

This is a preprint of an article whose final and definitive form has been published in
Genetical Research, volume 81: issue 3, 2003.

© 2003 Cambridge University Press

Genetical Research online: <http://journals.cambridge.org/action/displayJournal?jid=GRH>

This article is also available online at: <http://dx.doi.org/10.1017/S0016672303006232>

An algorithm for sampling descent graphs in large complex pedigrees efficiently.

John M. Henshall*¹ and Bruce Tier**

* CSIRO Livestock Industries,

J. M. Rendel Laboratory,

Rockhampton QLD,

Australia.

** Animal Genetics and Breeding Unit²,

University of New England,

Australia.

Running head: Efficiently sampling descent graphs

Proofs to be set to:

John Henshall

CSIRO Livestock Industries, Private Mail Bag 1, Armidale,

New South Wales, 2350,

Australia.

Phone: +61 +2 6776 1302

Fax: +61 +2 6776 1333

E-mail: John.Henshall@csiro.au

¹Corresponding author. Current address: CSIRO Livestock Industries, Private Mail Bag 1, Armidale, NSW, 2350, Australia. Tel: +61 +2 6776 1302. Fax: +61 +2 6776 1333. e-mail: John.Henshall@csiro.au

²AGBU is a joint institute of NSW Agriculture and The University of New England.

SUMMARY

No exact method for determining genotypic and identity-by-descent probabilities is available for large, complex pedigrees. Approximate methods for such pedigrees cannot be guaranteed to be unbiased. A new method is proposed that uses the Metropolis-Hastings algorithm to sample a Markov Chain of descent graphs which fit the pedigree and known genotypes. Unknown genotypes are determined from each descent graph. Genotypic probabilities are estimated as their means. The algorithm is shown to be unbiased for small, complex pedigrees and feasible and consistent for large complex pedigrees.

Keywords: descent graph, Markov chain Monte Carlo, complex pedigree

1 INTRODUCTION

Many methods are currently used for estimating genotypic and identity by descent (IBD) probabilities in human and animal pedigrees. Genotypic and IBD probabilities are of interest to geneticists studying the transmission of genes through complex pedigrees, where a gene might be for a genetic disorder, a molecular test or marker. Commonly, the genotypes of some individuals in the population are known with certainty, partially known for other individuals (in that some genotypes can be excluded), and unknown for the remainder of the population. Most methods for estimating genotypic and IBD probabilities are suitable for small pedigrees, but have disadvantages when applied to large and complex pedigrees, where “complex” implies the presence of many marriage and inbreeding loops.

Gene dropping provides unbiased estimates, but it is only feasible for very small pedigrees. Methods based on peeling (e.g. [1]) provide unbiased estimates of genotypic probability for small pedigrees or pedigrees without loops. However, exact peeling is computationally infeasible for large complex pedigrees. Pedigree simplification and iterative peeling are two feasible methods used for large complex pedigrees, but estimates can no longer be guaranteed to be unbiased [2].

To solve the problem, a number of Markov chain Monte Carlo (MCMC) methods have been used to estimate genotypic and IBD probabilities (e.g. [3]). These methods sample either genotypes or descent graphs. They can produce unbiased samples for large complex pedigrees, provided that the sampling algorithm can traverse the parameter space efficiently; however, impediments to traversing the parameter space can be severe. Furthermore, when there are more than two alleles at the locus, MCMC methods for sampling genotypes are not necessarily irreducible [4]. MCMC methods which operate by sampling descent graphs need not be subject to irreducibility problems, but as noted by [5], “mixing” may be very poor. Adjacent samples are highly correlated and it may be infeasible to obtain sufficient samples to guarantee estimates have a low probability of error.

Attempts to improve the performance of descent graph sampling algorithms have focused on the correlation

of adjacent samples, and on drawing legal descent graphs through genotype elimination. At one extreme are MCMC algorithms such as those of ? where the autocorrelation of legal candidate samples is high. To reduce this autocorrelation composite transmission rules have been proposed (e.g. ???), in which structured groups of elements of the descent graph are changed together. However, changing more of the descent graph reduces the probability that the result will be legal, so more samples may be required. At the other extreme are methods for sampling uncorrelated descent graphs, such as those of ??. By applying the genotype elimination through inheritance constraint (GEIC) algorithm, all samples are legal, but the density of the samples drawn is due not only to the likelihood of the sample, but also to properties of the algorithm used to obtain the sample (the GEIC sampling density). While it is easy to use importance sampling or a Metropolis Hastings step to adjust for the GEIC sampling density, for pedigrees of reasonable size this adjustment results in a small number of effective samples, resulting in estimates of low accuracy.

In this paper a new MCMC method for sampling descent graphs is proposed in which the independent descent graph method of ? is placed into a MCMC context. A Metropolis-Hastings (MH) step is used to accept or reject candidate graphs which have far less autocorrelation than the candidate graphs used in other MCMC methods. The paper shows that the algorithm can produce unbiased estimates on small complex pedigrees and that it is feasible for large complex pedigrees.

2 METHOD

?? describe a method (GEIC) for estimating genotypic and IBD probabilities from independently sampled descent graphs. Each descent graph is sampled de novo and consequently adjacent samples are completely uncorrelated.

This new algorithm puts the GEIC algorithm into an MCMC framework. An initial descent graph is sampled using GEIC. Subsequent samples are obtained by using the GEIC and MH algorithms. A subset of the primary descent graph is retained from the previous sample, and the remainder of the the primary descent graph is sampled using GEIC. MH is used to accept or reject the candidate. This sampling procedure in the algorithm differs to that described in ? in two important ways.

Firstly, a partial primary descent graph is sampled from the inheritance constraints of the current sample primary descent graph. This is a legal subset of the current descent graph. GEIC is then used to complete a new primary descent graph. Base alleles and a secondary descent graph are then sampled as in ?.

Secondly, the MH algorithm is used to accept or reject candidate samples, based on the likelihood of the sample, and the probability of moving from the current descent graph to the candidate descent graph, rather than using importance sampling to weight samples. This probability is similar to the importance sampling density in the method of ?, and is a function of the number of elimination and base gamete sampling steps required to produce each sample. MH is of benefit here as adjacent samples are correlated.

The new algorithm is an MCMC descent graph sampling algorithm, but has the ability to make long jumps between adjacent samples. Accordingly it is referred to here as the long jumping descent graph sampler (LJDGS). A full description of the algorithm follows:

1. Obtain a legal descent graph and associated likelihood ($\pi(y)$) and GEIC sampling density ($g(y)$) using the method of ?.

2. Repeat

- (a) Set $\pi(x) = \pi(y)$ and $g(x) = g(y)$.
- (b) Sample a subset of the primary descent graph to retain.
- (c) Apply GEIC to the pedigree, constrained to the retained subset of the primary descent graph, to obtain a new descent graph, with likelihood $\pi(y)$ and GEIC sampling density $g(y)$.
- (d) Apply MH, with $q(x,y) = g(y)$, $q(y,x) = g(x)$, and acceptance criterion

$$\min((\pi(y)q(y,x))/(\pi(x)q(x,y)), 1).$$

- (e) If the candidate sample is rejected, then set $\pi(y) = \pi(x)$ and $g(y) = g(x)$.
- (f) Accumulate the most recently accepted descent graph and associated genotypes.

3. Summarise parameters of interest as means of the samples.

(i) Subset sampling algorithm

It is critical to the success of the algorithm that the method of sampling subsets of the primary descent graph to retain (step ??) is balanced, and does not affect the candidate generating density $g(y)$. It may not be possible to prove that a particular subset sampling algorithm satisfies this criterion, but exhaustive testing has failed to find fault with the strategy described below.

A primary descent graph consists of the set of paths connecting informative gametes to base gametes. Before choosing the subset to be retained, a binary variable is initialised to “save” for all gametes in the primary descent graph. All paths connecting informative gametes to base gametes are traversed. With each step (gamete) on the traverse, a random number between zero and one is drawn. If the random number exceeds a predetermined non-zero probability b , then the binary variable associated with the gamete switches from the “save” state to a “discard” state. All binary variables associate with gametes on the path between, and including, it and the base gamete are set to “discard” (the initial change to “discard” breaks the link between

the informative and base gametes). It is possible for paths connecting a number of informative gametes to intersect at the same gamete. If when traversing a path a gamete with a variable set to “discard” is found then the remainder of that path has also been set to “discard” and is not reset again. For any individual on a path where one of its gametes’ variables has been set to “discard”, the state of the variable associated with its other gamete (and all gametes between it and the connected base gamete) is also set to “discard”. Gametes that remain set to “save” retain their current inheritance state when a new primary descent graph is sampled.

The variable b need not remain constant, and the algorithm may be “tuned” to different pedigrees by varying the method of selecting b .

3 TEST ANALYSES

Two test pedigrees were used to test the LJDGS algorithm.

A pedigree with 11 individuals (pedigree A, table 1) was used to validate that the method was able to produce unbiased estimates. This pedigree is small, allowing the calculation of exact genotypic probabilities for comparison purposes. A single locus, with 4 alleles with founder allele frequencies (0.5, 0.25, 0.2, 0.05) was assumed, and test data sets obtained by sampling base gametes and Mendelian transmission. On each test data set genotypes were made available for four randomly chosen individuals, and assumed unknown on the remaining seven individuals.

From pedigree A, 1000 random datasets were analysed twice, once with adjacent samples independent - essentially the algorithm used by ? but with a MH step instead of importance sampling (IDGS), and once with the LJDGS, with correlated descent graph samples. With LJDGS, the parameter b did not remain constant, but for each sample was drawn from a $\beta(1, 1)$ distribution. Genotypic and IBD probabilities were estimated as the mean of 10,000 samples. As the MH algorithm is used, the effective number of samples

is less than 10,000. A simple estimate of the effective number of samples was used. The sample which was retained for the most MH cycles (r_{max}) was assumed to have contributed one to the effective number of samples, the effective number of samples then being $\frac{n_s}{r_{max}}$, where n_s is the total number of samples. This measure is less appropriate for LJDGS than IDGS, because in LJDGS adjacent samples are correlated. Exact genotypic probabilities were obtained using MENDEL (?), which uses a peeling based algorithm.

To compare the sampled genotypic probabilities to the exact probabilities, the test statistic used in ? was used, $\chi^2 = \sum_{E_{kl} \neq 0} \frac{(O_{kl} - E_{kl})^2}{E_{kl}}$, where k relates to the individual, l is the unordered (no distinction between paternal and maternal) genotype, E_{kl} is the number of samples expected to occur for genotype l in individual k (calculated from the probabilities obtained using MENDEL and the effective number of samples) and O_{kl} is the effective number of samples which were observed for genotype l in individual k . This statistic has an approximate χ^2 distribution, with $n - 11$ degrees of freedom, where n is the number of non-zero E_{kl} in the sum. The distribution of the test statistic is only approximate as both within and across individuals, genotype probabilities are not independent, and the effective number of samples is only an approximation.

A larger pedigree with 1600 individuals was used to evaluate the performance of the LJDGS algorithm on more challenging data. This pedigree, modeled on the simulated pedigree of ?, consisted of 20 discrete generations, each with 80 individuals. These were the ten progeny from each of 8 matings between males and females from the previous generation. A single locus with 16 alleles, with uniform frequencies in the base individuals, was assumed. A single data set was obtained by sampling base gametes and Mendelian transmission. Analyses were then performed with varying proportions of the simulated genotypes made available. For all analyses genotypes for all individuals born in the final generation were available. Three datasets were constructed, with genotypes made available on 0%, 25% or 50% of the remaining individuals, with individuals to be genotyped chosen at random (pedigrees B0, B25 and B50).

Each of pedigrees B0, B25 and B50 was analysed using two methods, once with no correlation between adjacent samples (IDGS) and once with correlated descent graph samples (LJDGS). Each analysis involved

drawing 10,000 samples, and was repeated five times, using different seeds for the random number generator each time. With LJDGS, the parameter b did not remain constant, but for each sample was drawn from a $\beta(80, 1)$ distribution. Genotypic and IBD probabilities were estimated as means, and the effective number of samples calculated as described above.

Mixing was assessed in three ways. First, by examining the degree of symmetry in the inheritance of base gametes, as the probability of a base gamete being sampled as “paternal” should equal the probability of it being sampled as “maternal”. The test statistic used was

$$\text{SYM} = 1 - \frac{\sum_{i=1}^{n_b} \sum_{j=1}^{n_a} \sum_{k=1}^j |p_{ijk} - p_{ikj}|}{n_b},$$

where n_b is the number of base individuals (= 80), n_a is the number of alleles (= 16) and p_{ijk} is the probability that individual i inherited allele j from its sire and allele k from its dam. This statistic can take values from zero to one, with lower values less symmetric in the inheritance of base alleles.

For the second measure of mixing unordered genotypic probability estimates were considered across the five replicates. The number of genotypic probability estimates in which at least one replicate had a zero probability while at least one replicate had a probability greater than 0.00, 0.01, 0.02, 0.05 or 0.10 was calculated. The statistics $Z_0, Z_{01}, Z_{02}, Z_{05}$ and Z_{10} are these counts, expressed as percentages of the total number of cells with non-zero probabilities, excluding cells in which there was no variation.

For the third measure of mixing unordered genotypic probability estimates were again considered across replicates. The statistics S_{01}, S_{02}, S_{05} and S_{10} are number of cells in which the standard deviation (over replicates) of the genotypic probability estimate exceeded 0.01, 0.02, 0.05 or 0.10 respectively, expressed as percentages of the total number of cells with non-zero standard deviations.

While these tests may identify inadequate mixing, that inadequate mixing has not been identified is not a guarantee of adequate mixing.

4 RESULTS

The genotypes for pedigree A are well estimated (Figure 1). The test statistics obtained for the 1000 analyses of pedigree A are plotted against the approximate degrees of freedom, for both IDGS and LJDGS. The test statistics show chance deviations from expectation. Although the distributions of test statistics are similar for the two methods, there appear to be fewer extreme test statistics for LJDGS.

However, with large complex pedigrees LJDGS performs much better than IDGS (Table 2). Of the 10,000 samples, the number accepted is far higher with LJDGS than IDGS, and this is reflected in the much higher effective number of samples for LJDGS. The effective number of samples for IDGS is so low that genotypic probability estimates would be expected to be of very low accuracy while with LJDGS one would expect reasonable estimates. However, for LJDGS samples are not independent, so it is possible that the effective number of samples is an overestimate.

The measures of mixing, while of dubious value, indicate that LJDGS shows better mixing than IDGS. The first measure of mixing, symmetry in the base individuals, provides no evidence of mixing problems with LJDGS. However, reasonably high levels of symmetry also occur for IDGS for pedigree B0, despite only 2.5 effective samples, suggesting that base symmetry is not in itself an indicator of good mixing. The second measure of mixing, the percentage of genotypes in which at least one zero probability was observed while the maximum probability observed was greater than 0.00, 0.01, 0.05 or 0.10, suggests that mixing has been good with LJDGS for all pedigrees. Again however, no problem with mixing has been observed for IDGS for one pedigree, in this case B50. This casts some doubt on the worth of this statistic as an indicator of mixing. The third measure of mixing, the standard deviation of genotype probabilities across replicates, suggests that LJDGS mixes better than IDGS. Standard deviations of less than 5 (%) suggest that genotypic probability estimates are accurate to around 5%, which may be acceptable for some applications. A relatively small percentage of genotypic probabilities had standard deviations above 5 (%), especially for pedigrees B25 and B50.

The behaviour of the MH step shows that samples are more likely to be accepted if a larger proportion of the primary descent graph is retained. In table 3 the mean and minimum percentage of the primary descent graph saved is provided for both accepted and rejected samples. These are for the last 9,000 of the 10,000 samples, to increase the probability that the algorithm was sampling from the equilibrium distribution. It is clear that samples are more likely to be accepted if a large proportion of the current primary descent graph is retained. However, it is also evident that on occasions samples are accepted which retain little of the current primary descent graph.

Runs of 10,000 samples took on average 16.6, 3.4 and 4.5 hours for LJDGS on pedigrees B0, B25 and B50 respectively, using a Pentium III Xeon 1.7GHz processor.

5 DISCUSSION

By using the GEIC (genotype elimination by inheritance constraint) algorithm the method described in this paper generalises the method of ? for sampling descent graphs. The GEIC algorithm ensures that all candidate descent graphs are legal, and there is no need to “tunnel through” illegal descent graphs. As with all descent graph sampling algorithms, LJDGS is suitable for loci with more than two alleles to be evaluated without concern about the Markov chain being reducible.

The results from pedigree B suggest that this method is feasible on moderately sized pedigrees, with large numbers of alleles per locus. The large variation in time taken for pedigrees B0, B25 and B50 suggests that size and number of alleles are not the only factors affecting feasibility, the proportion and distribution of genotyped individuals in the pedigree is also very important, with pedigrees rich in genotyped individuals more quickly analysed than pedigrees with sparse genotype information.

The evaluation of the small pedigrees described in table 1 indicates that the results from this algorithm are unbiased. Nevertheless, we found that the choice of too few nodes for resampling could limit mixing. The

variation in the MH acceptance rate between pedigrees B0, B25 and B50 suggests that the algorithm may be tuned by varying the method of sampling b , the proportion of the primary descent graph to retain. Here, b was sampled from a $\beta(a, 1)$ distribution, where $a = \max(\frac{n}{20}, 1)$ and n was the number of individuals. It may be more appropriate to use a function of the number of genotyped individuals as the first parameter in the beta distribution. This tuning may be very important. If b is consistently small, then fewer candidate graphs will be accepted, and the effective number of samples will be reduced. If b is consistently large, then adjacent samples will be more correlated, again reducing the number of effective samples. This could be determined while the algorithm is running.

The method for sampling descent graphs, by randomly drawing a number of nodes to be resampled, permits rapid exploration of the parameter space, and enhances mixing. While it has not been shown that mixing is a problem in descent graph sampling methods such as ? , it is difficult to be sure that it is not. These methods are roughly equivalent to LJDGS but with a value of b very close to 1.0. The methods used here to test mixing are not perfect, for example, the methods which use the similarity between replicates would give a favourable statistic if all replicates were similar, even if this similarity was solely due to the use of similar starting values. This could be a problem with any MCMC method that requires a valid descent graph as a starting value, as the descent graph sampling density of ? shows that some descent graphs are thousands of times more likely to be found than others, and this variation is not due to the likelihood of the descent graph. Therefore, there may be a significant chance that sampled descent graphs, drawn for use as “fresh” starting values, all share some characteristic. It would appear that this is less likely to be a problem with LJDGS, as the descent graph sampling density is explicitly included in the MH step.

The execution times presented are for development software, and it is likely that significant speedups could be made through enhancements to the GEIC algorithm such as those proposed by ? . As they are cheap to obtain, it is possible to sample many secondary descent graphs for each primary descent graph, and use a weighted average for determining genotypic and IBD probabilities. Even with significant speedups

it will not be possible to draw as many samples as is possible with other MCMC descent graph sampling algorithms, such as those of ?. However, as adjacent samples obtained with LJDGS should be less correlated than those from other MCMC algorithms, fewer samples need be drawn to get good genotypic probability estimates. A composite method, using a conventional MCMC descent graph sampler to sample in the region of each LJDGS sample is also possible.

While the results presented here are for single loci, the extension to multiple loci is straight forward, using the likelihoods in ?. Descent graph sampling methods for quantitative trait loci (QTL), such as that of ? can also be combined with LJDGS, to sample QTL linked to markers.

6 CONCLUSIONS

By combining the best elements of existing MCMC descent graph sampling algorithms with the best elements of independent descent graph sampling methods, the method proposed here has the potential to be of use in estimating genotypic probabilities and IBD probabilities in large complex pedigrees. Adjacent samples are far less correlated than those produced using other MCMC descent graph sampling algorithms, reducing the number of samples required to produce reliable estimates. At the same time, allowing some correlation ensures that enough samples are accepted to make the analysis of large pedigrees feasible.

References

- Du, F.-X. and Hoeschele, I. (2000). A note on algorithms for genotype and allele elimination in complex pedigrees with incomplete genotype data. *Genetics* **156**, 2051–2062.
- Elston, R. C. and Stewart, J. (1971). A general model for the genetic analysis of pedigree data. *Human Heridity* **21**, 523–542.
- Fernando, R. L., Stricker, C., and Elston, R. C. (1993). An efficient algorithm to compute the posterior genotypic distribution for every member of a pedigree without loops. *Theoretical and Applied Genetics* **87**, 89–93.
- Guo, S. W. and Thompson, E. A. (1994). Monte Carlo estimation of mixed models for large complex pedigrees. *Biometrics* **50**, 417–432.
- Heath, S. C. (1998). Generating consistent genotypic configurations for multi-allelic loci and large complex pedigrees. *Human Heridity* **48**, 1–11.
- Henshall, J. M., Tier, B., and Kerr, R. J. (1999). Inferring genotype probabilities for untyped individuals in complex pedigrees. In *Proceedings of the Thirteenth Conference of the Association for the Advancement of Animal Breeding and Genetics* pages 329–332.
- Henshall, J. M., Tier, B., and Kerr, R. J. (2001). Estimating genotypes with independently sampled descent graphs. *Genetical Research* **78**, 281–288.
- Janss, L. L. G., Thompson, R., and van Arendonk, J. A. M. (1995a). Application of Gibbs sampling for inference in a mixed major gene-polygenic inheritance model in animal populations. *Theoretical and Applied Genetics* **91**, 1137–1147.
- Janss, L. L. G., van Arendonk, J. A. M., and van der Werf, J. H. J. (1995b). Computing approximate monogenic model likelihoods in large pedigrees with loops. *Genetics Selection Evolution* **27**, 567–579.

- Kerr, R. J. and Kinghorn, B. P. (1996). An efficient algorithm for segregation analysis in large populations. *Journal of Animal Breeding and Genetics* **113**, 457–469.
- Lange, K. and Sobel, E. (1991). A random walk method for computing genetic location scores. *American Journal of Human Genetics* **49**, 1320–1334.
- Lange, K., Weeks, D. E., and Boehnke, M. (1988). Programs for pedigree analysis: MENDEL, FISHER and dGENE. *Genetic Epidemiology* **5**, 471–472.
- Sheehan, N. and Thomas, A. (1993). On the irreducibility of a markov chain defined on a space of genotype configurations by a sampling scheme. *Biometrics* **49**, 163–175.
- Sobel, E. and Lange, K. (1993). Metropolis sampling in pedigree analysis. *Statistical Methods in Medical Research* **2**, 263–282.
- Sobel, E. and Lange, K. (1996). Descent graphs in pedigree analysis: Applications to haplotyping, location scores, and marker-sharing statistics. *American Journal of Human Genetics* **58**, 1323–1337.
- Stricker, C., Fernando, R. L., and Elston, R. C. (1995). An algorithm to approximate the likelihood for pedigree data with loops by cutting. *Theoretical and Applied Genetics* **91**, 1054–1063.
- Tier, B. and Henshall, J. M. (2001). A sampling algorithm for segregation analysis. *Genetics Selection Evolution* **33**, 587–603.
- van Arendonk, J. A. M., Smith, C., and Kennedy, B. W. (1989). Method to estimate genotype probabilities at individual loci in farm livestock. *Theoretical and Applied Genetics* **78**, 735–740.

(Legend for table 1)

Table 1. Pedigree A. For each analysis, genotype was sampled for base individuals and Mendelian transmission was sampled for non-base individuals. Unordered genotype was then made available for only four randomly chosen individuals.

Table 1.

id	father	mother
1	0	0
2	0	0
3	0	0
4	2	1
5	2	3
6	2	4
7	5	4
8	5	3
9	7	6
10	7	8
11	10	9

(Legend for table 2)

Table 2. Results of five repeated analyses of a pedigree with 1600 individuals. A 16 allele locus was simulated, with genotypes available on the last generation and on 0% (B0), 25% (B25) or 50% (B50) of the remaining individuals. 10,000 samples were drawn using the independent descent graph sampler (IDGS) and the long jumping descent graph sampler. The number of samples accepted (N_{ACC}) and an approximation of the effective number of samples (N_{EFF}) is provided. For these, along with a measure of mixing, base allele symmetry (SYM), larger values are desirable. Two other measures of mixing are provided, for which smaller values are desirable. They are the percentage of genotypes in which at least one replicate had a zero probability while at least one replicate had a probability greater than 0.00 (Z_0), 0.01 (Z_{01}), 0.02 (Z_{02}), 0.05 (Z_{05}) or 0.10 (Z_{10}), and the percentage of genotypes in which the standard deviation of the probability estimate across replicates exceeded 0.01 (S_{01}), 0.02 (S_{02}), 0.05 (S_{05}), 0.10 (S_{10}) or 0.20 (S_{20}). These percentages are calculated using only genotypes in which there was some variation across replicates.

Table 2.

Pedigree	B0		B25		B50	
	IDGS	LJDGS	IDGS	LJDGS	IDGS	LJDGS
N_{ACC}	17	4132	16	3229	37	5996
N_{EFF}	2.5	538.7	1.9	397.8	3.4	848.5
SYM	0.958	0.986	0.793	0.979	0.834	0.990
Z_0	9.5	1.2	5.2	2.5	1.0	1.4
Z_{01}	7.9	0.3	3.4	0.6	0.4	0.1
Z_{02}	7.5	0.2	3.3	0.4	0.2	0.0
Z_{05}	7.1	0.2	3.0	0.3	0.2	0.0
Z_{10}	6.6	0.2	2.5	0.3	0.1	0.0
S_{01}	75.3	57.2	74.5	38.0	85.0	7.7
S_{02}	68.5	35.1	68.0	16.6	77.9	1.2
S_{05}	53.8	13.3	56.7	4.8	60.8	0.1
S_{10}	39.9	4.0	50.6	1.4	39.5	0.0
S_{20}	18.7	0.4	36.0	0.5	10.1	0.0

(Legend for table 3)

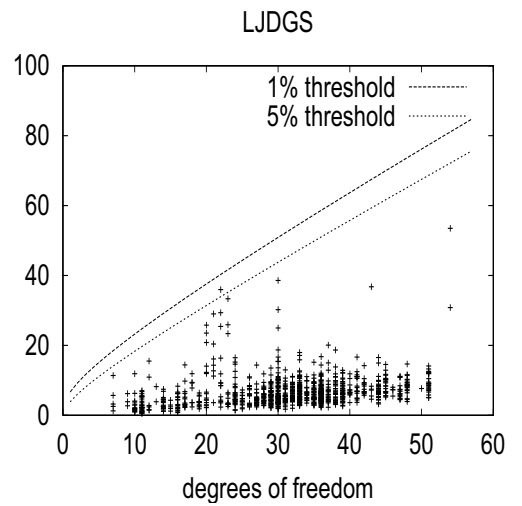
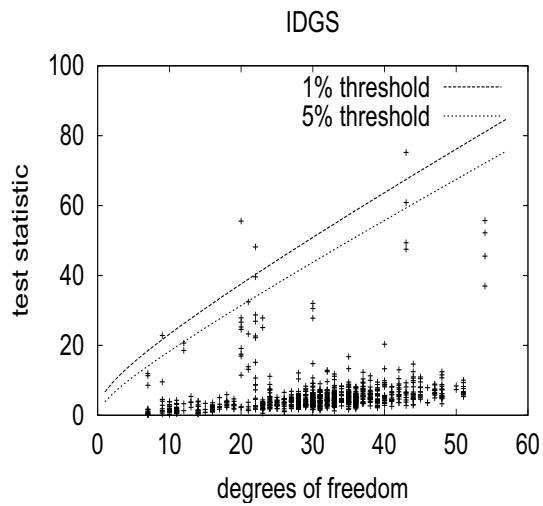
Table 3. Summary statistics for accepted and rejected samples for pedigrees B0, B25 and B50. The percentage of samples accepted (ACC) is provided, along with the mean and minimum percentage of the primary descent graph retained for both the accepted samples (μ_a and \min_a) and rejected samples (μ_r and \min_r).

Table 3.

	ACC	μ_a	μ_r	\min_a	\min_r
B0	0.41	0.93	0.61	0.23	0.09
B25	0.32	0.92	0.57	0.22	0.12
B50	0.60	0.78	0.43	0.14	0.09

(Legend for figure 1)

Figure 1 Distribution of test statistic obtained for 1,000 pedigrees analysed using the descent graph sampler with independent samples (IDGS) and the long jumping descent graph sampler with correlated samples (LJDGS).



This is a preprint of an article whose final and definitive form has been published in
Genetical Research, volume 81: issue 3, 2003.

© 2003 Cambridge University Press

Genetical Research online: <http://journals.cambridge.org/action/displayJournal?jid=GRH>

This article is also available online at: <http://dx.doi.org/10.1017/S0016672303006232>