Documentation for programs developed for analysing full-sibship and nested-sibship configurations with parents genotypes unknown (for unlinked loci.)

Ian Painter *

September 12 1995

This documents the following programs for evaluating likelihoods for full-sibship configurations and nested-sibship configurations.

conditions and corresponding programs			
configurations	*population allele frequencies*	*mutations*	*program name*
full-sibships	specified	no	eval_all_fulls_sibsets.c
full-sibships	specified	yes	eval_all_fulls_sibsets_with_mutations.c
full-sibships	unknown	no	integrate_all_fulls_sibsets.c
nested-sibships	specified	no	eval_all_nested_sibsets.c

The programs should come with the following sample input files:

test_eval_all_fulls_sibsets.data

test_eval_all_fulls_sibsets_with_mutations.data

test_integrate_all_fulls_sibsets.data

test_eval_all_nested_sibsets.data

and the corresponding correct results for these input files:

test_eval_all_fulls_sibsets.result

test_eval_all_fulls_sibsets_with_mutations.result

test_integrate_all_fulls_sibsets.result

test_eval_all_nested_sibsets.result

1. *Summary of programs*

*Department of Statistics,
University of Washington.
e-mail *painter@stat.washington.edu*

- eval_all_fulls_sibsets.c

  This program evaluates the likelihoods for all full-sibship configurations at specified population allele frequencies and no mutations.

- eval_all_fulls_sibsets_with_mutations.c

  This program evaluates the likelihoods for all full-sibship configurations, at specified population allele frequencies and allowing for a (specified) constant mutation rate.

- integrate_all_fulls_sibsets.c

  This program evaluates the integrated likelihoods for all full-sibship configurations, with no mutations possible. The likelihoods are integrated over the unknown population allele frequencies weighted by specified *Dirichlet* distributions.

- eval_all_nested_sibsets.c

  This program evaluated the likelihoods for all nested-sibship configurations, at specified population allele frequencies and no mutations.

2. Details

eval_all_fulls_sibsets.c

This program evaluates the likelihood for all full-sibship configurations at specified population allele frequencies and no mutations. Full-sibship configurations are those configurations consisting of full-sibships only.

Input: The input consists of the following:

- the number of loci.
- the number of alleles and the population allele frequencies for each loci.
- the number of offspring.
- an identifier for each offspring and that offsprings genotype.

The file test_eval_all_fulls_sibsets.data is a sample input file.

Output: Each line of the output consists of a log likelihood, a number for the configuration and the configuration. e.g.

```
                 .
                 .
-47.3840523   134   ( 1 2 4 )( 3 )( 5 6 )
-38.6893203   135   ( 1 2 4 )( 3 6 )( 5 )
-43.7275123   136   ( 1 2 4 6 )( 3 )( 5 )
-48.4085623   137   ( 1 2 4 )( 3 5 )( 6 )
0.0           138   ( 1 2 4 6 )( 3 5 )
                 .
                 .
```

The configuration (1 2 4)(3 5)(6) represents the relationships that offspring 1, 2 and 4 have both parents common, offspring 3 and 5 have both parents common (but different than the parents of 1, 2 and 4), and offspring 6 has a different set of parents. Configurations which are not possible have log-likelihoods recorded as 0. The file test_eval_all_fulls_sibsets.result is the correct output for the sample input file.

Notes: The program goes through the configurations in a lexographical order. The likelihoods for each full-sibship are calculated by summing over all possible genotypes for parents of the full sibship, and weighting the probability of the parents genotypes (using the specified population allele frequencies and assuming the Hardy-Weinberg principle.) The log-likelihoods for the full sibships in a configuration are added together to get the log-likelihood for that configuration. Works for up to 12 offspring. The maximum number of loci is 30 and the maximum number of alleles per loci is 12 (these can be changed by changing `MAXLOCI` and `MAXALLELES`.

eval_all_fulls_sibsets_with_mutations.c

This program evaluates the likelihood for all full-sibship configurations at specified population allele frequencies and with a constant mutation rate.

Input: The input should consist of the mutation rate, followed by input in the same format as for eval_all_fulls_sibsets.c. The file test_eval_all_fulls_sibsets_with_mutations.data is a sample input file.

Output: The output is in the same form as that for eval_all_fulls_sibsets.c. The file test_eval_all_fulls_sibsets_with_mutations.result is the correct output for the sample input file.

Notes: The program works the same was as in eval_all_fulls_sibsets.c, except that segregation probabilities are calculated under the assumption of the specified constant mutation rate. The mutation rate assumes that for each gene on each loci there is a constant probability of mutating into any of the alleles for that loci. The chance of mutating from any allele to any other allele is the same for each allele and each loci. Works for up to 12 offspring. The maximum number of loci is 30 and the maximum number of alleles per loci is 12 (these can be changed by changing `MAXLOCI` and `MAXALLELES`.

integrate_all_fulls_sibsets.c

This program evaluates the likelihood for all full-sibship configurations at unknown population allele frequencies and no mutations, by integrating across the population allele frequencies using specified *Dirichlet* distributions as weights.

Input: The input consists of the following:

- the number of loci.
- the number of alleles and the parameters of the Dirichlet distribution for each loci.
- the number of offspring.
- an identifier for each offspring and that offsprings genotype.

The file test_integrate_all_fulls_sibsets.data is a sample input file.

Output: The output is in the same form as that for eval_all_fulls_sibsets.c. The file test_integrate_all_fulls_sibsets.result is the correct output for the sample input file.

Notes: The likelihoods for each full-sibship at known population allele frequencies are polynomials in the population allele frequencies. For each loci, the products of the polynomials for each full-sibship in a configuration is calculated using symbolic manipulation and the resulting polynomial is then integrated, weighting by the specified Dirichlet distribution. Works for up to 12 offspring. The maximum number of loci is 30 and the maximum number of alleles per loci is 12 (these can be changed by changing `MAXLOCI` and `MAXALLELES`.

eval_all_nested_sibsets.c

This program evaluates the likelihood for all nested-sibship configurations at specified population allele frequencies and no mutations. Nested-sibship configurations are those configurations consisting of half sibships within full sibships and full sibships only.

Input: The form of the input is the same as for eval_all_fulls_sibsets.c. The file test_eval_all_nested_sibsets.data is a sample input file.

Output: Each line of the output consists of a log likelihood, a number for the configuration and the configuration. e.g.

```
                  .
                  .
-205.5332953     124    [ ( 1 4 ) ] [ ( 2 ) ] [ ( 3 ) ]
-200.9403693     125    [ ( 1 4 ) ] [ ( 2 )( 3 ) ]
-185.5818333     126    [ ( 1 4 )( 3 ) ] [ ( 2 ) ]
-185.4015503     127    [ ( 1 4 )( 2 ) ] [ ( 3 ) ]
-164.8208163     128    [ ( 1 4 )( 2 )( 3 ) ]
-207.5091093     129    [ ( 1 ) ] [ ( 2 3 ) ] [ ( 4 ) ]
                  .
                  .
```

The configuration [(1 4)(3)] [(2)] represents the relationship that offspring 1, 3 and 4 had one common parent, the other parent for offspring 1 and 4 being the same but different than the other parent for offspring 3, while offspring 2 had both parents different than any of the parents of offspring 1, 3 or 4.

Configurations which are not possible have log-likelihoods recorded as 0. The file `test_eval_all_nested_sibsets.result` is the correct output for the sample input file.

Notes: The program goes through the configurations in a lexographical order. At one level, offspring are assigned to sets with both parents common, and then all possible groupings for these sets into full-sibships within half sibships are examined.

The likelihoods for each half-sibship are calculated by summing over all possible genotypes for the parent of the half sibship weighting by the probability of the parent's genotype, and for each possible genotype for the parent, the possible genotypes for the other parent in each full sibship are summed over, weighting by the probability of the parent's genotype, and for each possible genotype for the parent. The weights are calculated from the specified population allele frequencies using the Hardy-Weinberg assumption. The log-likelihoods for the half sibships in a configuration are added together to get the log-likelihood for that configuration.

Works for up to 10 offspring. The maximum number of loci is 30 and the maximum number of alleles per loci is 12 (these can be changed by changing `MAXLOCI` and `MAXALLELES`.

3. Accuracy of programs

No guarantee can be given that these programs will work as stated, or that the output from the programs is correct or accurate. The output from all of the programs has been compared for very small examples against the results calculated by hand. Three variants of `eval_all_fulls_sibsets.c` have been written, and all give the same result as `eval_all_fulls_sibsets.c`. `eval_all_fulls_sibsets_with_mutations.c` with a zero mutation rate gives the same result as for `eval_all_fulls_sibsets.c`, as expected. `integrate_all_fulls_sibsets.c` with Dirichlet distributions which are essentially point distributions gives the same results as `eval_all_fulls_sibsets.c` evaluated at those points. One variant of `eval_all_nested_sibsets.c` has been written and gives the same results as `eval_all_nested_sibsets.c`. The results for those nested sibship configurations which are also full sibship configurations are identical under `eval_all_nested_sibsets.c` and `eval_all_fulls_sibsets.c`, as expected. No results have been compared with results obtained by programs written by people other than the author.