Notes on PPGtk

Paul Blischak

E-mail: blischak.4@osu.edu

Contents

1 Allele frequency estimation

1.1 The likelihood

The likelihood of an individuals' read data given the population allele frequency can be computed by summing over the possible genotypes:

$$\mathcal{L}_i(p) = P(r_i|p) = \sum_{j=0}^{m_i} P(r_i|j)P(j|p), \tag{S1}$$

$$\text{where} \quad P(r_i|j) = \binom{t_i}{r_i} [\mathcal{G}_\epsilon(j)]^{r_i} [1 - \mathcal{G}_\epsilon(j)]^{(t_i - r_i)}, \quad \text{and} \quad P(j|p) = \binom{m_i}{j} p^j (1-p)^{m_i - j}.$$

For multiple samples, we take the product of the individual likelihoods:

$$\mathcal{L}(p) = \prod_i \mathcal{L}_i(p) = \prod_i \left(\sum_{j=0}^{m_i} P(r_i|j)P(j|p) \right). \tag{S2}$$

Taking the natural log gives us the log likelihood of the population allele frequency at a single site:

$$\ell(p) = \log \mathcal{L}(p) = \sum_i \log \left(\sum_{j=0}^{m_i} P(r_i|j)P(j|p) \right). \tag{S3}$$

1.2 Metropolis-Hastings algorithm

$$P(p|r) \propto P(r|p)P(p) \tag{S4}$$

$$\alpha = \min \left\{ 1, \frac{P(r|p^*)P(p^*)}{P(r|p)P(p)} \right\} \tag{S5}$$

2 Inbreeding

2.1 The likelihood

2.2 Metropolis-Hastings algorithm