Most Powerful Permutation Invariant Tests for Relatedness Hypotheses Using Genotypic Data

Anthony Almudevar*

August 1, 2000

Abstract

A class of tests for permutation invariant relatedness hypotheses using genotypic data is proposed, which are proven to be of maximum power among permutation invariant tests. These tests lead naturally to "locally most powerful tests", in the sense that power is maximized for alternatives statistically close to a null hypothesis of unrelatedness. Although the resulting statistic is a U-statistic, normal approximation theory is found to be inapplicable, due to high skewness. As an alternative it is found that a conditional procedure based on the most powerful test statistic can calculate accurate significance levels without much loss in power. Examples are given in which this type of test proves to be more powerful than a number of alternatives considered in the literature, including Queller and Goodknight's (1989) estimate of genetic relatedness, the average number of shared alleles (Blouin, 1996) and the number of feasible sibling triples (Almudevar and Field, 1999).

*Department of Mathematics and Computing Science, St. Mary's University, Halifax, N. S., Canada, B3H 3C3, e-mail: anthony.almudevar@stmarys.ca

Key words: most powerful invariant test, genotypic data, relatedness inference, conditional tests

1 Introduction

The ability to infer relatedness among individuals of a species of unknown pedigree is important to the study of wild populations, as well as to captive breeding programs. Such inference can proceed by identifying joint, or at least pairwise, kinship relationships between specific individuals at the multigenerational level as in Thompson (1976) and Meagher and Thompson (1986, 1987). Techniques specialized for the single generation cases are proposed in Blouin (1996), Painter (1994, 1996) and Almudevar and Field (1999). Alternatively, aggregate measures of group relatedness have been developed in Queller and Goodknight (1989) and Pamilo (1989) in which individual kinships are not inferred.

The approach proposed here can be considered as an intermediate methodology, in that tests can be constructed for particular types of relationships, but the inference is ultimately of an aggregate nature. Formally, we test a null hypothesis of unrelatedness against a particular group of relatedness hypotheses. Inference is permutation invariant with respect to sampled individuals. We use the generalized form of the Neyman-Pearson lemma to derive tests which are most powerful against a general class of such alternative hypotheses. The motivating example, developed here, is the case of sibship inference. If we wish simply to determine whether or not sibships exist in a sample, without the necessity of individual kinship inference, then a suitably constructed hypothesis test is appropriate. This leads to the development of a locally most powerful test; that is, one which is most powerful among all other invariant tests against alternative hypotheses "near" the null hypothesis. In the case of sibling inference we may define as "near" the hypothesis that there exist exactly two

individuals in a sibling relationship, with all others mutually unrelated.

In Section 2 the class of hypothesis tests is defined, and the most powerful tests derived. In Section 3 the theory is specialized to the case of sibling inference. Some guidelines regarding the calculation of significance levels are developed. In particular, normal approximations are seen to be of no applicability, despite the fact that the statistics are asymptotically normal. A conditional approach is developed, which gives accurate significance levels with little sacrifice of power. In addition, the test is compared to a number of alternative measures of relatedness appearing in the literature. A further example, using a set of DNA collected from a population of sperm whales, is analyzed using the proposed methodologies in Section 4. The results are summarized and discussed in Section 5.

2 Most powerful invariant tests

Let $\mathcal{X}$ be a discrete sample space. We introduce a generalization of the Neyman-Pearson Lemma (Lehmann, 1986, pg. 96), stated without proof. It is presented here (in part) using notation appropriate for discrete domains.

Theorem 1 *(Neyman-Pearson) Suppose $g_1, \ldots, g_m$ are real valued integrable functions on $\mathcal{X}$, and $b_1, \ldots, b_m$ are finite constants. Let $\mathcal{C}$ be the class of functions $\phi : \mathcal{X} \to [0,1]$ which satisfy*

$$\sum_{x \in \mathcal{X}} \phi(x) g_i(x) = b_i, \;\; i = 1, \ldots, m.$$

If g_0 is an integrable function on $\mathcal{X}$ then

(i) There exists $\phi \in \mathcal{C}$ which maximizes $\sum_{\mathcal{X}} \phi(x) g_0(x)$.

(ii) A sufficient condition for ϕ to maximize $\sum_{\mathcal{X}} \phi(x) g_0(x)$ is the existence of constants

$k_1, \ldots, k_m$ *such that*

$$\phi(x) = 1 \quad when \quad g_0(x) > \sum_{i=1}^{m} k_i g_i(x)$$

$$\phi(x) = 0 \quad when \quad g_0(x) < \sum_{i=1}^{m} k_i g_i(x)$$

In the problem under consideration $\mathcal{X}$ is the set of all possible multilocus genotype arrays among a sample of N individuals, which identifies a multilocus genotype for each individual in the sample. Let $\mathcal{G}$ be the class of all permutations of individual labelings acting on elements of $\mathcal{X}$. We define a *relationship class* R to be a *type* of relationship among K individuals. We may, for example, define R to be a full sibling relationship among K individuals. The more detailed specification $R(i_1, \ldots, i_K)$ is taken to define a type R relationship specifically among (ordered) individuals $i_1, \ldots, i_K$. Suppose we let $\mathcal{R}$ denote all possible joint relationships among individuals $1, \ldots, N$. For convenience we adopt the convention that for $r \in \mathcal{R}$ we set $r = R(i_1, \ldots, i_K)$ if individuals $i_1, \ldots, i_K$ are in relationship type R and all other individuals are unrelated to the rest of the sample. The permutation group generates a set of hypotheses

$$\mathcal{G}R = \{R(i_1, \ldots, i_K) : (i_1, \ldots, i_N) \in \mathcal{G}\} .$$

Suppose $r \in \mathcal{R}$ is the true joint relationship of the sample. The present goal to develop tests for the hypotheses

$$H_o : \quad r = U$$

$$H_a : \quad r \in \mathcal{G}R$$

for some relationship type R based on observation $X \in \mathcal{X}$. For example if R is a relationship type specifying full sibship of two individuals, the hypothesis test would become

$$H_o : \quad \text{All individuals unrelated}$$

H_a : All individuals unrelated except for two full siblings

A hypothesis test is given by *critical function* $\phi : \mathcal{X} \to [0,1]$ where H_o is rejected with probability $\phi(X)$. If ϕ equals zero or one, the test is nonrandomized. If there is no *a priori* reason to favor one element of $\mathcal{G}R$ over another then, logically, we would require the hypothesis test ϕ to be permutation invariant in the sense that $\phi(X) = \phi(GX)$ for all $X \in \mathcal{X}$ and $G \in \mathcal{G}$. Suppose we denote Gr to be the relationship obtained by subjecting the individuals in relationship r to permutation G. For any set A, by group symmetry we have

$$P(GX \in A|r) = P(X \in A|Gr) \tag{2.1}$$

The power function is defined as

$$\beta(r) = E_r[\phi(X)]$$

so that by (2.1) for a permutation invariant test we have

$$\begin{aligned} \beta(Gr) &= E_{Gr}[\phi(X)] \\ &= E_r[\phi(GX)] \\ &= E_r[\phi(X)] \\ &= \beta(r). \end{aligned} \tag{2.2}$$

If $\beta(r) = \beta(Gr)$ for all $G \in \mathcal{G}$ the test is called *unbiased*. The approach taken below will be to find the most powerful unbiased test. As will be shown, the resulting test is also permutation invariant. Since by (2.2) permutation invariance implies unbiasedness, this test is also the most powerful permutation invariant test. See Lehmann (1986), chapter 6, for a comprehensive discussion of invariance in testing.

We write $X = (X_1, \ldots, X_N)$, where X_i is the multilocus genotype of individual i. Let $p(x)$ to be the population frequency of genotype x. Consider a relationship type R. Let

$Q_R(x_{i_1}, \ldots, x_{i_K})$ be the joint probability of genotype arrays $x_{i_1}, \ldots, x_{i_K}$ when individuals $i_1, \ldots, i_k$ are in relationship type R. The distribution of X given configuration $r = U$ or $r = R(i_1, \ldots, i_K)$ can be written

$$\begin{aligned} f_U(x_1, \ldots, x_N) &= \prod_{n=1}^{N} p(x_n) \\ f_r(x_1, \ldots, x_N) &= f_U(x_1, \ldots, x_N) L_R(x_{i_1}, \ldots, x_{i_K}) \end{aligned}$$

where

$$L_R(y_1, \ldots, y_K) = \frac{Q_R(y_1, \ldots, y_K)}{\prod_{j=1}^{K} p(y_j)}.$$

To determine the form of the test first fix the size by specifying $\alpha = \beta(U)$. Let r_1 be one particular member of $\mathcal{G}R$. Then the problem of finding the most powerful unbiased test is equivalent to finding critical function ϕ which maximizes

$$\beta(r_1) = \sum_{x \in \mathcal{X}} \phi(x) f_{r_1}(x)$$

subject to

$$\beta(r_1) - \beta(r) = \sum_{x \in \mathcal{X}} \phi(x) f_{r_1}(x) - \sum_{x \in \mathcal{X}} \phi(x) f_r(x) = 0, \text{ for all } r \in \mathcal{G}R \tag{2.3}$$

and

$$\beta(U) = \sum_{x \in \mathcal{X}} \phi(x) f_U(x) = \alpha. \tag{2.4}$$

Order the elements of $\mathcal{G}R = (r_1, \ldots, r_{m-1})$. The Neyman-Pearson Lemma may then be applied with

$$\begin{aligned} g_0(x) &= f_{r_1}(x) \\ g_i(x) &= f_{r_1}(x) - f_{r_i}(x), \quad i = 1, \ldots, m-1 \\ g_m(x) &= f_U(x) \end{aligned}$$

and

$$\begin{aligned} b_i &= 0, \quad i = 1, \ldots, m-1 \\ b_m &= \alpha \end{aligned}$$

The most powerful test then has rejection region

$$f_{r_1}(x) > k_m f_U(x) + \sum_{i=1}^{m-1} k_i (f_{r_1}(x) - f_{r_i}(x)) \tag{2.5}$$

provided values for $k_1, \ldots, k_m$ can be found which leave the constraints (2.3) and (2.4) satisfied. If (2.5) is divided by $f_U(x)$, and we set $k_i = 1/(m-1)$ for $i = 1, \ldots, m-1$, then (2.5) can be written

$$\Gamma_R(x) = \sum_{(i_1, \ldots, i_K) \in I_K^o} L_R(x_{i_1}, \ldots, x_{i_K}) > k \tag{2.6}$$

where I_K^o is the set of all size K ordered subsets of $\{1, \ldots, N\}$. If R is symmetric (e.g. siblings) then summation can be taken over I_K^u, the unordered subsets. A test using rejection region (2.6) is permutation invariant, and therefore unbiased, satisfying (2.3). We then need only select k to satisfy (2.4), verifying the applicability of Theorem 1, concluding that the most powerful permutation invariant test rejects H_o for large values of $\Gamma_R(x)$.

It should be noted that the first moment under H_o is easily derived. Each term in Γ_R is a likelihood ratio of simple hypotheses, with $H_o : r = U$ being the common denominator hypothesis. Hence under H_o the expected value of each term is unity, so that the expected value of Γ_R is equal to the number of terms.

We note that Γ_R is a U-statistic of order K with kernel L_R. The theory of U-statistics was introduced in Hoeffding (1948), in which general conditions under which U-statistics are asymptotcally normal are given. Although these conditions will likely apply to at least most cases of Γ_R, in the cases considered in this article very large variance and severe skewness were

consistently observed, rendering such asymptotic results inapplicable in this case. Therefore, nonparametric methods for obtaining significance levels are needed, at least in the absence of a distribution theory specialized to this problem.

3 Testing for sibling relationships

In Almudevar and Field (1999) an algorithm is proposed for identifying sibling groups from a sample known to be within a single generation. The algorithm proceeds largely by systematically applying the principle of exclusion. It was found that larger sibling groups could be identified with small error given genotypic evidence of sufficient quality, as measured by number of loci and allelic diversity within loci. Thus, while large sibling groups can be identified, there remains the problem of the existence of spurious observed sibling groups of smaller size. In fact, any two individuals may form a feasible sibling group even if they share no alleles. We therefore propose a strategy of identifying any large sibling groups, removing them from the sample, then testing for the presence of smaller sibling groups among the remaining individuals. The results in Almudevar and Field (1999) suggest that sibling groups of size 2 or 3 can usefully be thought of as small, with groups of size 4 or more considered large, although this ultimately depends on the quality of the genotypic data used.

We adopt the concept of the *locally most powerful test*. This is a test which is most powerful for members of the alternative hypothesis closest to the boundary between null and alternative hypotheses. When the hypotheses are subsets of Euclidean space this has an exact meaning. In the problem considered here a hypothesis boundary can be reasonably defined in more than one way. The approach taken here is to maximize power for alternatives consisting of one single small sibling group. Sizes of 2 and 3 will be both considered.

We let R_K define a sibship relationship among K individuals, where interest is in $K = 2, 3$.

Specifically, we test

$$H_o: \quad r = U$$

$$H_a: \quad r \in \mathcal{G}R_K$$

using statistic Γ_{R_K} as derived in the previous section.

We now consider the question of calculating Q_{R_K}. In general, the calculation of the joint distribution of the genotypes of related individuals is usually accomplished using the concept of *identity by descent* introduced by Cotterman (1940) and extended by many authors, including Malecot (1948), Li and Sacks (1954), Cockerham (1971), Jacquard (1972), Nadot and Vaysseix (1973), Denniston (1974), Thompson (1974), Karigl (1981, 1982), Amos, Dawson and Elston (1990) and Whittemore and Halpern (1994). An identity state for a vector of genes on a pedigree is a complete specification of which genes are identical by virtue of having been inherited from a single gene in a common ancestor. The probabilities of these identity states are not dependent on population allele frequencies. A joint genotype distribution is then calculated by conditioning on these identity states. The identity state method becomes complex very quickly, so we will develop a specialized argument for the joint genotype distribution of siblings.

Suppose individuals $1, \ldots, K$ are a sibling group, with unrelated parents. Let (x_{i1}, x_{i2}) be the *ordered* genotype of individual i, with genes inherited from parents labeled 1 and 2 respectively. When we wish to identify a genotype as being from locus l a superscript will used to write (x_{i1}^l, x_{i2}^l). The gene arrays $(x_{11}, \ldots, x_{K1})$ and $(x_{12}, \ldots, x_{K2})$ are stochastically independent, so we concentrate on deriving the joint distribution of one of these arrays. We will assume independent loci (linkage equilibrium).

Suppose alleles $1, \ldots, \nu$ exist in the population. Let N_j be the number of genes of type j in the genotype of parent 1. Then, $\tilde{N} = (N_1, \ldots, N_\nu)$ is a multinomial vector, assuming

Hardy-Weinberg equilibrium. We then have conditional probability

$$P\left\{(x_{11},\ldots,x_{K1}) = (a_1,\ldots,a_K) \,|\, \tilde{N}\right\} = \prod_{i=1}^{K} \frac{N_{a_i}}{2}. \tag{3.7}$$

The density function f_1 of $(x_{11},\ldots,x_{K1})$ is obtained by taking the expected value of (3.7). This is simplified somewhat by noting that the sum of the multinomial vector $\tilde{N}$ is 2. If we let m_j be the number of alleles in $(a_1,\ldots,a_K)$ equalling j, then

$$f_1(a_1,\ldots,a_K) = \begin{cases} 0 & m_j > 0 \text{ for more than 2 } j \\ \frac{p_j p_k}{2^{K-1}} & m_j > 0,\ m_k > 0,\ j \neq k,\ m_n = 0 \forall n \neq j,k \\ \frac{(2^K-2)p_j^2 + 2p_j}{2^K} & m_j > 0,\ m_k = 0 \forall k \neq j \end{cases}$$

where p_a is the population frequency of allele a. Then by independence the density of the ordered genotypes is given by the product $f_1(x_{11},\ldots,x_{K1})f_1(x_{12},\ldots,x_{K2})$. We assume genotypic observations are unordered, so that the density of the array of *unordered* genotypes $((x_{11},x_{12}),\ldots,(x_{K1},x_{K2}))$ is

$$f((x_{11},x_{12}),\ldots,(x_{K1},x_{K2})) = \sum f_1(x_{11},\ldots,x_{K1})f_1(x_{12},\ldots,x_{K2})$$

where the summation is over all distinct arrays of ordered genotypes compatible with the array of unordered genotypes of the argument. Suppose G_e is the group of vectors of length $2K$ obtained by exchanging elements $2i-1$ and $2i$ for any set of $i \in \{1,\ldots,k\}$ of the vector $(1,2,\ldots,1,2)$. Then, denoting $\tilde{j} = (j_{11},j_{12},\ldots,j_{K1},j_{K2})$

$$f((x_{11},x_{12}),\ldots,(x_{K1},x_{K2})) = \frac{1}{2^h} \sum_{\tilde{j}\in G_e} f_1(x_{1j_{11}},\ldots,x_{Kj_{K1}})f_1(x_{1j_{12}},\ldots,x_{Kj_{K2}})$$

where h is the number of homozygous genotypes among $(x_{11},x_{12}),\ldots,(x_{K1},x_{K2})$. An individual term in Γ_R is then

$$L_R\left(x_{i_1},\ldots,x_{i_K}\right) = \prod_{l=1}^{N_L} \frac{\frac{1}{2^h}\sum_{\tilde{j}\in G_e} f_1(x^l_{i_1 j_{11}},\ldots,x^l_{i_K j_{K1}})f_1(x^l_{i_1 j_{12}},\ldots,x^l_{i_K j_{K2}})}{2^{K-h} p_{x^l_{i_1 1}} p_{x^l_{i_1 2}} \times \ldots \times p_{x^l_{i_K 1}} p_{x^l_{i_K 2}}}$$

$$= \frac{1}{2^{N_L K}} \prod_{l=1}^{N_L} \frac{\sum_{\tilde{j} \in G_e} f_1(x^l_{i_1 j_{11}}, \ldots, x^l_{i_K j_{K1}}) f_1(x^l_{i_1 j_{12}}, \ldots, x^l_{i_K j_{K2}})}{p_{x^l_{i_1 1}} p_{x^l_{i_1 2}} \times \ldots \times p_{x^l_{i_K 1}} p_{x^l_{i_K 2}}}.$$

For R_2 we have

$$f_1(x, y) = \frac{p_x p_y + p_x I\{x = y\}}{2}$$

resulting in test statistic

$$\begin{aligned} \Gamma_{R_2} &= \tfrac{1}{8^{N_L}} \textstyle\sum_{i<j} \prod_{l=1}^{N_L} \left(1 + \dfrac{I\{x^l_{i1} = x^l_{j1}\}}{p_{x^l_{i1}}}\right)\left(1 + \dfrac{I\{x^l_{i2} = x^l_{j2}\}}{p_{x^l_{i2}}}\right) \\ &+ \left(1 + \dfrac{I\{x^l_{i1} = x^l_{j2}\}}{p_{x^l_{i1}}}\right)\left(1 + \dfrac{I\{x^l_{i2} = x^l_{j1}\}}{p_{x^l_{i2}}}\right) \end{aligned} \tag{3.8}$$

Similarly, for R_3 we have

$$\begin{aligned} f_1(x, y, z) &= \frac{1}{4}(p_x p_y I\{x \neq y\} I\{y = z\} + p_y p_z I\{x \neq y\} I\{x = z\} \\ &\quad + p_x p_z I\{y \neq z\} I\{x = y\} + (3p_x^2 + p_x) I\{x = y\} I\{y = z\}). \end{aligned}$$

Letting

$$\begin{aligned} q_1(x, y, z) &= \frac{1}{4}[\frac{1}{p_z} I\{x \neq y\} I\{y = z\} + \frac{1}{p_x} I\{x \neq y\} I\{x = z\} \\ &\quad + \frac{1}{p_y} I\{y \neq z\} I\{x = y\} + \left(\frac{3}{p_x} + \frac{1}{p_x^2}\right) I\{x = y\} I\{y = z\}] \end{aligned}$$

we obtain

$$\begin{aligned} \Gamma_{R_3} &= \tfrac{1}{4^{N_L}} \textstyle\sum_{i<j<k} \prod_{l=1}^{N_L} q_1(x^l_{i1}, x^l_{j1}, x^1_{k1}) q_1(x^l_{i2}, x^l_{j2}, x^1_{k2}) \\ &+ q_1(x^l_{i1}, x^l_{j2}, x^1_{k1}) q_1(x^l_{i2}, x^l_{j1}, x^1_{k2}) \\ &+ q_1(x^l_{i1}, x^l_{j1}, x^1_{k2}) q_1(x^l_{i2}, x^l_{j2}, x^1_{k1}) \\ &+ q_1(x^l_{i1}, x^l_{j2}, x^1_{k2}) q_1(x^l_{i2}, x^l_{j1}, x^1_{k1}). \end{aligned}$$

As a numerical example a test case was constructed involving 20 individuals. Ten loci were assumed, each with 5 equiprobable alleles. Hardy-Weinberg and linkage equilibrium

were assumed for the simulations. A total of 11 relatedness configurations were used, defined by 0 to 5 sibling groups of size 2, then 0 to 5 sibling groups of size 3. A power curve for a level 0.05 test was constructed for Γ_{R2} and Γ_{R3}, in each case using actual frequencies and then frequencies estimated from the data. In addition, the test statistic T_3 proposed in Almudevar and Field (1999), defined as the number of subsets of size three which form genetically compatible sibling groups, was included. The above scenarios were simulated 10,000 times each. For each trial the five statistics were calculated. Power curves were calculated by determining the 95th percentile of each statistic from the null hypothesis (zero sibling group) trials. The results for the size 2 and size 3 trials are given in Figures 1 and 2 respectively. In both sets of trials T_3 is noticeably less powerful than both Γ_{R_2} and Γ_{R_3}. For the size 2 trials Γ_{R_2} is noticebly more powerful than Γ_{R_3}. For the size 3 trials, while Γ_{R_3} would be expected to be more powerful than Γ_{R_2}, the difference is not as great as in the size 2 trials, suggesting that the strategy of adopting a locally most powerful test by maximizing the power against the alternative of a single sibling group of size 2 is a sound one. Figures 1 and 2 also suggest that the power lost in using empirical frequencies in place of the (usually unknown) population frequencies is not too great, although there is a loss of power of about 0.1 when using Γ_{R_2} to test against a single sibling group of size 3.

Figure 1 here

Figure 2 here

The power curves in Figures 1 and 2 are largely of theoretical interest, leaving open the problem of calculating the significance for any given set of data. If we calculate the variance of Γ_{R_2} under H_o we obtain

$$\text{var}\,(\Gamma_{R_2}) = \left(\begin{array}{c} N \\ 2 \end{array} \right) \left\{ \left[\prod_{l=1}^{N_L} \frac{(3+v_l)^2 + 15 + v_l}{32} \right] - 1 \right\}$$

where v_l is the number of alleles at locus l. That the variance depends on the allele frequencies only through v_l is a mathematical consequence of the appearance of the frequencies p_x in the denominator in (3.8). That the variance should undergo a disproportionately large increase with the appearance of a single example of a new allele in a population, independently of population size, should preclude the use of any test procedure dependent on this variance. In addition, large skewness was consistently observed. Thus, it would seem appropriate to use nonparametric methods, using the power curve obtained from perfect knowledge of the null distribution as a theoretical benchmark.

We consider two alternatives, the first a bootstrap procedure, the second a conditional approach. For the moment we assume Hardy-Weinberg equilibrium in addition to independence of loci, so that under the null hypothesis of unrelatedness, the individual genes constitute a random sample, with gene frequencies conforming to a multinomial distribution. For the bootstrap procedure resampling can proceed by randomly selecting $2N$ genes with replacement, independently for each loci, assigning them at random to the N individuals, then calculating the statistic. Then H_o is rejected with significance level α if the observed value of the statistic is above the (1-α)100th percentile of the resampled statistics. The observed value itself is included in the bootstrap sample. There are two alternative methods, depending on whether the statistic is calculated using the observed frequencies from the sample or those observed in the bootstrap replication.

As for the conditional test, under H_o we note that the distribution of the alleles conditional on the observed gene totals can be reproduced by a random permutation of the genes (Levene, 1949; Guo and Thompson, 1992). This is equivalent to randomly reassigning the observed genes to individuals without replacement. This may be done independently across the loci. It should be noted that the resulting distribution is independent of the population allele frequencies, so that the significance level is calculable exactly in principle. In practice,

we simulate independent random permutations, recalculating the test statistic each time. The significance level is estimated by the proportion of simulated test statistics greater than or equal to the observed value. Again, the observed value is itself included in the sample. Then H_o is rejected if the observed value of the statistic is above the (1-α)100th percentile of the resampled statistics. Since the gene frequencies remain constant there is no need to consider alternative substitutions for the population frequencies as for the bootstrap procedure. The distinction between the bootstrap and the conditional procedure then reduces to whether resampling is done with or without replacement.

If Hardy-Weinberg or linkage disequilibrium is assumed, then either procedure can be suitably modified. In the case of Hardy-Weinberg disequilibrium resampling should take place at the genotype rather than at the gene level. The theoretical justification remains unchanged, although the actual test statistic need no longer be most powerful. If significant linkage is assumed between two loci, than the resampling can be suitably paired by the loci in question.

The procedures were evaluated using 5,000 trials from 20 individuals. The number of sibships of size 2 was allowed to vary from 0 to 5. These sibships constituted all the relatedness structure. Ten loci with 5 equiprobable alleles were used. Hardy-Weinberg and linkage equilibrium were assumed for the simulations. All tests were based on Γ_{R_2}. Both suggested bootstraps were done for each trial, sampling at the gene level with independent loci. In addition, the conditional test was performed, using resampling at both the gene and genotype level. Loci were assumed to be independent.

The resulting power curves for a test of size $\alpha = 0.05$ are given in Figure 3. For comparison, the theoretical power curve for Γ_{R_2} is included (plot A). The bootstrap procedure proved to be highly conservative (plots B and C). The observed test size, although designed to be 0.05 proved to be 0.0114 and 0.0104 using trial and resampled frequencies respectively.

On the other hand, the conditional test when conditioned on gene totals (plot D) reproduces very closely the theoretical power curve. The observed test size was 0.0536, slightly above the designed test size of 0.05. The conditional test when conditioned on genotype totals produces a similar power curve, with slightly greater loss in power, which is generally to be expected when making fewer assumptions. The observed test size was 0.0516. The conclusion is therefore that a conditional approach can produce accurate significance levels while reproducing near optimal power.

Figure 3 here

Finally, we compare the conditional test based on Γ_{R_2} to an unconditional test based on two statistics for relatedness which appear in the literature. The first is Queller and Goodknight's (1989) statistic r_{xy}. This is used to estimate genetic relatedness, a quantity important to the concept of inclusive fitness (Hamilton, 1964a, 1964b). In the present example it may be defined for two individuals x and y as

$$r_{xy} = \frac{\sum_l \sum_m (p_m^y - p_m^{-xy})}{\sum_l \sum_m (p_m^x - p_m^{-xy})}$$

where the summation $\sum_l \sum_m$ is taken over all alleles in all observed loci present in x, p_m^x and p_m^y are the frequencies of allele m present in x and y (ie. 0, 0.5 or 1), and p_m^{-xy} is the estimate of the frequency of m obtained by excluding x and y. The aggregate value of the statistic is obtained by summing r_{xy} over all ordered pairs of individuals (note that in general $r_{xy} \neq r_{yx}$). This procedure is used in, for example, Taylor *et. al.* (1997). The second relatedness statistic considered is the average number of shared alleles per locus M_{xy} (Blouin *et. al.*, 1996). Again, the aggregate statistic is taken as the sum of M_{xy} over all pairs. The scenario considered is the one used above with 20 individuals and varying numbers of sibships of size 2. The power curve is calculated by first estimating the distribution under the null

hypothesis and rejecting when the statistic exceeds the 95th percentile of this distribution. One thousand trials were used, with the resulting power curves given in Figure 4. The power curve for the conditional test (conditioned on gene totals) based on Γ_{R_2} for the same scenario is included. Clearly, the conditional Γ_{R_2} test is of greater power.

Figure 4 here

4 Parent/offspring inference

As a further example DNA markers from 17 sperm whale from a Galapagos Islands breeding ground were considered. There were 10 loci with heterozygosities of 0.75, 0.68, 0.87, 0.89, 0.75, 0.89, 0.91, 0.83, 0.78 and 0.62. Only two parent/offspring pairs exist which are not genetically excluded. In addition, the maximal sibling group algorithm proposed in Almudevar and Field (1999) was applied to the data, from which it was concluded that any sibling group of size 3 or greater can be genetically excluded. We test against two types of alternatived, R_2 (sibships of size 2) and P (parent/offspring pairs). Tests based on Γ_{R_2} and Γ_P were performed. The conditional approach (conditioned on gene totals) proposed in the previous section was used. To further study the difference between the two tests they were performed using randomly chosen subsets of the 10 available loci (except that all 10 distinct subsets of size 1 and 9 were used, as was the single subset of size ten). The average observed significance levels grouped by number of loci used are given in Figure 5. The significance levels are noticably smaller for Γ_P than for Γ_{R_2}. In particular, the average observed significance level is below 0.05 for 7 loci or more for Γ_{R_2}, but for Γ_P this occurs only for 9 or more loci. This suggests greater evidence of parent/offspring pairs than of sibling groups, which is suggested by the application of the exclusion principle. Certainly, this conclusion cannot be taken as definitive given the available evidence, but it does suggest the possibity of a more refined

inference of relationship types.

Figure 5 here

5 Discussion

A class of statistics for testing permutation invariant relatedness hypotheses against unrelatedness using genotypic data was proposed, which were shown to be of maximum power among permutation invariant tests. These tests lead naturally to "locally most powerful tests", in the sense that power is maximized for alternatives to unrelatedness which are not at a great statistical difference from unrelatedness. Simluation studies suggest that these tests can be more powerful than those based on other relatedness statistics appearing in the literature examined here. The large skewness exhibited by the null hypothesis distribution suggests the use of nonparametric alternatives. To that effect, it is found that when the test statistic is conditioned on gene or genotype totals the resulting conditional test gives accurate significance levels with little sacrifice in power. Some slight loss in power was detected when sampling at the genotype rather than the gene level, but this would be the recommended procedure when Hardy-Weinberg equilibrium is not assumed.

An experiment done with sperm whale data suggests that the type of relationships dominant in the sample can be inferred by applying tests for different types of relationships, although the procedure by which tests can be combined for inferential statements would need to be developed.

The recommended method of obtaining significance levels depends on independent sampling of random permutations, and the repeated calculation of a statistic with order N^K terms, which can be computationally time consuming. However, the technique of Monte Carlo Markov chains has been applied to this type of problem in Guo and Thompson (1992)

and Lazzeroni and Lange (1997), providing computationally efficient alternatives to independent sampling. It is anticipated that similar techniques can be used in the problem considered here.

6 Acknowledgements

The author wishes to thank Jenny Christal for allowing the use of the sperm whale data. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

7 Bibliography

Almudevar, A. and Field, C. (1999). Estimation of single generation sibling relationships based on DNA markers. *Journal of Agricultural, Biological and Enviornmental Statistics* **4**, 2, 136-165

Amos, C.I., Dawson, D.V. and Elston, R.C. (1990). The probabilistic determination of identity-by-descent sharing for pairs of relatives from pedigrees. *American Journal of Human Genetics* **47**, 842-853.

Blouin, M.S., Parsons, M., Lacaille, V. and Lotz, S. (1996). Use of microsatellite loci to classify individuals by relatedness. *Molecular Ecology* **5**, 393-401.

Cockerham, C. C. (1971). Higher-order probability functions of identity of alleles by descent. *Genetics* **69**, 235-246.

Cotterman, C. W. (1940). A calculus for statistico genetics. Ph.D. Thesis, Ohio State University. Published in *Genetics and Social Structure*, P. Ballonoff (ed.). (1975). Benchmark papers in Genetics, Dowden, Hutchinson and Ross.

Denniston, C. (1974). An extension of the probability approach to genetic relationships, one locus. *Theoretical Population Biology* **6**, 58-75.

Guo, S.W., and Thompson, E.A. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* **48**, 361-372.

Hamilton, W.D. (1964a). The genetical evolution of social behaviour. *Journal of Theoretical Biology* **7**, 1-16.

Hamilton, W.D. (1964b). The genetical evolution of social behaviour, II. *Journal of Theoretical Biology* **7**, 17-52.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics* **19**, 293-325.

Jacquard, A. (1972). Genetic information given by a relative. *Biometrics* **28**, 1101-1114.

Karigl, G. (1981). A recursive algorithm for the calculation of identity coefficients. *Annals of Human Genetics* **45**, 299-305.

Karigl, G. (1982). A mathematical approach to multiple genetic relationships. *Theoretical Population Biology* **21**, 379-393.

Lazzeroni, L.C. and Lange, K. (1997). Markov chains for Monte Carlo tests of genetic equilibrium in multidimensional contingency tables. *Annals of Statistics* **25**, 138-168.

Lehmann, E.L. (1986) *Testing Statistical Hypotheses, 2nd Ed.* John Wiley and Sons.

Levene, H. (1949). On a matching problem arising in genetics. *Annals of Mathematical Statistics* **20**, 91-94.

Li, C. C. and Sacks, L. (1954). The derivation of joint distribution and correlation between relatives by the use of stochastic matrices. *Biometrics* **10**, 347-360.

Malecot, G. (1948). Les Mathematiques de l'heredité. Paris: Masson et Cie.

Meagher, R.M. and Thompson, E. (1986). The relationship between single parent and

parent pair genetic likelihoods in genealogy reconstruction. *Theoretical Population Biology* **29**, 87-106.

Meagher, R.M. and Thompson, E. (1987). Analysis of parentage for naturally established seedlings of *chamaelirium luteum* (liliacae). *Ecology* **68**, 803-812.

Nadot, R. and Vaysseix, G. (1973). Apparentement et identité. Algorithme de calcul des coefficients d'identité. *Biometrics* **29**, 347-359.

Painter, I. (1994) Analysis of full-sibship configurations in a maternal half-sibship. Technical Report No. 272, Dept. of Statistics, University of Washington, Seattle, Washington.

Painter, I. (1996). Sibling reconstruction without parental information. *Journal of Agricultural, Biological, and Environmental Statistics* **2**, 2, 212-229.

Pamilo, P. (1989). Estimating relatedness in social groups. *Trends in Ecology and Evolution* **4**, 353-355.

Queller, D.C. and Goodknight, K.F. (1989). Estimating relatedness using genetic markers. *Evolution* **43**, 258-275.

Taylor, A.C., Horsup, A., Johnson, C.N., Sunnucks, P. and Sherwin, B. (1997). Relatedness structure detected by microsatellite analysis and attempted pedigree reconstruction in an endangered marsupial, the northern hairy-nosed wombat *Lasiorhinus krefftii*. *Molecular Ecology* **6**, 9-19.

Thompson, E. A. (1974). Gene identities and multiple relationships. *Biometrics* **30**, 667-680.

Thompson, E. (1976). Inference of genealogical structure. *Social Science Information* **15**, 2/3, 477-526

Whittemore, A. S. and Halpern, J. (1994). Probability of gene identity by descent: computation and applications. *Biometrics* **50**, 109-117.

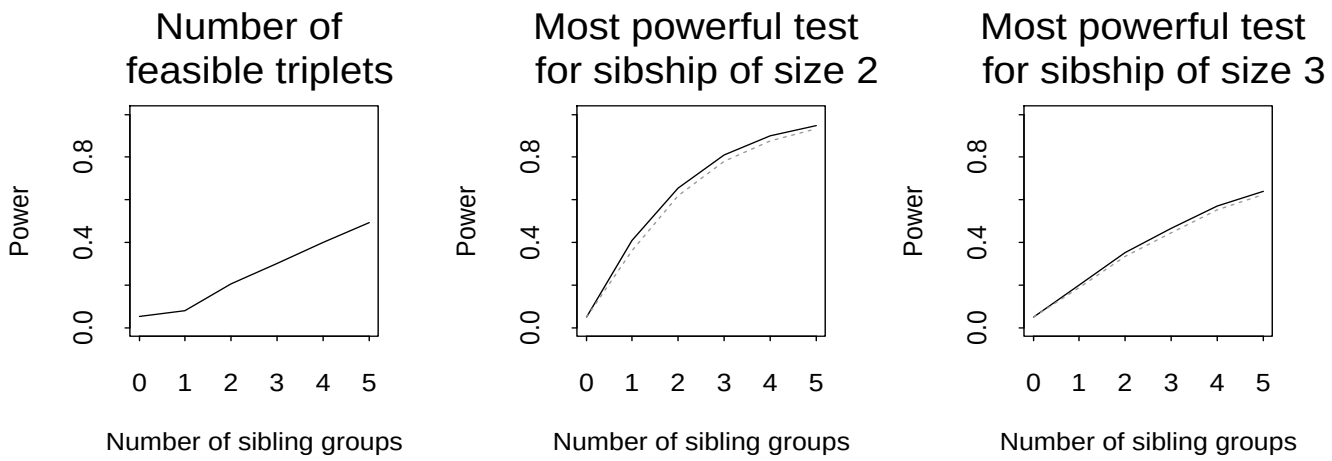

Figure 1: Power curves for sibships of size 2 from sample of size 20 using T_3 (plot 1), Γ_{R_2} (plot 2), Γ_{R_3} (plot 3). Test size is 0.05. Ten loci with 5 equiprobable alleles are assumed. The solid line represents the statistic using actual allele frequencies, while the dashed line represents the statistic with empirical frequencies substituted for the actual ones.

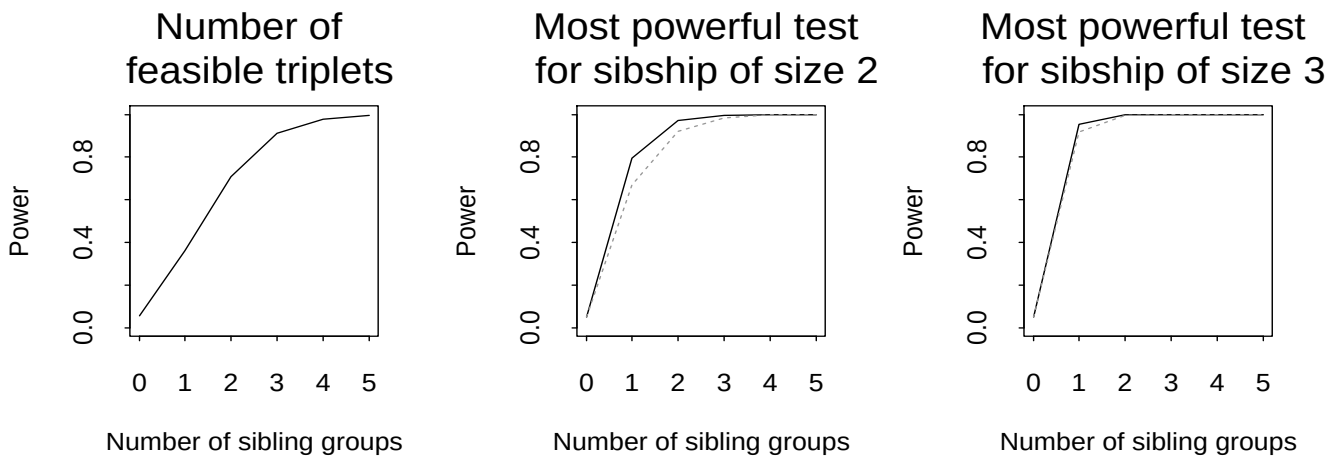

Figure 2: Power curves for sibships of size 3 from sample of size 20 using T_3 (plot 1), Γ_{R_2} (plot 2), Γ_{R_3} (plot 3). Test size is 0.05. Ten loci with 5 equiprobable alleles are assumed. The solid line represents the statistic using actual allele frequencies, while the dashed line represents the statistic with empirical frequencies substituted for the actual ones.

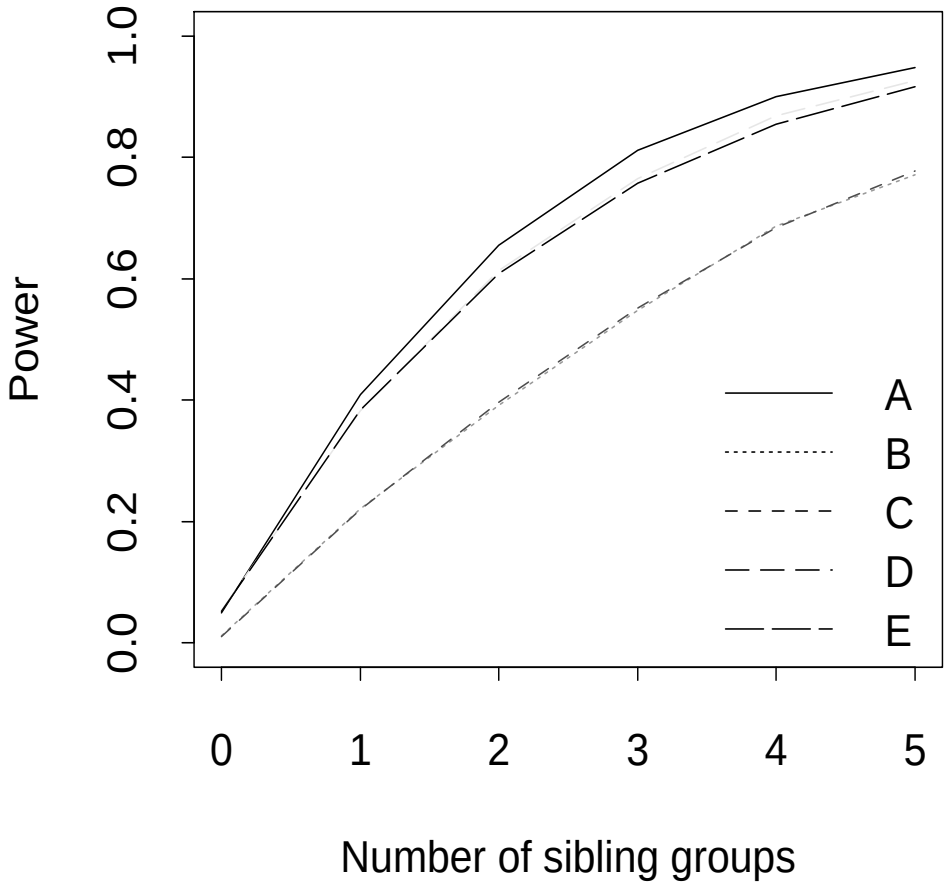

Figure 3: Power curves for sibships of size 2 from sample of size 20 using Γ_{R_2} (plot A), bootstrap procedure applied to Γ_{R_2} using trial frequencies and bootstrap frequencies (plots B and C), Γ_{R_2} conditioned on gene totals (plot D), Γ_{R_2} conditioned on genotype totals (plot E). Test size is 0.05.

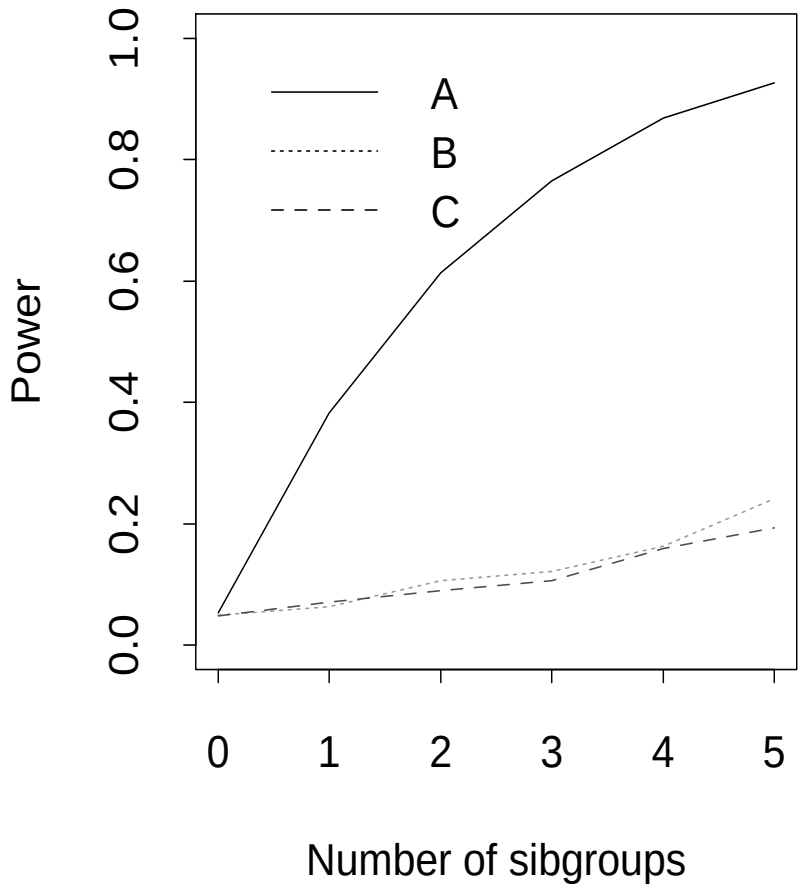

Figure 4: Power curves for sibships of size 2 from sample of size 20 using, Γ_{R_2} conditioned on gene totals (plot A), r_{xy} (plot B) and M_{xy} (plot C).

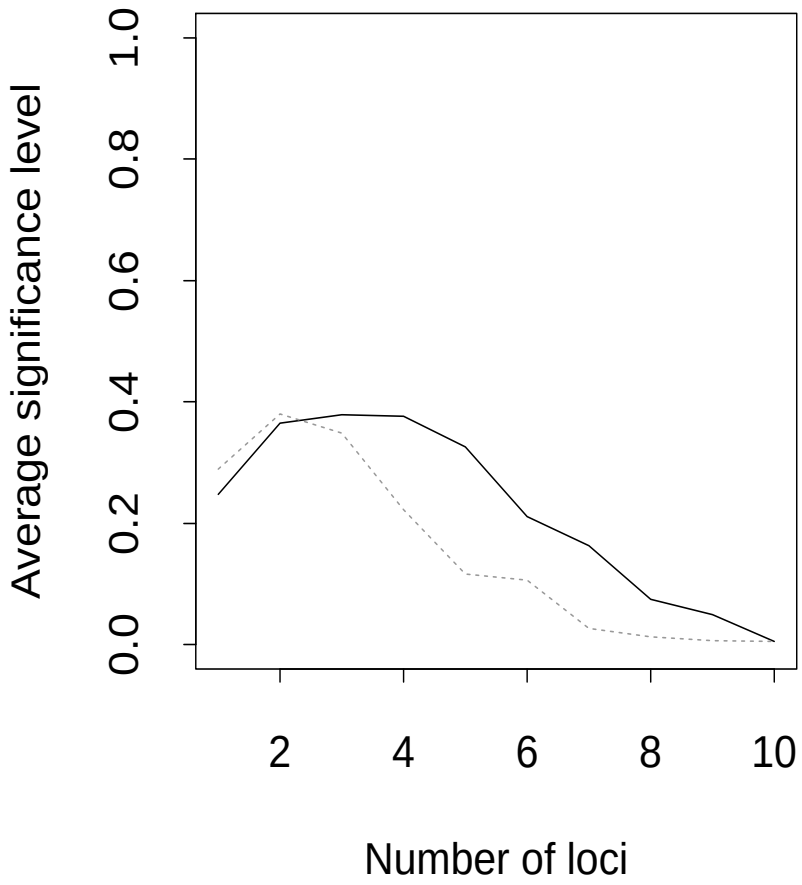

Figure 5: Average observed significance level for varying numbers of loci. The solid line represents Γ_{R_2} conditioned on gene totals. The dashed line represents Γ_P conditioned on gene totals.