphic loci in the sample of size M, the distribution of population frequencies is multinomial, i.e., $p_{M,S}(\{m_i\}|\{\bar{\phi}_{M,i}\}) = \binom{S}{m_1\, m_2 \ldots m_{M-1}} \prod_{i=1}^{M-1} (\bar{\phi}_{M,i})^{m_i}$ where $\bar{\phi}_{M,i} = \phi_{M,i} / \sum_{i=1}^{M-1} \phi_{M,i}$ is the normalized allele frequency spectrum. In other words, each locus is drawn independently from the expected distribution $\bar{\phi}_{M,i}$. Hence, the log-likelihood difference between the two models becomes

$$\begin{aligned}
\delta \ln \mathcal{L} &= \ln p_{M,S}(\{m_i\}|\{\bar{\phi}'_{M,i}\}) - \ln p_{M,S}(\{m_i\}|\{\bar{\phi}_{M,i}\}) \\
&= \sum_{i=1}^{M-1} m_i \ln \left(\frac{\bar{\phi}'_{M,i}}{\bar{\phi}_{M,i}} \right) = \sum_{i=1}^{M-1} m_i \ln \left(1 + \frac{\delta \bar{\phi}_{M,i}}{\bar{\phi}_{M,i}} \right) \\
&= \sum_{i=1}^{M-1} m_i \left[\frac{\delta \bar{\phi}_{M,i}}{\bar{\phi}_{M,i}} - \frac{1}{2} \left(\frac{\delta \bar{\phi}_{M,i}}{\bar{\phi}_{M,i}} \right)^2 + \mathcal{O}\Big((\delta \bar{\phi}_{M,i} / \bar{\phi}_{M,i})^3 \Big) \right]
\end{aligned} \tag{31}$$

where, in the last step, we used the Taylor expansion $\ln(1+x) \simeq x - x^2/2 + \mathcal{O}(x^3)$ for $x \ll 1$.

We now transform the functional dependence of $\delta \ln \mathcal{L}$ from $\delta \bar{\phi}_{M,i}$ (change in the normalized frequency spectrum) to $\delta \phi_{M,i}$ (change in the unnormalized frequency spectrum). Let $h = f/g$ where f, g, and h are three functions. For small changes in f and g, we can write the corresponding change in h as

$$h + \delta h = \frac{f + \delta f}{g + \delta g} = \frac{f}{g} \left(\frac{1 + \delta f / f}{1 + \delta g / g} \right) = h \big[1 + \delta f / f \big] \Big[1 - \delta g / g + (\delta g / g)^2 - \mathcal{O}\big((\delta g / g)^3 \big) \Big] \tag{32}$$

where we have Taylor expanded the denominator up to second order. We find

$$\begin{aligned}
\delta h / h &= \delta f / f - \delta g / g - \delta f \delta g / (f g) + (\delta g)^2 / g^2 + \mathcal{O}\Big((\delta f, \delta g)^3 \Big) \\
&= \big(\delta f / f - \delta g / g \big) \big(1 - \delta g / g \big) + \mathcal{O}\Big((\delta f, \delta g)^3 \Big).
\end{aligned} \tag{33}$$

Now, let us substitute $f \to \phi_{M,i}$ and $g \to \sum_{j=1}^{M-1} \phi_{M,j}$ (that is, $h \to \bar{\phi}_{M,i}$). Therefore, the relative change in normalized frequency, up to second order, becomes

$$\frac{\delta \bar{\phi}_{M,i}}{\bar{\phi}_{M,i}} = \left(\frac{\delta \phi_{M,i}}{\phi_{M,i}} - \frac{\sum_{j=1}^{M-1} \delta \phi_{M,j}}{\sum_{j=1}^{M-1} \phi_{M,j}} \right) \left(1 - \frac{\sum_{j=1}^{M-1} \delta \phi_{M,j}}{\sum_{j=1}^{M-1} \phi_{M,j}} \right) + \mathcal{O}\Big((\delta \phi_{M,i} / \phi_{M,i})^3 \Big). \tag{34}$$

Substituting Eq. (34) into Eq. (31) leads to the series expansion of $\delta \ln \mathcal{L}$ in powers of $\delta \phi_{M,i}$. We formally represent this expansion as $\delta \ln \mathcal{L} = \delta^{(1)} \ln \mathcal{L} + \delta^{(2)} \ln \mathcal{L} + \mathcal{O}\Big((\delta \phi_{M,i} / \phi_{M,i})^3 \Big)$ where

$$\delta^{(1)} \ln \mathcal{L} = \sum_{i=1}^{M-1} m_i \left(\frac{\delta \phi_{M,i}}{\phi_{M,i}} - \frac{\sum_{j=1}^{M-1} \delta \phi_{M,j}}{\sum_{j=1}^{M-1} \phi_{M,j}} \right) \tag{35}$$

and

$$\begin{aligned}
\delta^{(2)} \ln \mathcal{L} &= -\sum_{i=1}^{M-1} m_i \left[\left(\frac{\delta\phi_{M,i}}{\phi_{M,i}} - \frac{\sum_{j=1}^{M-1} \delta\phi_{M,j}}{\sum_{j=1}^{M-1} \phi_{M,j}} \right) \left(\frac{\sum_{j=1}^{M-1} \delta\phi_{M,j}}{\sum_{j=1}^{M-1} \phi_{M,j}} \right) + \frac{1}{2} \left(\frac{\delta\phi_{M,i}}{\phi_{M,i}} - \frac{\sum_{j=1}^{M-1} \delta\phi_{M,j}}{\sum_{j=1}^{M-1} \phi_{M,j}} \right)^2 \right] \\
&= -\frac{1}{2} \sum_{i=1}^{M-1} m_i \left(\frac{\delta\phi_{M,i}}{\phi_{M,i}} - \frac{\sum_{j=1}^{M-1} \delta\phi_{M,j}}{\sum_{j=1}^{M-1} \phi_{M,j}} \right) \left(\frac{\delta\phi_{M,i}}{\phi_{M,i}} + \frac{\sum_{j=1}^{M-1} \delta\phi_{M,j}}{\sum_{j=1}^{M-1} \phi_{M,j}} \right) \\
&= -\frac{1}{2} \sum_{i=1}^{M-1} m_i \left[\left(\frac{\delta\phi_{M,i}}{\phi_{M,i}} \right)^2 - \left(\frac{\sum_{j=1}^{M-1} \delta\phi_{M,j}}{\sum_{j=1}^{M-1} \phi_{M,j}} \right)^2 \right]
\end{aligned} \tag{36}$$

are, respectively, the first- and second-order changes in the log-likelihood expressed in terms of changes in the unnormalized frequency spectrum.

Next, we relate the relative change in the normalized frequency spectrum, given by Eq. (34), to a change in the demographic model. As shown in Appendix A, the change in frequency spectrum $\delta\phi_{M,i}$ depends only linearly on the change in the demographic model $\delta\tilde{n}(\tau)$. This fact allows us to use Eqs. (31), (34), and (25) to derive the change in the log-likelihood function in terms of the change in the demographic model. In other words, $\delta^{(1)} \ln \mathcal{L}$ and $\delta^{(2)} \ln \mathcal{L}$ also represent respectively the first- and second-order changes in the log-likelihood as a result of a small change in the demographic model $\delta\tilde{n}(\tau)$.

We now assume, as we did in Appendix A, that $\phi_{M,i}$ is the frequency spectrum obtained by maximizing the likelihood function $\mathcal{L}$ with respect to the model $\tilde{n}(\tau)$. Therefore, the first-order change in the log-likelihood function $\delta^{(1)} \ln \mathcal{L}$ due to the change $\delta\tilde{n}(\tau)$ around the best-fit model should vanish at the extremum, that is, $\delta^{(1)} \ln \mathcal{L} = 0$, and using Eq. (36), we can write

$$\begin{aligned}
\delta \ln \mathcal{L} &= \delta^{(2)} \ln \mathcal{L} + \mathcal{O}\Big([\delta\tilde{n}(\tau)/\tilde{n}(\tau)]^3 \Big) \\
&= -\frac{1}{2} \sum_{i=1}^{M-1} m_i \underbrace{\left[\left(\frac{\delta\phi_{M,i}}{\phi_{M,i}} \right)^2 - \left(\frac{\sum_{j=1}^{M-1} \delta\phi_{M,j}}{\sum_{j=1}^{M-1} \phi_{M,j}} \right)^2 \right]}_{\Delta_i} + \mathcal{O}\Big([\delta\tilde{n}(\tau)/\tilde{n}(\tau)]^3 \Big).
\end{aligned} \tag{37}$$

To derive the bounds on the change in log-likelihood, we now find the following bounds on Δ_i (defined in the previous equation)

$$\begin{aligned}
|\Delta_i| = \left| \left(\frac{\delta\phi_{M,i}}{\phi_{M,i}} \right)^2 - \left(\frac{\sum_{j=1}^{M-1} \delta\phi_{M,j}}{\sum_{j=1}^{M-1} \phi_{M,j}} \right)^2 \right| &\leq \max \left[\left(\frac{\delta\phi_{M,i}}{\phi_{M,i}} \right)^2, \left(\frac{\sum_{j=1}^{M-1} \delta\phi_{M,j}}{\sum_{j=1}^{M-1} \phi_{M,j}} \right)^2 \right] \\
&\leq \max \left[\left(\frac{|\delta\phi_{M,i}|}{\phi_{M,i}} \right)^2, \left(\frac{\sum_{j=1}^{M-1} |\delta\phi_{M,j}|}{\sum_{j=1}^{M-1} \phi_{M,j}} \right)^2 \right].
\end{aligned} \tag{38}$$

Using Eq. (22), both terms on the right-hand side are smaller than ϵ^2, so we can write

$$|\Delta_i| < \epsilon^2. \tag{39}$$

Finally, substituting (39) in Eq. (37) leads to the bound on the change in log-likelihood function

$$|\delta \ln \mathcal{L}| < \frac{1}{2} S \epsilon^2 + \mathcal{O}\big(\epsilon^3\big) \tag{40}$$

which is identical to that derived using the Poisson distribution for the frequencies.

C Effect of bottlenecks

Here, we investigate the effect of the existence of a bottleneck in the demography, e.g., one given by (8), on the resulting allele frequency spectrum. Our motivation is to determine whether a bottleneck would leave a measurable impact on the present-day allele frequency spectrum given the timing (recent or past) of the bottleneck. In Fig. 4, we show (black, solid line) a demographic history constructed from the history given in Eq. (8), with the cutoff $c = 0.5$, by removing all bottlenecks; the original history is shown in red, dashed line. The resulting frequency spectrum (obtained using ∂a∂i), shown in Fig. 5, indicates that

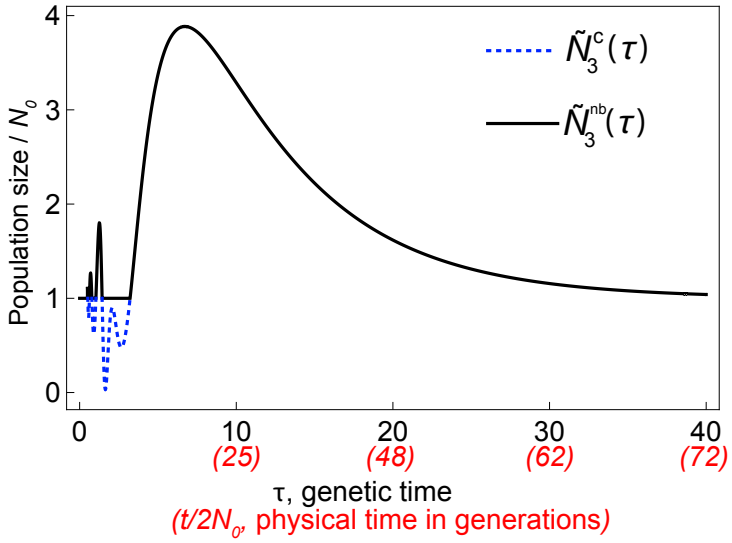

Figure 4: A population size history $\tilde{N}_3^{\mathrm{nb}}(\tau)$ derived by removing bottlenecks from $\tilde{N}_{3c}(\tau)$ for $c = 0.5$ with the bottlenecks removed. Removed bottlenecks in the original history are plotted in blue (dotted lines). Physical times corresponding to the indicated genetic times for $\tilde{N}_{3c}(\tau)$ are presented in units of $2N_0$ generations. The time axis is not linear in physical time.

removing the recent bottlenecks (which are more recent compared to those in the example presented in Myers *et al.* [24]) indeed leaves a detectable effect on the spectrum, especially for the common variants. Performing the same analysis using the example given by Myers *et al.* [24] leads to negligible difference in the frequency spectrum (results not shown for brevity), mainly due to the fact that the bottlenecks are in distant past and have no discernible effect on the frequency spectrum observed at present.

References

[1] R. C. Griffiths. The frequency spectrum of a mutation, and its age, in a general diffusion model. *Theoretical Population Biology*, 64(2):241–251, 2003.

[2] Gabor T. Marth, Eva Czabarka, Janos Murvai, and Stephen T. Sherry. The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. *Genetics*, 166(1):351–372, 2004.

[3] Ryan N. Gutenkunst, Ryan D. Hernandez, Scott H. Williamson, and Carlos D. Bustamante. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLoS Genetics*, 5(10):e1000695, October 2009.

[4] Sergio Lukić, Jody Hey, and Kevin Chen. Non-equilibrium allele frequency spectra via spectral methods. *Theoretical Population Biology*, 79(4):203–219, 2011.

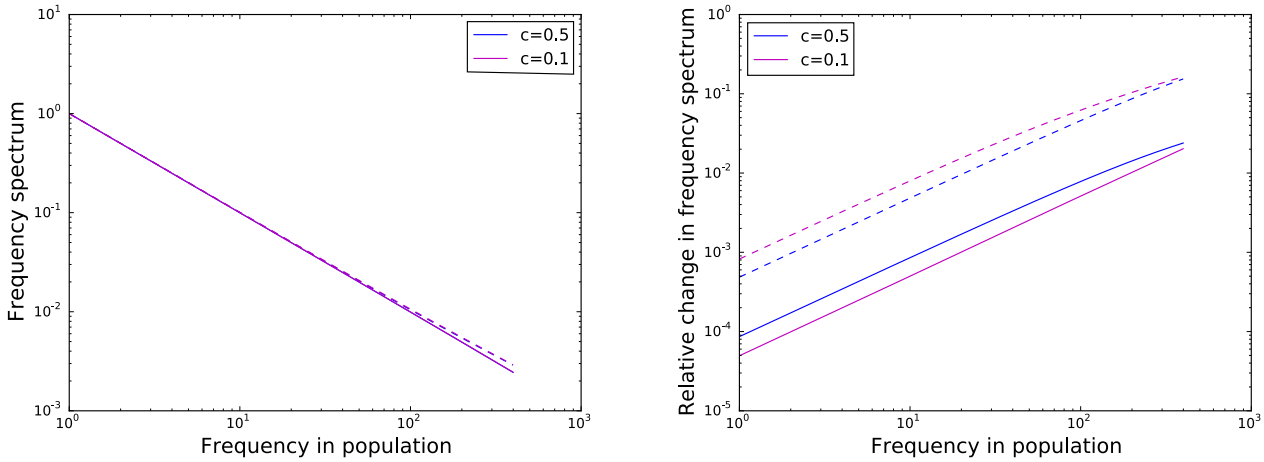

Figure 5: The effect of the removal of bottlenecks on the allele frequency spectra for different values of the cutoff c in two scenarios: (I) using the original history $\tilde{N}_c(\tau)$ defined in Eq. (8) (solid lines), and (II) using the aforementioned history without the bottlenecks, as discussed in Fig. 4 (dashed lines). Simulations are performed with ∂a∂i for a sampled population of size 400. Left panel: Removing bottlenecks causes an observable change in the frequency spectrum, especially for more common variants. Note that in both cases (that is, with or without bottlenecks) the effect of the small cutoff c on the spectrum is practically negligible. Right panel: Relative change in frequency spectrum (measured with respect to that of a population of constant size).

[5] Junhyong Kim, Elchanan Mossel, Miklós Z. Rácz, and Nathan Ross. Can one hear the shape of a population history? *Theoretical Population Biology*, 100:26–38, 2014.

[6] Jonathan Terhorst and Yun S. Song. Fundamental limits on the accuracy of demographic inference based on the sample frequency spectrum. *Proceedings of the National Academy of Sciences*, 112(25):7677–7682, 2015.

[7] Benjamin F. Voight, Alison M. Adams, Linda A. Frisse, Yudong Qian, Richard R. Hudson, and Anna Di Rienzo. Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proceedings of the National Academy of Sciences*, 102(51):18508–18513, 2005.

[8] Joseph K. Pickrell and Jonathan K. Pritchard. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, 8(11):e1002967, 2012.

[9] Sergio Lukić and Jody Hey. Demographic inference using spectral methods on snp data, with an analysis of the human out-of-africa expansion. *Genetics*, 192(2):619–639, 2012.

[10] Laurent Excoffier, Isabelle Dupanloup, Emilia Huerta-Sánchez, Vitor C. Sousa, and Matthieu Foll. Robust demographic inference from genomic and snp data. *PLoS Genetics*, 9(10):e1003905, 2013.

[11] Simon Gravel, Fouad Zakharia, Andres Moreno-Estrada, Jake K. Byrnes, Marina Muzzio, Juan L. Rodriguez-Flores, Eimear E. Kenny, Christopher R. Gignoux, Brian K. Maples, Wilfried Guiblet, Julie Dutil, Marc Via, Karla Sandoval, Gabriel Bedoya, Taras K. Oleksyk, Andres Ruiz-Linares, Esteban G. Burchard, Juan Carlos Martinez-Cruzado, Carlos D. Bustamante, and The 1000 Genomes Project. Reconstructing native american migrations from whole-genome and whole-exome data. *PLoS Genetics*, 9(12):e1004023, 12 2013.

[12] Feng Gao and Alon Keinan. Inference of super-exponential human population growth via efficient computation of the site frequency spectrum for generalized models. *Genetics*, 115:180570, 2015.

[13] Julien Jouganous, Will Long, Aaron P Ragsdale, and Simon Gravel. Inferring the Joint Demographic History of Multiple Populations: Beyond the Diffusion Approximation. *Genetics*, page 117.200493, 2017.

[14] Simon Gravel, Brenna M. Henn, Ryan N. Gutenkunst, Amit R. Indap, Gabor T. Marth, Andrew G. Clark, Fuli Yu, Richard A. Gibbs, The 1000 Genomes Project, and Carlos D. Bustamante. Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences*, 108(29):11983–11988, 2011.

[15] Jacob A. Tennessen, Abigail W. Bigham, Timothy D. O'Connor, Wenqing Fu, Eimear E. Kenny, Simon Gravel, Sean McGee, Ron Do, Xiaoming Liu, Goo Jun, Hyun Min Kang, Daniel Jordan, Suzanne M. Leal, Stacey Gabriel, Mark J. Rieder, Goncalo Abecasis, David Altshuler, Deborah A. Nickerson, Eric Boerwinkle, Shamil Sunyaev, Carlos D. Bustamante, Michael J. Bamshad, Joshua M. Akey, and Broad GO, and Seattle GO and on behalf of the NHLBI Exome Sequencing Project. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*, 337(6090):64–69, 2012.

[16] Alon Keinan and Andrew G. Clark. Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science*, 336(6082):740–743, 2012.

[17] Elodie Gazave, Li Ma, Diana Chang, Alex Coventry, Feng Gao, Donna Muzny, Eric Boerwinkle, Richard A. Gibbs, Charles F. Sing, Andrew G. Clark, and Alon Keinan. Neutral genomic regions refine models of recent rapid human population growth. *Proceedings of the National Academy of Sciences*, 111(2):757–762, 2014.

[18] Scott H. Williamson, Ryan Hernandez, Adi Fledel-Alon, Lan Zhu, Rasmus Nielsen, and Carlos D. Bustamante. Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proceedings of the National Academy of Sciences*, 102(22):7882–7887, 2005.

[19] Kirk E. Lohmueller, Amit R. Indap, Steffen Schmidt, Adam R. Boyko, Ryan D. Hernandez, Melissa J. Hubisz, John J. Sninsky, Thomas J. White, Shamil R. Sunyaev, Rasmus Nielsen, Andrew G. Clark, and Carlos D. Bustamante. Proportionally more deleterious genetic variation in european than in african populations. *Nature*, 451(7181):994–997, 2008.

[20] Roy Ronen, Nitin Udpa, Eran Halperin, and Vineet Bafna. Learning natural selection from the site frequency spectrum. *Genetics*, 195(1):181–193, 2013.

[21] Yuval B Simons, Michael C Turchin, Jonathan K Pritchard, and Guy Sella. The deleterious mutation load is insensitive to recent population history. *Nature Genetics*, 46:220–224, 2014.

[22] Ron Do, Daniel Balick, Heng Li, Ivan Adzhubei, Shamil Sunyaev, and David Reich. No evidence that selection has been less effective at removing deleterious mutations in europeans than in africans. *Nature Genetics*, 47:126–131, 2015.

[23] Simon Gravel. When Is Selection Effective? *Genetics*, 203(1):451–462, May 2016.

[24] Simon Myers, Charles Fefferman, and Nick Patterson. Can one learn history from the allelic spectrum? *Theoretical Population Biology*, 73(3):342–348, January 2008.

[25] Anand Bhaskar and Yun S. Song. Descartes' rule of signs and the identifiability of population demographic models from genomic variation data. *The Annals of Statistics*, 42(6):2469–2493, 2014.

[26] Motoo Kimura. Diffusion models in population genetics. *Journal of Applied Probability*, 1(2):177–232, 1964.

[27] Stanley A Sawyer and Daniel L Hartl. Population genetics of polymorphism and divergence. *Genetics*, 132:1161–1176, 1992.

[28] Anand Bhaskar, Andrew G Clark, and Yun S. Song. Distortion of genealogical properties when the sample is very large. *Proceedings of the National Academy of Sciences*, 111(6):2385–2390, 2014.