

Polygenic Multiplicative Fitness & MCMC

Moi Exposito-Alonso

2017-11-21

Genotypes and selection model

Given two loci, A and B, with alleles A and a, and B and b, there must be nine possible gametes. If the A allele and B allele are selected with coefficients s_A and s_B , and we assume no dominance and no epistasis, we can write the average fitness per genotype. For simplicity we use an haploid model, which would be similar to the diploid homozygotes case (also the case of highly self-fertilizing species):

	B	b
A	$(1 + s_A)(1 + s_B)$	$(1 + s_A)(1 - s_B)$
a	$(1 - s_A)(1 + s_B)$	$(1 - s_A)(1 - s_B)$

We could then expect that the observed fitness y of one of N number of haplotype of p loci is expressed as:

$$y_h = w(s, X) = \prod_{i=1}^p (1 + s_i X_{i,h})$$

The probabilistic model

The observed fitness of a haplotype will have a sampling variance that follows a Gamma distribution (typically with α and β parameters or $\alpha = k$ and $\theta = 1/\beta$):

$$Y \sim \Gamma(\alpha, \beta)$$

The first moment is:

$$E[Y] = \alpha/\beta$$

The second moment is then:

$$Var(y) = E[Y^2] - E[Y]^2 = \alpha/\beta^2$$

We can take the second moment as a constant that we can calculate from the observed variance across replicates, v . The first moment, μ will be the function $w(s, X)$ dependent on the genotypes and selection coefficients. Then we can rewrite the shape and rate parameters in terms of the moments $\alpha = \mu^2/v$ and $\beta = \mu/v$.

The probability density function of Gamma in terms of $k = \alpha$ and $\theta = 1/\beta$ is:

$$f(y) = \frac{1}{\Gamma(k)\theta^k} y^{k-1} e^{-\frac{y}{\theta}}$$

In log form this can be written as:

$$\ln f(y) = (k-1) \ln(y) - \frac{y}{\theta} - \ln \Gamma(k) - k \ln \theta$$

And in terms of moments:

$$\ln f(y) = \left(\frac{\mu^2}{v} - 1\right) \ln(y) - y \frac{\mu}{v} - \ln \Gamma\left(\frac{\mu^2}{v}\right) - \frac{\mu^2}{v} \ln \frac{v}{\mu}$$

And in terms of selection coefficients and genotypes

$$\begin{aligned} \ln f(y) = & \left(\frac{\prod_{i=1}^p (1 + s_i X_{i,h})^2}{v} - 1 \right) \ln(y) \\ & - y \frac{\prod_{i=1}^p (1 + s_i X_{i,h})}{v} \\ & - \ln \Gamma\left(\frac{\prod_{i=1}^p (1 + s_i X_{i,h})^2}{v} \right) \\ & - \frac{\prod_{i=1}^p (1 + s_i X_{i,h})^2}{v} \ln \frac{v}{\prod_{i=1}^p (1 + s_i X_{i,h})} \end{aligned} \quad (1)$$

For the full log likelihood we sum over all haplotypes:

$$\ln \mathcal{L}(\vec{s}) = \sum_{h \in \text{haps}} \ln f(y | X_h, \vec{s}, v_h)$$

Prior

We assume that the array of selection coefficients follow a multivariate Normal:

$$\vec{s} = (s_1, \dots, s_n) \sim \mathcal{N}(u, v)$$

So the prior probability is:

$$f(s) = \frac{1}{\sqrt{2\pi v^2}} e^{-\frac{(s-u)^2}{2v^2}}$$

And the log probability for all selection coefficients:

$$\ln f(\vec{s}) = -\frac{p}{2} \ln(2\pi) - \frac{p}{2} \ln v^2 - \frac{1}{2v^2} \sum_{i=1}^p (s_i - u)^2$$

$$\ln \mathcal{L}(\vec{s}) = \sum_{i=1}^p \ln f(s_i; u, v)$$

The conditional probability

$$\mathcal{P}(\vec{s}, u, v | y) = \mathcal{L}(y | \vec{s}) + \ln \mathcal{L}(\vec{s} | u, v)$$

MCMC

The updating/acceptance scheme is:

$$\min\left\{1, \frac{\mathcal{P}(y, \vec{s}^*, \alpha^*, \beta^*)q(\vec{s}^*, \vec{s})}{\mathcal{P}(y, \vec{s}, \alpha, \beta)q(\vec{s}, \vec{s}^*)}\right\}$$

By having a symmetric transition probability matrix (the Metropolis method), i.e. $q(\vec{s}^*, s) = q(\vec{s}, s)$, we can express the acceptance just as

$$\frac{\mathcal{P}(y, \vec{s}^*, \alpha^*, \beta^*)}{\mathcal{P}(y, \vec{s}, \alpha, \beta)}$$

After running a number of steps, an stationary distribution is obtained.