perm.pub/dsi:krP1W87DCQajl5bJABhHW8HLq7c/0.1
Additional formats and editions available online.

Edition 0.1

Author date: 2023-03-03

Archive date: 2023-05-30

A Simple Estimator of Lineal Admixture Time

E. Castedo Ellerman (castedo@castedo.com)

Citation:
E. Castedo Ellerman (2023) "A Simple Estimator of Lineal Admixture Time" perm.pub https://perm.pub/dsi: krP1W87DCQajl5bJABhHW8HLq7c/0.1

Copyright:
creativecommons.org/licenses/by/4.0/
2023 © The Authors. This document is distributed under a Creative Commons Attribution 4.0 International license.

Abstract

STAGE: Early Draft

DOCUMENT TYPE: Mathematical Results

OBJECTIVES

- Outline simple stochastic process of gametic lineages
- Deduce estimator of lineal admixture generation number

Introduction

Lineal admixture time is a microscale measure of admixture time. In this document, we derive an estimator of lineal admixture time. This derivation is from a stochastic process where the underlying random object is a gametic lineage.

A simple stochastic process

We define a simple stochastic process with the following properties:

- discrete regular time steps,
- infinite population size,
- proportion α_i from i-th ancestral group,
- stationary, and
- underlying random object is a gametic lineage.

We further define an "island subpopulation" of individuals for which at each time step:

1. there is random mating and then
2. immigration such that ϕ of the population is new non-admixed immigrants

The motivation for this stochastic process is to derive an estimator of lineal admixture time.

Lineal admixture time requires that there be a time horizon at which all individuals can be categorized into separate source populations.

Since the stochastic process in this document is stationary, this condition of full categorization of all lineages into ancestral source populations is only achieved asymptotically. In other words, for any lineage, there exists some time prior to which all gametes in the lineage are within one of the source populations. (Formally, this condition requires some measurable subset of exceptions, but of zero probability.)

Main Result

Formal notation

From the underlying random gametic lineage, two random variables are defined by the stochastic process:

- M_t as lineal admixture time (or generation number) and
- A_t as ancestral origin, represented as one-hot vector.

For notational brevity we use the notation

$$\mathrm{P}_t\{\ldots\}$$

and

$$\mathrm{E}_t[\ldots]$$

to denote probability and expectation, respectively, of expressions where M and A denote M_t and A_t respectively.

Base Facts

From the definition of the "island subpopulation" we deduce that

$$\mathrm{P}_{t+1}\{M = 0 \wedge A = \mathrm{e}_i\} = \phi\alpha_i + (1-\phi)\,\mathrm{P}_t\{M = 0 \wedge A = \mathrm{e}_i\}^2$$

From the definition of lineal admixture generation number,

$$\begin{aligned}\mathrm{E}_{t+1}[M] = {} & (\mathrm{E}_t[M|M>0]+1)\,\mathrm{P}_t\{M>0\}^2\,(1-\phi) \\ & + 2\left(\frac{1}{2}\,\mathrm{E}_t[M|M>0]+1\right)\mathrm{P}_t\{M>0\}\,\mathrm{P}_t\{M=0\}\,(1-\phi) \\ & + \left(\mathrm{P}_t\{M=0\}^2 - \sum_i \mathrm{P}_t\{M=0\wedge A=e_i\}^2\right)(1-\phi)\end{aligned}$$

Derivation

Given the assumptions of stationarity, we can define:

Let $x_i := \mathrm{P}_t\{M = 0 \wedge A = \mathrm{e}_i\}$.

$$x_i = \phi\alpha_i + (1-\phi)x_i^2$$

where

$$\vec{\alpha} := \mathrm{E}[A]$$

By theorem 1, the only quadratic solution for x_i is

$$x_i = \frac{1 - \sqrt{1 - 4\phi(1-\phi)\alpha_i}}{2(1-\phi)}$$

Let $q := \mathrm{P}_t\{M = 0\}$, thus

$$q = \phi + (1-\phi)\sum_i x_i^2$$

We define

$$\mu := \mathrm{E}_t[M]$$

which is the expected lineal admixture time (and generation number).

Given the base facts, we make the following deduction using the newly defined variables μ, ϕ and x_i.

$$\begin{aligned}
\mu &= (\mathrm{E}[M|M>0]+1)(1-q)^2(1-\phi) \\
&\quad + 2\left(\frac{1}{2}\mathrm{E}_t[M|M>0]+1\right)(1-q)q(1-\phi) \\
&\quad + \left(q^2 - \sum_i x_i^2\right)(1-\phi) \\
&= \mu(1-q)(1-\phi) + (1-q)^2(1-\phi) \\
&\quad + \mu q(1-\phi) + 2(1-q)q(1-\phi) \\
&\quad + \left(q^2 - \sum_i x_i^2\right)(1-\phi) \\
&= \mu(1-\phi) + ((1-q)+q)^2(1-\phi) - (1-\phi)\sum_i x_i^2 \\
0 &= -\mu\phi + 1 - q \\
\mu &= \frac{1-q}{\phi}
\end{aligned}$$

Replacing q gets

$$\mu = \frac{1-\phi}{\phi}\left(1 - \sum_i x_i^2\right)$$

We conjecture that this formula serves as a consistent maximum likelihood estimator.

Estimation of ϕ

Let $\dot{\alpha}_i$ denote the expected frequency of an i-th ancestral source. Let $\ddot{\alpha}_{i,j}$ denote the frequency of a diploid genotype with an i-th maternal ancestral source and j-ith paternal ancestral source.

Thus

$$\begin{aligned}
\ddot{\alpha}_{i,i} &= \phi\dot{\alpha}_i + (1-\phi)\dot{\alpha}_i^2 \\
&= \phi\dot{\alpha}_i(1-\dot{\alpha}_i) + \dot{\alpha}_i^2 \\
\phi &= \frac{\ddot{\alpha}_{i,i} - \dot{\alpha}_i^2}{\dot{\alpha}_i(1-\dot{\alpha}_i)}
\end{aligned}$$

Consider the case of only two ancestral sources. With $\beta := \ddot{\alpha}_{0,1} + \ddot{\alpha}_{1,0}$ we deduce that

$$\ddot{\alpha}_{0,0} + \ddot{\alpha}_{1,1} = 1 - \beta$$

$$\dot{\alpha}_1 = 1 - \dot{\alpha}_0$$

$$\dot{\alpha}_0^2 + \dot{\alpha}_1^2 = 1 - 2\dot{\alpha}_0(1 - \dot{\alpha}_0)$$

$$\begin{aligned}\phi &= \frac{\phi + \phi}{2} \\ &= \frac{\ddot{\alpha}_{0,0} + \ddot{\alpha}_{1,1} - \dot{\alpha}_0^2 - \dot{\alpha}_1^2}{2\dot{\alpha}_0(1 - \dot{\alpha}_0)} \\ &= 1 - \frac{\beta}{2\dot{\alpha}_0(1 - \dot{\alpha}_0)}\end{aligned}$$

This form is the same as the *inbreeding coefficient* but with ancestral source as the allele state rather than haplotype.

Theorem 1

The solution to x_t given stationarity etc... can not be

$$x_i = \frac{1 + \sqrt{1 - 4\phi(1 - \phi)\alpha_i}}{2(1 - \phi)}$$

when $\alpha_i < 1$.

PROOF

Assume the contrary. Since $\phi < 1$ and $x_i \leq 1$, we have

$$\begin{aligned}1 &\geq \frac{1 + \sqrt{1 - 4\phi(1 - \phi)\alpha_i}}{2(1 - \phi)} \\ 2(1 - \phi) &\geq 1 + \sqrt{1 - 4\phi(1 - \phi)\alpha_i} \\ 1 - 2\phi &\geq \sqrt{1 - 4\phi(1 - \phi)\alpha_i} \\ 1 - 4\phi + 4\phi^2 &\geq 1 - 4\phi(1 - \phi)\alpha_i \\ -4\phi(1 - \phi) &\geq -4\phi(1 - \phi)\alpha_i \\ 1 &\leq \alpha_i\end{aligned}$$

which can not be true given $\alpha_i < 1$.