IM_hypotest
Computes hypothesis tests from IM surface files (*.surf).
Code written by C. R. Young
Last updated: 17 Nov 2005

This utility computes the hypothesis tests in Young *et al*. (2005).
The tests include:

1) Different effective population sizes (q_1 and q_2)
2) Different immigration proportions (m_1 and m_2)
3) Different effective numbers of immigrants (M_1 and M_2)

Hypothesis tests using IM
To compute the probabilities that effective sizes or immigration rates are different, we defined an indicator function $I(j)$, that takes a value of one if the hypothesis is true (e.g., $I(j) = 1$ if $m_1 > m_2$) , and zero if false (e.g., $I(j) = 0$ if $m_1 < m_2$) for step j in the MCMC dataset. An estimate of the probability of interest (e.g., $\Pr(m_1 > m_2) = p$) is the expectation of these indicator variables over s samples:

$$\Pr(m_1 > m_2) = \frac{1}{s}\sum_{j=1}^{s} I(j). \qquad (1)$$

Eq. 1 computes the posterior probability that $m_1 > m_2$ (see fig. 1).

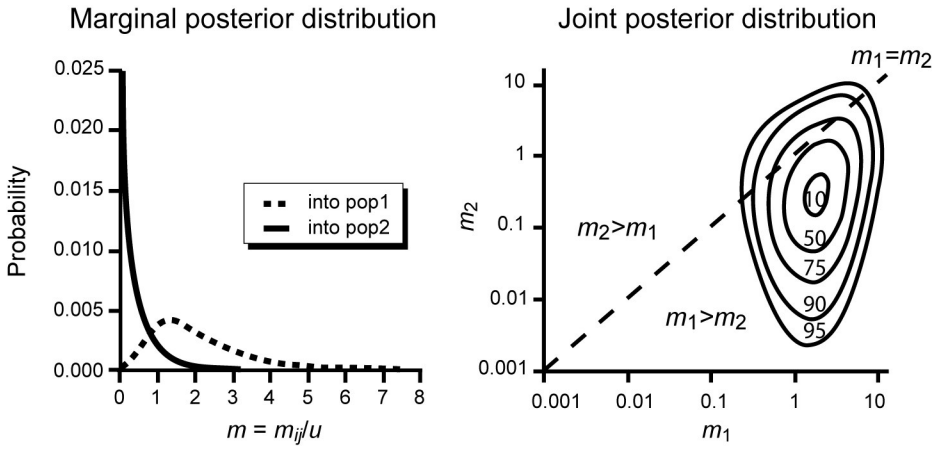

Figure 1. Marginal and joint posterior distributions for immigration proportions (m_1 and m_2). The contours of the joint posterior distribution contain the specified percentage of the distribution. The indicator function to test the hypothesis $m_1 > m_2$ computes the mass of the joint posterior distribution below the dashed line ($m_1 = m_2$). In this case, $\Pr(m_1 > m_2) = 0.93$, and the posterior odds ratio is about 14:1 in favor of the hypothesis that $m_1 > m_2$.

Posterior odds ratios are computed as $p/(1-p)$, where p is the posterior probability of the hypothesis with the highest statistical support (i.e., $p > 0.5$). In this manner, we express statistical support for the most probable hypothesis relative to the probability of the less favored hypothesis.

Rescaling IM parameter estimates
Immigration estimates are rescaled to find the posterior probability distributions of the effective number of immigrants in population *i*, $M_i = \theta_i m_i/2 = 2N_{e(i)}m_{ij}$ for each population, *i*. M_i is evaluated for each step in the MCMC dataset. We then compute the hypothesis test as defined above to determine the posterior probability that the effective numbers of immigrants are different between the two populations.

Running the utility
IM_hypotest expects *at least* two files to be present in the same directory as the executable. The first file, "filelist.txt" is a text file containing a list of the names of all of the surface files that you wish to thin. *This file needs to end with a new line character.* All of the surface files listed in "filelist.txt" need to be present in the directory also.

Input files (in same directory as IM_hypotest):

```
"filelist.txt"                Contains a list of the names of the files
                              that you want to read

"outfile.txt_0000.surf"       All of the surface files listed in
"outfile.txt_0001.surf"       "filelist.txt"
...
```

Useage:

```
IM_hypotest <burnin>
```

Options:

```
<burnin>          integer >= 0.     Ignores this many samples at the
                                    beginning of the run
```

Example of usage for 100,000 iteration burnin:

```
IM_hypotest 100000 > outfile.txt
```

The output generated by IM_hypotest consists of three posterior probabilities (for q, m, and M) and the associated posterior odds ratios.

Reference
Young CR, Fujio S, Vrijenhoek RC (2005) One-way dispersal between mid-ocean ridges: deep-ocean circulation and gene flow in *Ridgeia piscesae*. *Molecular Ecology* In review.