**Categorical Method**

1. Installation

*R CMD SHLIB McCOIL\_categorical\_code.c llfunction\_het.c*

2. Input format

Assuming there are *n* individuals and *k* loci, SNP calling information is stored in a matrix,



, where each element *Sij* represents SNP information at locus *j* of individual *i*, and can be 0 [homozygous minor allele], 0.5 [heterozygous], 1 [homozygous major allele] or -1 [missing data].

3. Run

See test\_R\_code.R for an example.

**Usage**

McCOIL\_categorical(data, ...)

**Arguments**

|  |  |
| --- | --- |
| data | An **R** data frame of SNP calling information. Row names are names of samples and column names are names of assays. |
| maxCOI | Upper bound for COI. The default is 25. |
| threshold\_ind | The minimum number of sites for a sample to be considered. The default is 20. |
| threshold\_site | The minimum number of samples for a locus to be considered. The default is 20. |
| totalrun | The total number of MCMC iterations. The default is 10000. |
| burnin | The total number of burnin iterations. The default is 1000. |
| M0 | Initial COI. The default is 15. |
| e1 | The probability of calling heterozygous loci homozygous. The default is 0.05. |
| e2 | The probability of calling homozygous loci heterozygous. The default is 0.05. |
| path | The default is the current directory. |
| output | The name of output file. The default is “output.txt”. |
| err\_method | The default is 1.  1: use pre-specified e1 and e2 and treat them as constants.  3: e1 and e2 are estimated with COI and allele frequencies |