

Some stuff regarding quantitative genetics from a coalescent perspective

Evan Koch, Joe Marcus

November 3, 2016

① Introduction

② Mathematical results

- A general moment generating function

- Connection to previous results

- A central limit theorem to get the infinitesimal model

③ Further research

- Is this just trivial?

- Alternate mutational distributions

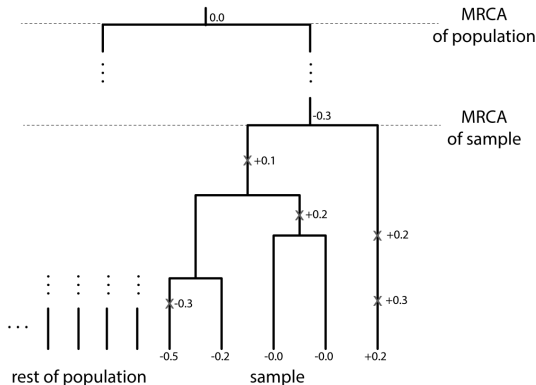
$$Q_{ST} = \frac{V_{among}}{V_{among} + 2V_{within}}$$

$$F_{ST} = \frac{t_{species} - t_{within}}{t_{species}}$$

- Q_{ST} is a commonly used statistic to test for excessive (or excessively uniform) phenotypic differentiation among populations.
- Showed that the expectations of Q_{ST} and F_{ST} are equal
- Didn't really make that much sense

Schraiber and Landis approach

Schraiber and Landis (2015) developed a rigorous coalescent approach to studying the sampling distribution of quantitative traits based on the coalescent.



Schraiber and Landis approach

- Main result is a characteristic function for the sampling distribution of a quantitative trait under the standard coalescent model for an arbitrary stepping stone mutational model.
- Derive expected moments from this
- Derive limiting distribution when the mutational effects follow a power law

Generating functions for coalescent models have been examined in some detail

Lohse, Harrison, and Barton (2011)

Derive generating functions for

- Structured coalescent
- Rate changes
- Linked loci

Extending Schraiber and Landis (2015) results to arbitrary genealogical distributions

- T** : random vector containing lengths of all possible branches in the genealogy
(ex: $\mathbf{T} = \{T_a, T_b, T_c, T_{a,b}, T_{a,c}, T_{b,c}\}$).
- O** : all possible configurations that coalescent branches can subtend
(ex: $\mathcal{O} = \{\{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}\}$).
- Y** : quantitative trait 'values' in the sampled individuals
(ex: $\mathbf{Y} = \{Y_a, Y_b, Y_c\}$).

The moment generating function for trait values

$$\varphi_{\mathbf{Y}}(\mathbf{k}) = E \left[e^{\mathbf{k} \cdot \mathbf{Y}} \right] = \int e^{\mathbf{k} \cdot \mathbf{Y}} P(\mathbf{Y} = \mathbf{y}) d\mathbf{y},$$

k is a vector a dummy variables for each sample.

Conditioning on the genealogy...

$$\varphi_{\mathbf{Y}}(\mathbf{k}) = \int e^{\mathbf{k} \cdot \mathbf{Y}} \int P(\mathbf{Y} = \mathbf{y} | \mathbf{T} = \mathbf{t}) P(\mathbf{T} = \mathbf{t}) d\mathbf{t} d\mathbf{y} \quad (1)$$

$$= \int \int e^{\mathbf{k} \cdot \mathbf{Y}} P(\mathbf{Y} = \mathbf{y} | \mathbf{T} = \mathbf{t}) d\mathbf{y} P(\mathbf{T} = \mathbf{t}) d\mathbf{t} \quad (2)$$

Splitting trait change across branches

$$Y_i = \sum_{\omega \in \mathcal{O}} Y_{i,\omega}$$

$Y_{i,\omega}$ are conditionally independent given \mathbf{T} .

$$\mathbf{k} \cdot \mathbf{y} = \sum_{\omega \in \mathcal{O}} \left(\sum_{i \in \omega} k_i y_{i,\omega} \right)$$

$$P(\mathbf{Y} = \mathbf{y} | \mathbf{T} = \mathbf{t}) = \prod_{\omega \in \mathcal{O}} P(\mathbf{Y}_\omega = (y_{i,\omega})_{i \in \omega} | \mathbf{T} = \mathbf{t}).$$

$$\int e^{\mathbf{k} \cdot \mathbf{y}} P(\mathbf{Y} = \mathbf{y} | \mathbf{T} = \mathbf{t}) d\mathbf{y} = \prod_{\omega \in \mathcal{O}} \int \exp \left(\sum_{i \in \omega} k_i y_{i,\omega} \right) P(\mathbf{Y}_\omega = (y_{i,\omega})_{i \in \omega} | \mathbf{T} = \mathbf{t}) d(y_{i,\omega})_{i \in \omega}.$$

Two important moment generating function facts

mgf of compound poisson process

$$\exp(\lambda t(\psi(k) - 1))$$

λ is the rate that events occur and ψ is the distribution that the effects of events are drawn from.

mgf of completely correlated random variables

If $X_1 = X_2$ then the mgf of their joint distribution is $\varphi_{X_1}(k_1 + k_2)$.

This is relevant because the effects of mutations along branches are shared by all descendants.

The mgf for a general distribution of genealogies

$$\varphi_{\mathbf{T}}(\mathbf{k}) = \prod_{\omega \in \mathcal{O}} \int \exp \left(\frac{\theta}{2} t_{\omega} \left(\psi \left(\sum_{a \in \omega} k_a \right) - 1 \right) \right) P(\mathbf{T} = \mathbf{t}) d\mathbf{t}.$$

This is simply the moment generating function for the genealogy \mathbf{T} with $\frac{\theta}{2} (\psi(\sum_{i \in \omega} k_{\omega}) - 1)$ substituted for the dummy variable of branch T_{ω} . Or,

$$\varphi_{\mathbf{T}}(\mathbf{s}) \Big|_{s_{\omega} = \frac{\theta}{2} (\psi(\sum_{a \in \omega} k_a) - 1)}$$

Would multiply by L to get a trait affected by L different loci.

Lohse et al. 2011

$$\varphi_T(\mathbf{s}) = \frac{\sum_i \lambda_i \varphi_i(\mathbf{s})}{\sum_i \lambda_i - \sum_{|\omega|=1} s_\omega}.$$

Schraiber and Landis 2015

$$\varphi_Y(\mathbf{k}) = \frac{2}{n(n-1) - \theta(\sum_{u=1}^n \psi(k_u) - n)} \sum_{u < v} \varphi_{n-1}(\mathbf{k}^{(u,v)})$$

Can obtain 2nd result from substitution into the first.

An example in structured populations

$$\varphi_{\mathbf{T}}^{\Omega}(\mathbf{s}) = \left(\sum_{i=1}^M \binom{|\Omega_i|}{2} \eta_i + \sum_{(i,j): i \neq j} m_{i,j} |\Omega_i| - \sum_{i=1}^M \sum_{\omega \in \Omega_i} s_{\omega} \right)^{-1} \\ \times \left(\sum_{i=1}^M \eta_i \sum_{(a,b) \in \Omega_i: a \neq b} \varphi_{\mathbf{T}}^{\Omega(i:a \cup b)}(\mathbf{s}) + \sum_{(i,j): i \neq j} m_{i,j} \sum_{\omega \in \Omega_i} \varphi_{\mathbf{T}}^{\Omega(i: -\omega, j: +\omega)}(\mathbf{s}) \right).$$

Central limit theorem

$$\varphi_T(\mathbf{s}) = \int \exp\left(\sum_{\omega \in \mathcal{O}} s_\omega t_\omega\right) P(\mathbf{T} = \mathbf{t}) d\mathbf{t}.$$

Taking a Taylor series we get

$$\begin{aligned} & \left(\int \exp\left(\sum_{\omega \in \mathcal{O}} s_\omega t_\omega\right) P(\mathbf{T} = \mathbf{t}) d\mathbf{t} \right)^L = \\ & \left(1 + \sum_{\omega \in \mathcal{O}} s_\omega E[t_\omega] + \sum_{\omega_1 \neq \omega_2} s_{\omega_1} s_{\omega_2} E[t_{\omega_2} t_{\omega_1}] + \sum_{\omega \in \mathcal{O}} \frac{s_\omega^2 E[t_\omega^2]}{2} + \dots \right)^L. \end{aligned}$$

Central limit theorem

Also taking the Taylor series of ψ and making the appropriate substitution we get

$$\left(1 + \frac{\theta}{2} \sum_{\omega \in \mathcal{O}} E[t_\omega] \left(m \left(\sum_{a \in \omega} k_a \right) + \frac{\tau^2}{2} \left(\sum_{a \in \omega} k_a \right)^2 + \dots \right) \right)^L.$$

Taking the limit such that $Lm \rightarrow \mu$ and $L\tau^2 \rightarrow \sigma^2$ as $L \rightarrow \infty$ we get

$$\exp \left(\frac{\theta}{2} \sum_{\omega \in \mathcal{O}} E[t_\omega] \left(\mu \left(\sum_{a \in \omega} k_a \right) + \frac{\sigma^2}{2} \left(\sum_{a \in \omega} k_a \right)^2 \right) \right),$$

Central limit theorem

Y is multivariate normal with

$$E[T_i] = \frac{\theta}{2} E[T_{MRCA}] \mu$$

$$Var[T_i] = \frac{\theta}{2} E[T_{MRCA}] \sigma^2$$

$$Cov[T_i, T_j] = \frac{\theta}{2} (T_{MRCA} - E[t_{1,2}]) \sigma^2$$

The distribution of individual differences

$$\text{Cov}[Y_1 - Y_2, Y_3 - Y_4] = \text{Cov}[Y_1, Y_3] + \text{Cov}[Y_2, Y_4] - \text{Cov}[Y_1, Y_4] - \text{Cov}[Y_2, Y_3]$$

$$\text{Cov}[Y_1 - Y_2, Y_3 - Y_4] \propto E[\tau_{1,4}] + E[\tau_{2,3}] - E[\tau_{1,3}] - E[\tau_{2,4}].$$

Is this just trivial?

maybe?

Other potential mutational distributions

Random walk mutation model (Crow and Kimura, 1964)

$$u(x, y) = u_{RW}(y - x)$$

House-of-cards mutation model (Kingman 1979)

$$u(x, y) = u_{HC}(y)$$

Zeng and Cockerham (1993)

$$u(x, y) = u_{ZC}(y - cx)$$

These destroy the compound Poisson process portion of things.