Journal of Data Science **9**(2011), 23-41

Maximum Likelihood Estimation for Ascertainment Bias in Sampling Siblings

Balgobin Nandram[1], Jai-Won Choi[2] and Hongyan Xu[2]
[1] *Worcester Polytechnic Institute and* [2] *Medical College of Georgia*

Abstract: When there is a rare disease in a population, it is inefficient to take a random sample to estimate a parameter. Instead one takes a random sample of all nuclear families with the disease by ascertaining at least one affected sibling (proband) of each family. In these studies, an estimate of the proportion of siblings with the disease will be inflated. For example, studies of the issue of whether a rare disease shows an autosomal recessive pattern of inheritance, where the Mendelian segregation ratios are of interest, have been investigated for several decades. How do we correct for this ascertainment bias? Methods, primarily based on maximum likelihood estimation, are available to correct for the ascertainment bias. We show that for ascertainment bias, although maximum likelihood estimation is optimal under asymptotic theory, it can perform badly. The problem is exasperated in the situation where the proband probabilities are allowed to vary with the number of affected siblings. We use two data sets to illustrate the difficulties of maximum likelihood estimation procedure, and we use a simulation study to assess the quality of the maximum likelihood estimators.

Key words: Expectation-maximization algorithm, Nelder-Mead algorithm, population genetics, segregation ratio, truncated binomial distribution.

1. Introduction

When there is a rare disease in a population, it is inefficient to take a random sample to estimate a parameter of interest. Instead one takes a random sample of all nuclear families with the disease by ascertaining at least one sibling (proband) of each family. In such studies, an estimate of the proportion of siblings with the disease will be inflated. Sometimes the situation is even worse; the investigator takes all families that appear in the hospital. Thus, there is a selection bias (e.g., Patil and Rao, 1978).

Fisher (1934) illustrated the importance of adjusting for the selection bias. For a discussion of the problems of ascertainment bias in the analysis of family data, see Crow (1965). For example, studies of the issue of whether a rare disease shows an autosomal recessive pattern of inheritance, where the Mendelian segregation ratios are of interest, have been investigated for several decades. For a rare disease, the Mendelian segregation ratio is $p = 0.5$ for an autosomal dominant disease and $p = .25$ for an autosomal recessive disease. These follow from the first law of Mendel. For a rare disease one would be interested to know whether it is autosomal dominant or recessive. That is, whether $p = 0.5$ or $p = .25$ respectively. But because the disease is rare, the investigator will select all those nuclear families that appear. Then there is a selection bias; specifically the estimates will be inflated. How do we correct for this ascertainment bias? Methods, primarily based on maximum likelihood estimation, are available to correct for the ascertainment bias. See Lange (2002, chap. 2) and Sham (1998, chap. 2) for very clear pedagogy on this problem.

Table 1 gives a set of data which was presented by Fisher (1934) to illustrate the need to take account of the method of ascertainment in segregation analysis. The data consist of 340 families all with five offspring. The family was ascertained through at least one affected offspring. One can count the total number of offspring to be 1700, the total number of affected offspring to be 623, and the total number of probands to be 432. [Sham (1998) gave an incorrect total of 434.] Thus, one might estimate the segregation ratio to be $623/1700 = .3665$, and the ascertainment probability to be $432/623 = .6934$. Unfortunately, these simple estimates are too inflated. Sham (1998) also used these data for illustration. We note that Fisher (1934) did not state that the data are on albinism, but one might believe so because his work was motivated by the study of albinism. It is currently known that there are various forms of albinism in which chromosomes (11, 15, 13, 9, 10 and X) may become damaged or incomplete during mutation so that the proper proteins may not form, making the person albino. So that albinism does not come from a single chromosome. For illustration using these data, we will treat it as autosomal recessive as Fisher (1934) did.

Table 2 gives a set of data on cystic fibrosis which was presented by Crow (1965) to illustrate the need to take account of the method of ascertainment in segregation analysis. Cystic fibrosis is a hereditary disease that affects the mucus glands of the lungs, liver, pancreas, and intestines, causing progressive disability due to multisystem failure. The CFTR gene, found in Chromosome 7, is the cause of cystic fibrosis, where mutations result in proteins that are too short because of premature end to production. One can count the total number of offspring to be 269, the total number of affected offspring to be 124, and the total number of probands to be 90. Thus, one might estimate the segregation ratio to be $124/269 = .4610$, and the ascertainment probability to be $90/124 = .7258$. Again, these simple estimates are too inflated. Note that 46.1% which is far in excess of the 25% expected on simple recessive inheritance (cystic fibrosis is autosomal recessive). One reason for the excess is the ascertainment bias - the exclusion of families where the parents are heterozygous, but fail to have a homozygous recessive child. These would add to the number of normal children and thereby reduce the proportion affected. This data set was also used in Lange (2002) for illustration. Current data on cystic fibrosis of the same form from the state of Georgia are available, but because of confidentiality they cannot be used.

There are two major differences between the two data sets. First, in Fisher's data the family sizes are all the same, but in Crow's data the sample sizes vary from 1 to 10. There are 340 families in Fisher's data, but there are only 80 families in Crow's data. Therefore, because maximum likelihood estimation has optimal *asymptotic* properties, it may be more appropriate in Fisher's data.

We describe the ascertainment bias problem in the study of rare autosomal recessive disorders. It is almost always the case that a disease is inherited from carrier parents when the disease is rare in the entire population. The number of at-risk parents is usually small (i.e., the number of parents capable of producing affected siblings is very small relative to the number not capable of producing affected siblings). So if a sample is taken at random from the entire population, there could be no at-risk families. Hence, at-risk families are divided into two groups, those with at least one affected sibling and the other with no affected siblings. A sample is then drawn from the families with at least one affected sibling, thereby introducing an ascertainment bias. Thus, our two examples can be viewed in this manner, and as is evident in both examples, a direct estimate of the proportion of affected siblings will be too large; one needs to adjust for the ascertainment bias.

When all families with affected offspring are ascertained, we say that there is complete ascertainment. When there are families with affected offspring who are not probands, we say that there is incomplete ascertainment. Fisher (1934) first analyzed the data in Table 1 using complete ascertainment. His analysis was done using a truncated binomial distribution. However, Fisher (1934) also described a simpler method for the more appropriate incomplete ascertainment for these data. This discussion was further developed by Bailey (1951) and Morton (1959). In this paper, we will focus on incomplete ascertainment as is evident in data in both Tables 1 and 2. Crow (1965) pointed out the need to adjust for ascertainment bias and incomplete ascertainment for the cystic fibrosis data.

The key idea for the correction of ascertainment bias is to find the correct sampling distribution under the ascertainment bias. Let x represent the quantity being measured, A denote the ascertainment event, and θ a parameter. Without the ascertainment bias, $f(x \mid \theta)$ is the sampling distribution for a random sample. However, when there is an ascertainment bias, we need $f(x \mid \theta, A) = f(x, A \mid \theta)/f(A \mid \theta)$.

In general, the two sampling distributions $f(x \mid \theta, A)$ and $f(x \mid \theta)$ are different; $f(x \mid \theta, A)$ being the more appropriate sampling distribution. Correcting for ascertainment bias means that we need to construct the sampling distribution, $f(x \mid \theta, A)$. A simple example, introduced by Fisher (1934) for complete ascertainment, is on the number (r) of affected siblings in a family of size (s) in a binomial model with $r > 0$. Then, $f(r \mid \theta, A) = s!\theta^r(1-\theta)^{s-r}/\{r!(s-r)![1-(1-\theta)^s]\}$, $r = 1, \ldots, s$, where θ is the proportion of affected siblings, and A is the event that $r > 0$, leading to the binomial distribution truncated at 0. More importantly the binomial probabilities are being re-weighted (increased in this case) so that the mass points are $1, \ldots, s$; 0 is excluded.

The problem of ascertainment is not new to survey samplers. For finite population sampling, Sverchkov and Pfeffermann (2004) defined the sample and sample-complement distributions as two separate weighted distributions (Patil and Rao, 1978) for developing design consistent predictors of the finite population total; see also the more recent presentation (Pfeffermann and Sverchkov, 2007). Malec, Davis and Cao (1999) used a hierarchical Bayesian method to estimate a finite population mean for binary data. These works are not directly applicable to our situation, but the ideas they portray are important for the issues associated with ascertainment bias. For probability proportional to size (PPS) sampling Nandram (2007) implemented surrogate sampling techniques to provide simulated random samples by using a model which reverses the selection bias. Under PPS sampling, Nandram *et al.* (2006) used a method, developed by Chambers, Dorfman and Wang (1998), to do Bayesian predictive inference when a transformation is needed.

We wish to study how inference about the segregation ratio changes with the proband probability. So we consider two cases. In the first case we consider a single proband probability, and we discuss extensively maximum likelihood estimation. In the second case, we consider how inference about the segregation parameter will change when there are different proband probabilities. In fact, we allow the proband probabilities to depend on the number of affected siblings in each family. Because there are more parameters in the analysis of the same data, maximum likelihood estimation should be relatively inefficient.

In this paper we provide some new distribution results and algorithms on maximum likelihood estimation of the ascertainment bias problem in which we assume incomplete ascertainment. The plan of the rest of the paper is as follows. In the next section we review maximum likelihood estimation, and we present some new analytical results. Specifically, we discuss existence of maximum likelihood estimators, how to compute them, and what inferential difficulties exist. In the third section, we present numerical results and a simulation study. We also show how to incorporate different proband probabilities. As we will show, this task is particularly challenging for maximum likelihood estimation. The final section has a discussion where, in addition to a summary, we discuss how one might fix the problems associated with the maximum likelihood estimation procedure.

2. Theories and Methods

Thompson (1986) discussed many ascertainment models. In this paper, we discuss the simplest ascertainment model (Sham, 1998; Lange, 2002). Essentially Lange (2002) showed how to adjust for the ascertainment bias using the expectation-maximization (EM) algorithm (Dempster, Laird and Rubin, 1977); Sham (1998) used Fisher's scoring. We will introduce a couple new methods as well.

Suppose there are n families selected through ascertainment sampling. Letting the k^{th} ascertained family have s_k siblings, we assume that there are r_k affected siblings and a_k ascertained siblings. In Fisher's data the s_k are all equal 5, and for Crow's data s_k vary from 1 to 10. Let p denote the segregation probability and π the proband probability. Here, p is primarily the parameter of interest. The simplest ascertainment model specifies that

$$a_k \mid r_k, \pi \overset{ind}{\sim} \text{Binomial}(r_k,\ \pi) \text{ and } r_k \mid s_k, p \overset{ind}{\sim} \text{Binomial}(s_k,\ p),$$

$k = 1, \ldots, n$. Thus, let $u_k = s_k!/[a_k!(s_k - r_k)!(r_k - a_k)!]$, the joint probability mass function of (a_k, r_k) is

$$p(a_k,\ r_k \mid \pi, p, s_k) = u_k p^{r_k}(1-p)^{s_k - r_k}\pi^{a_k}(1-\pi)^{r_k - a_k}, \qquad (2.1)$$

$a_k = 0, \ldots, r_k, r_k = 0, \ldots, s_k,\ k = 1, \ldots, n$. Note that (2.1) provides the likelihood for any family without conditioning on whether it is ascertained or not. To incorporate the ascertainment bias, we need to adjust (2.1) to the support $1 \le a_k \le r_k \le s_k,\ k = 1, \ldots, n$. That is, the ascertainment event, $1 \le a_k \le r_k \le s_k,\ k = 1, \ldots, n$, is denoted by A.

Now, the probability that a family with s_k siblings is ascertained is $1 - (1 - p\pi)^{s_k}$, leading to the truncated probability mass function

$$p(a_k,\ r_k \mid \pi, p, s_k, A) = u_k p^{r_k}(1-p)^{s_k - r_k}\pi^{a_k}(1-\pi)^{r_k - a_k}/[1 - (1 - p\pi)^{s_k}], \quad (2.2)$$

$a_k = 1, \ldots, r_k,\ r_k = a_k, \ldots, s_k$. Note that in (2.2) $1 - (1 - p\pi)^{s_k}$ is simply the probability that $1 \le a_k \le r_k \le s_k,\ k = 1, \ldots, n$. Also, note that (2.2) provides the likelihood for a family that has been ascertained. Thus, in the terminology of missing data, while (2.1) is the complete data likelihood, (2.2) is the incomplete data likelihood.

This is a fairly long section, so we present a plan. First we present some new theoretical results on the properties of the joint probability mass function. Then we present maximum likelihood estimation; most of this is known, but we present some new ideas as well.

Henceforth, unless otherwise stated all inference will be conditioned on s_k and A. However, for convenience we will drop this notation.

2.1 Properties of the joint probability mass function

We present some properties of the joint probability mass function $p(a_k, r_k \mid p, \pi)$ in (2.2). Again, note that we still have the conditioning on s_k and A, but it will be eliminated for convenience. We provide some interpretations as well. We note that some of the results are new.

First, we consider the marginal distribution of r_k. Using (2.2), let $v_k = s_k!/[r_k!(s_k - r_k)!]$, the marginal probability mass function of r_k is

$$p(r_k \mid p, \pi) = v_k p^{r_k}(1-p)^{s_k - r_k}[1-(1-\pi)^{r_k}]/[1-(1-p\pi)^{s_k}], \ r_k = 1, \ldots, s_k;$$

all other points have zero probability. In Appendix A we show that

$$E(r_k \mid p, \pi) = s_k p \left\{1 + \pi(1-p)(1-\pi p)^{s_k - 1}/[1-(1-\pi p)^{s_k}]\right\}. \tag{2.3}$$

Thus, $E(r_k \mid p, \pi)$ is bigger than $s_k p$ with the discrepancy related to p, π and s_k. With some cumbersome algebraic manipulation, we also show in Appendix A that

$$\text{Var}(r_k \mid p, \pi) = s_k p(1-p)(1-Q_k),$$

where

$$\begin{aligned} Q_k &= \pi^2 p(1-p)(1-\pi p)^{s_k - 2}[1-(1-\pi p)^{s_k}]^{-1} \\ &\quad \times\{s_k/[1-(1-\pi p)^{s_k}] - (1-\pi)(2\pi - 1)/[\pi^2 p(1-p)]\}. \end{aligned}$$

Note that $Q_k \leq 1$ (i.e., Q_k is an adjustment factor). So that if $Q_k \geq 0$, then $\text{Var}(r_k \mid p, \pi) \leq s_k p(1-p)$, the situation in which $r_k \mid p \sim \text{Binomial}(s_k,\ p)$. For example, if $s_k = 1$, then $Q_k = \{1 - \pi p - 2\pi(1-\pi)\}/\pi p(1-\pi p)$. If, in addition, $\pi - \frac{3}{4} < \pi p < \frac{1}{2}$ (reasonable for autosomal recessive), then $0 \leq Q_k \leq 1$ and $\text{Var}(r_k \mid p, \pi) \leq p(1-p)$.

Also, for a family that has not been ascertained (i.e., $a_k = 0$), it is easy to show that

$$p(a_k = 0,\ r_k \mid \pi, p) = \frac{s_k! p^{r_k}(1-p)^{s_k - r_k}(1-\pi)^{r_k}}{r_k!(s_k - r_k)!\{(1-\pi p)^{s_k} - [(1-\pi)p]^{s_k}\}},$$

$r_k = 1, \ldots, s_k$. Here, $(1-\pi p)^{s_k} - [(1-\pi)p]^{s_k}$ is the probability of having at least one affected sibling in the k^{th} family with $a_k = 0$.

The marginal probability mass function of a_k is

$$p(a_k \mid p, \pi) = s_k!(\pi p)^{a_k}(1-\pi p)^{s_k-a_k}/\{a_k!(s_k-a_k)![1-(1-\pi p)^{s_k}]\},\ a_k = 1, \ldots, s_k;$$

all other points have zero probability. That is, $p(a_k \mid p, \pi)$ is a truncated binomial probability mass function. It is easy to show that

$$E(a_k \mid p, \pi) = s_k \pi p/[1-(1-\pi p)^{s_k}] \tag{2.4}$$

and

$$\mathrm{Var}(a_k \mid p, \pi) = s_k \pi p(1-\pi p)\{1-\{1+(s_k-1)\pi p\}(1-\pi p)^{s_k-1}/[1-(1-\pi p)^{s_k}]\}.$$

Thus, as expected, $\mathrm{E}(a_k \mid p, \pi)$ increases from $s_k \pi p$, and $\mathrm{Var}(a_k \mid p, \pi)$ decreases from $s_k \pi p(1-\pi p)$.

In Appendix B, we show that

$$\mathrm{Cov}(r_k, a_k \mid p, \pi) = \frac{\left[1-(1-\pi p)^{s_k-1}\{1+(s_k-1)\pi p\}\right] s_k \pi p(1-p)}{\{1-(1-\pi p)^{s_k}\}^2}.$$

We also show that $(1-\pi p)^{s_k-1}\{1+(s_k-1)\pi p\}$ is nonnegative. Thus, the correlation between a_k and r_k is nonnegative, and therefore, there may be important information about p (via the r_k) in π (via the a_k).

In fact, the conditional probability mass function of r_k given a_k is also interesting. It is easy to show that

$$p(r_k \mid a_k, p, \pi) = \frac{(s_k-a_k)!\{(1-\pi)p\}^{r_k-a_k}(1-p)^{s_k-r_k}}{(r_k-a_k)!(s_k-r_k)!(1-\pi p)^{s_k-a_k}},$$

$r_k = a_k, \ldots, s_k$. Therefore, $r_k - a_k \mid a_k, p, \pi \sim \mathrm{Binomial}\{s_k - a_k, (1-\pi)p/(1-\pi p)\}$. Then

$$\mathrm{E}(r_k \mid a_k, p, \pi) = p(1-\pi)s_k/(1-\pi p) + a_k[1-p(1-\pi)/(1-\pi p)]$$

and

$$\mathrm{Var}(r_k \mid a_k, p, \pi) = (s_k - a_k)p(1-p)(1-\pi)/(1-\pi p)^2.$$

Thus, in the conditional probability mass function, expectation increases with a_k and variance decreases with a_k [i.e., a knowledge of a_k is informative, consistent with Sham (1998)]. Sham (1998) used Fisher's data to illustrate this issue, but here we have obtained an analytical argument.

Finally, letting $a = \sum_{k=1}^n a_k$, $r = \sum_{k=1}^n r_k$ and $s = \sum_{k=1}^n s_k$, without selection bias the maximum likelihood estimators of p and π are $\hat{p} = r/s$ and $\hat{\pi} = a/r$ respectively. These are the MLEs under the model without the ascertainment bias in (2.1). We will denote the MLEs with selection bias by $\tilde{p}$ and $\tilde{\pi}$, which are to be determined. These are the MLEs under the model with the ascertainment bias in (2.2).

2.2 Estimation procedures

We discuss maximum likelihood estimation of p and π under the reasonable assumption that the families are sampled independently. This is the same assumption used throughout the historical development since the pioneering work of Fisher (1934); see Lange (2002, chap. 2) and Sham (1998, chap. 2). Then, the likelihood function for all ascertained families is

$$\text{Likelihood}(p,\ \pi) = \prod_{k=1}^{n} \{p^{r_k}(1-p)^{s_k-r_k}\pi^{a_k}(1-\pi)^{r_k-a_k}/[1-(1-p\pi)^{s_k}]\}. \quad (2.5)$$

It is pertinent for us to show that if $a > n$, the maximum likelihood estimators (MLE) of p and π exist. For example, if $a_k = 1,\ k = 1, \ldots, n$, MLEs may not exist. That is, if exactly one sibling is ascertained in each family, MLEs may not exist. Also, if each family has exactly one sibling, the likelihood function is a constant in the unit square (i.e., $0 \le p, \pi \le 1$), and every point in the unit square is an MLE (i.e., the MLE is not unique). It is true in both Fisher's and Crow's examples that $a > n.$ To prove the existence of the MLEs, we note that because $1-(1-\pi p)^{s_k}$ is increasing in s_k and $s_k \ge 1,\ \ \pi p \le 1-(1-\pi p)^{s_k}$. Thus, using (2.5)

$$\prod_{k=1}^{n} p^{r_k}(1-p)^{s_k-r_k}\pi^{a_k}(1-\pi)^{r_k-a_k}/[1-(1-p\pi)^{s_k}] \le p^{r-n}(1-p)^{s-r}\pi^{a-n}(1-\pi)^{r-a},$$

where $a = \sum_{k=1}^{n} a_k$, $r = \sum_{k=1}^{n} r_k$ and $s = \sum_{k=1}^{n} s_k$. The maximum point (p, π) of the likelihood function exists (inside the unit square) if $r > n$, $s > r$, $a > n$ and $r > a$. This is true because the function $x^{g-1}(1-x)^{h-1}, 0 \le x \le 1$ if $g > 1$ and $g + h > 2$. But because $a \le r \le s$, $a > n$ suffices.

There are at least four methods to find the maximum likelihood estimators of p and π. One can use an optimization routine such as Nelder-Mead algorithm or Newton's method directly. Sham (1998) used a Fisher scoring algorithm, and Lange (2002) used the EM algorithm. We have developed a much simpler algorithm.

It is worth noting here that if we differentiate the log-likelihood function in (2.5) to obtain the maximum likelihood estimators of p and π, we need to solve the two equations simultaneously

$$s(\hat{p}-p) = \pi p(1-p)q \text{ and } r(\hat{\pi}-\pi) = p\pi(1-\pi)q, \quad (2.6)$$

where $q = \sum_{k=1}^{n}\{s_k(1-p\pi)^{s_k-1}\}/\{1-(1-p\pi)^{s_k}\}$. These are the equations that constitute our new iterative method. We start with p set at $\hat{p}$ and π set at $\hat{\pi}$ in the left-hand sides of these equations to update the right-hand sides of these equations. Thus, it is mathematically clear that $\hat{p}$ and $\hat{\pi}$ are inflated by $\pi p(1-p)q/s$ and $p\pi(1-\pi)q/r$ respectively, thereby accounting for the ascertainment bias. More importantly, it is easy to solve these equations iteratively by simply replacing p and π in the right-hand sides of these equations and updating the left-hand sides accordingly; it is sensible to start with $p = \hat{p}$ and $\pi = \hat{\pi}$.

In fact, we have maximized the logarithm of the likelihood function of $(p,\ \pi)$ in (2.5) directly using the Nelder-Mead algorithm (Nelder and Mead, 1965) to get the maximum likelihood estimators $(\tilde{\pi},\ \tilde{p})$. Unlike Newton's and the Fisher scoring algorithm, the Nelder-Mead algorithm is derivative-free; both Newton's and Fisher scoring need the first derivative and while Newton's method need the Hessian matrix, the Fisher scoring algorithm needs the information matrix (i.e., expected value of the negative Hessian matrix). Both of these methods are rather inefficient near the boundaries of the parametric space (e.g., p or π near 0 or 1).

Lange (2002) used the expectation-maximization (EM) algorithm. However, he has used an additional assumption in the EM algorithm. His key argument is, "If we view ascertainment as a sampling process in which unascertained families of size s_k are discarded one by one until the k^{th} ascertained family is finally ascertained, then the number of unascertained families discarded before reaching the k^{th} ascertained family follows a shifted geometric distribution with success probability $1-(1-\pi p)^{s_k}$." His EM algorithm gives the MLEs by solving

$$p = \sum_{k=1}^{n}\{r_k + s_k p(1-\pi)(1-\pi p)^{s_k-1}/[1-(1-\pi p)^{s_k}]\}\ /\sum_{k=1}^{n}\{s_k/[1-(1-\pi p)^{s_k}]\},$$

$$\pi = \sum_{k=1}^{n} a_k/\sum_{k=1}^{n}\{r_k + s_k p(1-\pi)(1-\pi p)^{s_k-1}/[1-(1-\pi p)^{s_k}]\} \tag{2.7}$$

iteratively as in (2.6). No measure of variability was presented, and any measure of variability will be too small because of the additional assumption. Essentially, Lange (2002) assumes that the missing sibship sizes are known, but he did no say this explicitly. In fact, no EM algorithm exists in the original model with missing sibling sizes.

However, we observe that it is much easier to solve the MLE equations in (2.6) by first updating p only. Using (2.6) we get

$$\pi p = 1 - \{(1-p)/(1-\hat{p}\}(1-\hat{\pi}\hat{p}).$$

Substituting πp into (2.6), and solving for p and π, we get

$$p = \hat{p} - \{p - \hat{p}(1 - \hat{\pi})/(1 - \hat{\pi}\hat{p})\} \, s^{-1} \sum_{k=1}^{n} s_k w/[(1 - \hat{p})^{s_k} - w] \tag{2.8}$$

and

$$\pi = \hat{\pi} - (1 - \hat{\pi})(\hat{p} - p)/[p(1 - \hat{p})], \tag{2.9}$$

where $w = \{(1 - p)(1 - \hat{\pi}\hat{p})\}^{s_k}$. Thus, we start with p set at $\hat{p}$ in the right-hand side of (2.8) to obtain p on the left-hand side, and iterate until convergence to $\tilde{p}$. Then, we substitute $\tilde{p}$ into (2.9) to get $\tilde{\pi}$ without iterations. Of course, convergence is much faster than updating (p, π) simultaneously.

It is also easy to find the negative inverse Hessian matrix to get an approximation for the covariance matrix of $(\tilde{\pi},\ \tilde{p})$. Sham (1998) gave a form for the standard errors from his Fisher scoring, but he did not present the correlation between the estimators. Lange (2002) gave the EM algorithm, but he did not present any measure of precision of his estimators. It is a standard practice to use the inverse negative Hessian matrix, evaluated at the MLEs to get an approximation of the covariance matrix. Thus, it does not matter which method is used to get the MLEs, the covariance matrix is the same. In Appendix B we present the covariance matrix. We note in Appendix B that if the MLEs exist, the covariance matrix will be positive definite, and therefore, the MLEs are unique.

In Appendix B we have also shown that a sufficient condition for the correlation between $\tilde{p}$ and $\tilde{\pi}$ to be nonnegative is $s_k - 1 \geq 4\pi p$. In the study of autosomal recessive typically $\pi, p \leq .50$ and πp is greater than a number which is smaller than 1/4. So that a sufficient condition for nonnegativity is that $s_k \geq 2$. This excludes families with one sibling, but if there are not too many of these, the correlation will be nonnegative.

3. Results

This section has three parts. First, we present numerical results for Fisher's data and Crow's data. Second, we perform a simulation study to assess the performance of the maximum likelihood estimators of the segregation ratio and proband probabilities. Third, we show that there are further difficulties of the maximum likelihood estimators when the proband probabilities vary with the number of affected siblings within a family.

3.1 Numerical results

Essentially we have used all the numerical methods we have discussed, and we have found that they gave the same estimates of the MLEs. Specifically, it is good that they agree with the Nelder-Mead algorithm.

For Fisher's data, the EM algorithm gives $\tilde{p} = .253$, $\tilde{\pi} = .475$; these are consistent with the estimates provided by Sham (1998). Their standard errors of .0129 and .0310 are also consistent with the ones we obtained. We also have a reasonable correlation of .250 between $\tilde{p}$ and $\tilde{\pi}$. We have used both the Fisher's information and the negative inverse Hessian matrix (without taking expectation) to compute the standard errors and correlation; they are in perfect agreement. Sham (1998) used Fisher's information to obtain the standard errors of $\tilde{p}$ and $\tilde{\pi}$, but he did not present the correlation between $\tilde{p}$ and $\tilde{\pi}$. We note that for the case in which the information on the probands are ignored, Sham (1998) reported standard errors of .0286 and .2400, showing large gains in precision in the method that includes the probands.

For Crow's data, the EM algorithm gives $\tilde{p} = .268$, $\tilde{\pi} = .359$; the standard errors are respectively .0347 and .0814 with a small correlation of .248. These are consistent with the estimates given by Lange (2002); the standard errors were not provided. As pointed by Lange (2002), these estimates are consistent with the theoretical value of .25 for an autosomal recessive as in cystic fibrosis.

It is possible to provide approximate 95% confidence interval for p by using the asymptotic normality of maximum likelihood estimators. We have used the intervals $\tilde{p} \pm 1.96STE(\tilde{p})$ where $\tilde{p}$ is the maximum likelihood estimator and $STE(\tilde{p})$, the standard error, obtained from the information matrix. Similarly for π, we have used the interval $\tilde{\pi} \pm 1.96STE(\tilde{\pi})$, where $\tilde{\pi}$ is the maximum likelihood estimator and $STE(\tilde{\pi})$, the standard error, obtained from the inverse negative Hessian matrix.

For Fisher's data the approximate 95% confidence interval for p is $.253 \pm 1.96 \times .0129$ which gives (.228, .278). Note that the 95% credible interval for p contains .250, consistent with an autosomal recessive inheritance. For Crow's data the approximate 95% confidence interval for p is $.268 \pm 1.96 \times .0347$ which gives (.200, .336). We also consider inference about π. For Fisher's data the approximate 95% confidence interval is $.475 \pm 1.96 \times .031$ which gives (.414, .536). For Crow's data the approximate 95% confidence interval is $.359 \pm 1.96 \times .081$ which gives (.200, .519). Note, as for Fisher's data, the 95% credible interval for p contains .250, consistent with an autosomal recessive model.

3.2 Simulation study

We have performed a small simulation study to assess the performance of the maximum likelihood estimation procedure. We have generated data from the model with ascertainment bias in (2.2), and we have fit the model using maximum likelihood estimation procedure. We have taken $p = .257$, $\pi = .371$ to obtain data similar to Crow's data. To study the effect of the sample size n, we have taken $n = 25, 50, 100, 200$; smaller values of n should challenge the maximum likelihood procedure.

We have generated 1000 data sets from the model that includes the ascertainment bias. From Crow's data, we have obtained the distribution of the ten family sizes $1,\ 2,\ldots,10$. The frequencies of the family sizes are $9, 24, 16, 13, 9, 2, 4, 1, 1, 1$. Thus, using the table method, we draw n family sizes for each of the 1000 simulated data sets. Now, noting that

$$p(a_k, r_k \mid p, \pi) = p(a_k \mid p, \pi) p(r_k \mid a_k, p, \pi),$$

we use the composition method to draw a_k from $p(a_k \mid p, \pi)$, and with this value of a_k, we draw r_k from $p(r_k \mid a_k, p, \pi)$, where

$$p(a_k \mid p, \pi) = s_k!(\pi p)^{a_k}(1{-}\pi p)^{s_k-a_k}/a_k!(s_k-a_k)![1-(1-\pi p)^{s_k}],\ a_k = 1,\ldots,s_k,$$

and

$$p(r_k \mid a_k, p, \pi) = \frac{(s_k-a_k)!\{(1-\pi)p\}^{r_k-a_k}(1-p)^{s_k-r_k}}{[(r_k-a_k)!(s_k-a_k)!(1-\pi p)^{s_k-a_k}]},$$

$r_k = a_k,\ldots,s_k$. Here $p(a_k \mid p, \pi)$ is a truncated binomial probability mass function; so we draw a_k from Binomial$(s_k, \pi p)$ and accept it if it is larger than 0. Because $r_k - a_k \mid a_k, p, \pi \sim$ Binomial$\{s_k - a_k, (1-\pi)p/(1-\pi p)\}$, we first draw $r_k - a_k$ from this binomial probability mass function and add a_k to it. We repeat this process for all n families.

In Table 3 we present the results for the simulation study. We consider each measure in turn. As expected, the MLE's for p and π converge respectively to the true values. However, $\tilde{p}$ is closer to the true value than $\tilde{\pi}$ for all sample sizes; π is noticeable far away for $n = 25$. As it must be, the standard errors go down with increasing n (so must be the widths). The coverage is not so good for $n = 25$ or $n = 50$ when p is estimated; this is worse when π is estimated. The MSEs seem fine for p, but off for π especially at $n = 25$ and $n = 50$.

Therefore, as expected for small sample sizes the maximum likelihood estimation does not perform well. However, as expected maximum likelihood estimation procedure does perform well for larger sample sizes. In fact, we have found that for small sample sizes, the lower end of the 95% confidence intervals under maximum likelihood estimation are smaller than 0, a standard problem with maximum likelihood estimation. An interval extended below zero has to be truncated at zero, and extended beyond two standard errors from the maximum likelihood estimate to get the nominal coverage of 95%; in practice just the truncation is done.

3.3 Unequal proband probabilities

We now show how to allow the proband probabilities to vary with the number of affected siblings. In Fisher's data there are five different values (1, 2, 3, 4, 5) for the number affected, and in Crow's data there are four values (1, 2, 3, 4) for the number affected. So for Fisher's data there are five different parameters $(\pi_1, \ldots, \pi_5)$, and for Crow's data there are four different parameters $(\pi_1, \ldots, \pi_4)$. Thus, generally let π_{r_k} denote the proband probabilities, and d be the number of distinct proband probabilities $(\pi_1, \ldots, \pi_d)$.

Then, with this simple adjustment the likelihood function for n families is

$$\text{Likelihood}(p,\ \tilde{\pi}) = \prod_{k=1}^{n} p^{r_k}(1-p)^{s_k - r_k}\pi_{r_k}^{a_k}(1-\pi_{r_k})^{r_k - a_k}/[1-(1-p\pi_{r_k})^{s_k}]. \quad (3.1)$$

We consider finding the MLEs of $p, \pi_1, \ldots, \pi_d$. We have used the Nelder-Mead minimization algorithm and an iterative method like the one we developed for a single π based on setting the first derivatives equal 0; see (2.8) and (2.9).

For Crow's data, $\tilde{p} = .299$, $\tilde{\pi}_1 = 1.000$, $\tilde{\pi}_2 = .301$, $\tilde{\pi}_3 = .361$, $\tilde{\pi}_4 = .000$ after 487 iterations when the Nelder-Mead algorithm is used, and $\tilde{p} = .294$, $\tilde{\pi}_1 = 1.000$, $\tilde{\pi}_2 = .300$, $\tilde{\pi}_3 = .361$, $\tilde{\pi}_4 = .000$ (about 10 iterations for our iterative method). We have virtually the same answers; there is a small difference for p. Unfortunately, both methods converge at the boundary of the parameter space. Crow's data are very sparse and the maximum likelihood estimation procedure is fallible. For Fisher's data, when the Nelder-Mead algorithm and our iterative method are applied, the MLEs are exactly the same, and they are $\tilde{p} = .274$, $\tilde{\pi}_1 = 1.000$, $\tilde{\pi}_2 = .528$, $\tilde{\pi}_3 = .298$, $\tilde{\pi}_4 = .375$, $\tilde{\pi}_5 = .278$. The Nelder-Mead algorithm converges after 669 iterations, and our method converges in about 10 iterations. But again convergence occurs on boundary of the parameter space.

Finally, we make two important observations. First, for maximum likelihood estimation, it is difficult to do further inference (e.g., standard errors and 95% confidence intervals cannot be obtained). The situation is worse for both Newton's method and the Fisher scoring algorithm because they need respectively the Hessian matrix and the information matrix which can not be computed; see Appendix E for some insight. Second, the issue of using different proband probabilities is an important one. For Crow's data when a single proband probability is assumed, $\tilde{p} = .268$ and $\tilde{\pi} = .359$; .299 differs considerably from .268. Also, for Fisher's data when a single proband probability is assumed, $\tilde{p} = .253$ and $\tilde{\pi} = .475$; again .274 differs considerably from .253.

4. Discussion

When one wants to find out about the proportion of people with a rare disease, one cannot take a random sample from the population. It is convenient to take a random sample of the cases that appear in a doctor's office. Thus, clearly this sample is biased (i.e., there is an ascertainment bias). An important example in genetics occurs when one is interested in the segregation ratio for a rare recessive disease. This problem exists over a century, and there are many solutions depending on the sampling scheme. More generally the selection bias problem is important when a non-random sample is taken from a population as in population genetics.

We have considered the problem of estimating the segregation ratio and the proband probabilities when there is an autosomal recessive disease. We have summarized some approaches for finding MLEs in the ascertainment bias problem, and we have provided some new theoretical results. We have also provided a new algorithm, potentially faster, and we have presented some new interpretations of the associated formulas. We also considered the case in which the proband probabilities change with the number of affected siblings in each at-risk family. This is a challenge for asymptotic theory as in maximum likelihood estimation. The two popular methods, Newton's method or the Fisher scoring method, can not be used in this case. The use of different proband probabilities can lead to changes in inference about the segregation parameter over the case of a single proband probability. This is true for both Fisher's data and Crow's data, more so for Fisher's data.

Finally, we discuss an alternative Bayesian procedure to maximum likelihood estimation. The basic difference between Bayesian method and maximum likelihood estimation is that in the Bayesian method the parameters are random, and they have prior distributions which arise from historical data and may not be informative. The prior distributions permit some flexibility with small sample sizes and many parameters (e.g., when the proband probabilities are allowed to vary with the number of affected siblings). For a single proband probability, we can take $p, \pi \overset{iid}{\sim} \text{Beta}(\alpha,\ \beta)$ where α and β are to be specified. For example, $\alpha = 1,\ \beta = 1$ gives a noninformative and proper prior, and $\alpha = 1/2,\ \beta = 1/2$ gives Jeffrey's prior. In general, one can incorporate important prior information using this prior distribution, and this can remove all the difficulties associated with maximum likelihood estimation procedure. Once a decision on a model is made, all the information about the parameters exist in their joint posterior distribution which is obtained using Bayes' theorem. The Bayesian method is more versatile than standard maximum likelihood estimation. Two advantages of Bayesian methods are they have simple interpretation, and they enjoy the recent development in Markov chain Monte carlo methods, the workhorse in Bayesian data analysis. We have shown, not presented in this paper because of space restriction, that the Bayesian procedure does overcome some of the difficulties associated with maximum likelihood estimation procedure, especially in the case where there are different proband probabilities.

In addition, not only the problem with unequal proband probabilities the Bayesian procedure can solve, but it will permit us to solve two more problems. First, we can include a familiar correlation in our model; for example, see Nandram and Choi (2005). The number of affected siblings in each family is not quite a binomial random variable. It is expected that one sibling getting affected will be related to the other siblings because they are in the same nuclear family sharing the same genes. For this problem, we have seen some improvements of the Bayesian procedure over the maximum likelihood procedure as well. Second, we can consider ascertainment bias that occurs in single nucleotide polymorphism (SNP) discovery, one of the issues that motivated this work. In SNP discovery a small sample of people is taken from the population, and these individuals are sequenced for a large number ($\approx 10^6$) of nucleotides. However, because of the low density of polymorphisms, many of the nucleotides of the panel are not polymorphic in the panel, and so they are eliminated from the panel. The discovery goes on to sequence a larger sample for the variable nucleotides (i.e., the remaining nucleotides). But, if the panel sample was larger, some of the discarded nucleotides could have been polymorphic. Thus, there is an ascertainment bias; for example, see Signorovitch (2003) for a description of this problem. The Bayesian procedure can be implemented to solve this problem, and we have some on-going activities in this area.

Appendix A. Derivation of Var$(r_k \mid p,\ \pi)$ and Cov$(a_k, r_k \mid p,\ \pi)$

To derive $\mathrm{Var}(r_k \mid p,\ \pi)$, we use the well-known identity

$$\mathrm{Var}(r_k \mid p, \pi) = \mathrm{E}\{r_k(r_k - 1) \mid p, \pi\} + \mathrm{E}(r_k \mid p, \pi)\{1 - \mathrm{E}(r_k \mid p, \pi)\}. \tag{A.1}$$

After some algebraic simplification, we have

$$\mathrm{E}(r_k \mid p, \pi) = s_k p\{1 + \pi(1-p)(1-\pi p)^{s_k - 1}/[1-(1-\pi p)^{s_k}]\} \tag{A.2}$$

and

$$\mathrm{E}\{r_k(r_k - 1) \mid p, \pi\} = s_k(s_k - 1)p^2[1-(1-\pi)^2(1-\pi p)^{s_k-2}]/[1-(1-\pi p)^{s_k}]. \tag{A.3}$$

By substituting (A.2) and (A.3) into (A.1) with further algebraic simplification, we get

$$\mathrm{Var}(r_k \mid p, \pi) = s_k p(1-p)(1-Q_k), \tag{A.4}$$

where

$$\begin{aligned} Q_k &= \pi^2 p(1-p)(1-\pi p)^{s_k-2}[1-(1-\pi p)^{s_k}]^{-1} \\ &\quad \times\{s_k/[1-(1-\pi p)^{s_k}]-(1-\pi)(2\pi-1)/[\pi^2 p(1-p)]\}. \end{aligned}$$

To derive $\mathrm{Cov}(a_k, r_k \mid p,\ \pi)$, we use the well-known identity

$$\mathrm{Cov}(a_k, r_k \mid p,\ \pi) = \mathrm{E}(a_k r_k \mid p,\ \pi) - \mathrm{E}(a_k \mid p,\ \pi)\mathrm{E}(r_k \mid p,\ \pi). \tag{A.5}$$

It is easy to show that

$$\mathrm{E}(a_k r_k \mid p,\ \pi) = s_k p\pi\{1+(s_k-1)p\}/[1-(1-p\pi)^{s_k}]. \tag{A.6}$$

Then, substituting (A.6), the formulae for $\mathrm{E}(a_k \mid p,\ \pi)$ in (2.3) and $\mathrm{E}(r_k \mid p,\ \pi)$ in (2.4) into (A.5), and simplifying, we get

$$\begin{aligned} \mathrm{Cov}(r_k, a_k \mid p, \pi) &= \left[1-(1-\pi p)^{s_k-1}\{1+(s_k-1)\pi p\}\right] \\ &\quad \times\ s_k \pi p(1-p)\{1-(1-\pi p)^{s_k}\}^{-2}. \end{aligned} \tag{A.7}$$

It is interesting to show that $\mathrm{Cov}(r_k, a_k \mid p, \pi)$ is nonnegative. We only need to show that $(1-\pi p)^{s_k-1}\{1+(s_k-1)\pi p\} \le 1$. But because $(1-\pi p)^{s_k-1}\{1+(s_k-1)\pi p\} = 1$ when $s_k = 1$, we only need to show that $(1-\pi p)^t\{1+t\pi p\}$ is strictly decreasing in t for $t \ge 0$. To show that $(1-\pi p)^t\{1+t\pi p\}$ is decreasing in t, we show that its first derivative is nonpositive. It is easy to show that the first derivative of $(1-\pi p)^t\{1+t\pi p\}$ is $(1-\pi p)^t\{\pi pt\ln(1-\pi p)+\ln(1-\pi p)+\pi p\}$. Then, we need to show that $a \ge -1/(\pi p) - 1/\ln(1-\pi p)$. But because $-1/\ln(1-\pi p)$ is nonnegative, $t \ge -1/(\pi p)$; in fact, $t \ge 0$. Thus, it is true that the function is nondecreasing, and $(1-\pi p)^{s_k-1}\{1+(s_k-1)\pi p\} \le 1$.

Appendix B. Covariance matrix of the MLEs of $(p,\ \pi)$

We approximate the covariance by inverse of Fisher's information matrix which is the expected value of the negative Hessian matrix over the truncated joint probability mass function in (2.2).

The Hessian matrix, $H(\tilde{r}, \tilde{a})$, is the 2×2 matrix of the second derivatives of the truncated joint probability mass function in (2.2), and

$$H(\tilde{r}, \tilde{a}) = \begin{pmatrix} H_{11} & B(p,\pi)+\pi p A(p,\pi) \\ B(p,\pi)+\pi p A(p,\pi) & H_{22} \end{pmatrix}, \tag{B.1}$$

where $A(p,\pi) = \sum_{k=1}^n [s_k(1-\pi p)^{s_k-2}\{(1-\pi p)^{s_k}+s_k-1\}]/\{1-(1-\pi p)^{s_k}\}^2$, $B(p,\pi) = -\sum_{k=1}^n s_k(1-\pi p)^{s_k-1}/[1-(1-\pi p)^{s_k}]$, $H_{11} = -rp^{-2}-(s-r)(1-p)^{-2}+\pi^2 A(p,\pi)$ and $H_{22} = -a\pi^{-2}-(r-a)(1-\pi)^{-2}+p^2 A(p,\pi)$.

The 2×2 information matrix is $I(p, \pi) = -\mathrm{E}\{H(\tilde{r}, \tilde{a} \mid p, \pi)\}$, where

$$I(p,\pi) = \begin{pmatrix} C(p,\pi) - \pi^2 A(p,\pi) & -B(p,\pi) - \pi p A(p,\pi) \\ -B(p,\pi) - \pi p A(p,\pi) & D(p,\pi) - p^2 A(p,\pi) \end{pmatrix}, \tag{B.2}$$

where $C(p,\pi) = \tilde{r}(p,\pi)p^{-2} + [s - \tilde{r}(p,\pi)](1-p)^{-2}$, $D(p,\pi) = \tilde{a}(p,\pi)\pi^{-2} + [\tilde{r}(p,\pi) - \tilde{a}(p,\pi)](1-\pi)^{-2}$ and $\tilde{r}(p,\pi) = \mathrm{E}(r \mid p, \pi)$ and $\tilde{a}(p,\pi) = \mathrm{E}(a \mid p, \pi)$ are given in (2.3) and (2.4) respectively.

Hence, letting $d = C(p,\pi) - \pi^2 A(p,\pi)$, $e = D(p,\pi) - p^2 A(p,\pi)$ and $f = -B(p,\pi) - \pi p A(p,\pi)$, the covariance matrix of $(\tilde{p}, \tilde{\pi})$, approximately $I(p,\pi)^{-1}$ with elements given in (B.2) evaluated at $(\tilde{p}, \tilde{\pi})$, is

$$I(\tilde{p},\tilde{\pi})^{-1} = \begin{pmatrix} \sigma_p^2 & \rho\sigma_p\sigma_\pi \\ \rho\sigma_p\sigma_\pi & \sigma_\pi^2 \end{pmatrix}, \tag{B.3}$$

where $\sigma_p^2 = \{d(1-\rho^2)\}^{-1}$, $\sigma_\pi^2 = \{e(1-\rho^2)\}^{-1}$ and $\rho = -f(de)^{-1/2}$. Note that in this approximation σ_p^2 is the variance of $\tilde{p}$, σ_π^2 is the variance of $\tilde{\pi}$, and ρ is the correlation between $\tilde{p}$ and $\tilde{\pi}$.

Finally, we show that the correlation ρ is nonnegative. It is easy to show that $-f = \sum_{k=1}^n A_k B_k$ where $A_k = \pi p(1-\pi p)s_k(1-\pi p)^{s_k}/\{1-(1-\pi p)^{s_k}\}$ and $B_k = s_k - \{1 + \pi p(1-\pi p)^{-1}\}\{1-(1-\pi p)^{s_k}\}$. Thus, we need the condition for B_k to be nonnegative for each k, and this is the same as $s_k - 1 \geq \{1-(1-\pi p)^{s_k}\}/\pi p(1-\pi p)$. This leads to the condition for nonnegativity that $s_k - 1 \geq 4\pi p$.

References

Bailey, N. T. J. (1951). The estimation of frequencies of recessives with incomplete multiple selection. *Annals of Eugenics* **16**, 215-222.

Chambers, R., Dorfman, A. and Wang, S. (1998). Limited information likelihood analysis of survey data. *Journal of the Royal Statistical Society, Series B* **60**, 397-411.

Crow, J. F. (1965). Epidemiology and genetics of chronic disease. In *Public Health Service Publication 1163* (edited by J. V. Neal, M. W. Shaw and W. J. Schull) Department of Health, Education, and welfare, Washington,DC pp. 23-44.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B* **39**, 1-38.

Fisher, R. A. (1934). The effect of methods of ascertainment upon the estimation of frequencies. *Annals of Eugenics* **6**, 13-25.

Lange, K. (2002). *Mathematical and Statistical Methods for Genetic Analysis*, 2ed. Springer-Verlag.

Malec, D., Davis, W. W. and Cao, X. (1999). Model-based small area estimates of overweight prevalence using sample selection adjustment. *Statistics in Medicine* **18**, 3189-3200.

Morton, N. E. (1959). Genetic tests under incomplete ascertainment. *American Journal of Human Genetics* **11**, 1-16.

Nandram, B. (2007). Bayesian predictive inference under informative sampling via surrogate samples. In *Bayesian Statistics and Its Applications* (Edited by S. K. Upadhyay, U. Singh and D. K. Dey). Anamaya, New Delhi: Anshan.

Nandram, B. and Choi, J. W. (2005). A bayesian analysis of a two way categorical table incorporating intra-class correlation. *Journal of Statistical Computation and Simulation* **76**, 233-249.

Nandram, B., Choi, J. W., Shen, G. and Burgos, C. (2006). Bayesian predictive inference under informative sampling and transformation. *Applied Stochastic Models in Business and Industry* **22**, 559-572.

Nelder, J. A. and Mead, R. (1965). A Simplex Method for Function Minimization. *Computer Journal* **7**, 308-313.

Nielsen, R. and Signorovitch, J. (2003). Correcting for ascertainment biases when analyzing snp data: Applications to the estimation of linkage disequilibrium. *Theoretical Population Biology* **63**, 245-255.

Patil, G. P. and Rao, C. R. (1978). Weighted distributions and size biased sampling with applications to wildlife populations and human families. *Biometrics* **34**, 179-189.

Pfeffermann, D. and Sverchkov, M. (2007). Small-area estimation under informative probability sampling of areas and within areas. *Journal of the American Statistical Association* **102**, 1427-1439.

Sham, P. (1998). *Statistics in Human Genetics.* Arnold.

Sverchkov, M. and Pfeffermann, D. (2004). Prediction of finite population totals based on the sample distribution. *Survey Methodology* **30**, 79-92.

Thompson, E. A. (1986). *Pedigree Analysis in Human Genetics.* Johns Hopkins.

Received March 19, 2009; accepted May 6, 2009.

Balgobin Nandram
Department of Mathematical Sciences
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA 01609, USA
balnan@wpi.edu

Jai-Won Choi
Department of Biostatistics
Medical College of Georgia
1469 Laney Walker Blvd.
Augusta, GA 30912, USA
jchoi@mcg.edu

Hongyan Xu
Department of Biostatistics
Medical College of Georgia
1469 Laney Walker Blvd.
Augusta, GA 30912, USA
hxu@mcg.edu