Estimating effective paternity number in social insects and the effective number of alleles in a population

RASMUS NIELSEN,* DAVID R. TARPY†‡ and H. KERN REEVE†
*Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14853, USA; †Department of Neurobiology and Behaviour, Seeley G. Mudd Hall, Cornell University, Ithaca, NY 14853, USA

Abstract

Estimating paternity and genetic relatedness is central to many empirical and theoretical studies of social insects. The two important measures of a queen's mating number are her actual number of mates and her effective number of mates. Estimating the effective number of mates is mathematically identical to the problem of estimating the effective number of alleles in population genetics, a common measure of genetic variability introduced by Kimura & Crow (1964). We derive a new bias-corrected estimator of effective number of types (mates or alleles) and compare this new method to previous methods for estimating true and effective numbers of types using Monte Carlo simulations. Our simulation results suggest that the examined estimators of the true number of types have very similar statistical properties, whereas the estimators of effective number of types have quite different statistical properties. Moreover, our new proposed estimator of effective number of types is approximately unbiased, and has considerably lower variance than the original estimator. Our new method will help researchers more accurately estimate intracolony genetic relatedness of social insects, which is an important measure in understanding their ecology and social behaviour. It should also be of use in population genetic studies in which the effective number of alleles is of interest.

Keywords: effective number of alleles, genotyping, number of species problem, paternity analysis, paternity number, polyandry, social insects

Received 14 May 2003; revision received 29 July 2003; accepted 28 August 2003

Introduction

Paternity analyses using molecular markers has enabled ecologists to better understand wide-ranging issues such as male mating success (Milligan & McMurry 1993; Nielsen *et al*. 2001), reproductive tactics (Rico *et al*. 1992; Neff 2001), reproductive skew (Nonacs 2000; Reeve *et al*. 2000), and population gene flow (Devlin & Ellstrand 1990; Konuma *et al*. 2000; Gonzalez Martinez *et al*. 2002). Paternity within colonies of social insects determines directly values of genetic relatedness among nestmates, which, in turn, influences a host of social interactions that are pertinent to their evolution and ecological success, such as optimal sex ratios (Trivers & Hare 1976; Queller 1993), division of labour (Beshers & Fewell 2001), worker policing (Ratnieks &

Visscher 1989; Ratnieks & Reeve 1992; Oldroyd 1999), and kin selection (Hamilton 1964; West-Eberhard 1975; Queller & Strassmann 1998). Although polyandry is relatively widespread across social insect taxa (Crozier & Pamilo 1996), polyandrous queens in most social insect species have low mating numbers (Strassmann 2001) and thus the distinction in paternity frequency is often between one vs. two fathers (Pedersen & Boomsma 1999). Extremely high mating frequencies are only commonly observed in highly eusocial taxa such as fungus-growing ants (*Acromyrmex* and *Atta*), honeybees (*Apis*), and yellow jackets (*Vespula*) (reviewed by Strassmann 2001).

When a female reproductive, or queen, mates with more than one male, her worker offspring are diverse genotypically. Thus her colony consists of genetically distinct subfamilies, the workers in each being sired from her respective mates, which can be detected using molecular markers. Several statistical techniques (e.g. Queller & Goodnight 1989) can be used to calculate the average colony genetic relatedness directly from such data, but

Correspondence: R. Neilsen. Fax: (607) 255–4698; E-mail: rn28@cornell.edu. ‡Present address: Department of Entomology, Campus Box 7613, North Carolina State University, Raleigh, NC 27495–7613, USA.

understanding relatedness alone provides little insight into the actual number of fathers or the skews in their paternity. Therefore, a common practice for researchers has been to estimate directly the number of mates of a social insect queen, the 'paternity number' (e.g. Estoup *et al.* 1994; Estoup *et al.* 1995; Boomsma & Van der Have 1998; Peters *et al.* 1999).

There are many measures of the number of mates in social insects, two of which are of particular interest in empirical studies (Boomsma & Ratnieks 1996). First, the 'actual paternity number' of a queen is the total number of males that sire her offspring (k). This measure is obviously important to better understand issues such as queen mating behaviour, male mating success, and reproductive systems. Second, the 'effective paternity number' of a queen (k_e) is her mating number if all males are represented equally within her offspring, and thus is a function of the number of different mates and their relative contributions within a queen's brood. A mathematical definition of effective paternity number is (Starr 1984)

$$k_e = 1 \bigg/ \sum_{i=1}^{k} p_i^2, \qquad (1)$$

where p_i is the proportion of offspring sired by male i. Effective paternity number is important to determine the identity by descent of workers; that is, the probability of sampling, at random, two individuals from a colony and having them share the same father. Identity by descent and differing degrees of relatedness have implications toward numerous nestmate interactions and the evolution of sociality by kin selection (Crozier & Pamilo 1996; Crozier *et al.* 1997; Ross 2001).

The definition of k_e in Equation 1 is similar to the definitions in the population genetic literature of the effective number of alleles proposed originally by Kimura & Crow (1964), with p_i now representing the frequency in the population of allele i. The effective number of alleles is often used as a measure of genetic diversity (e.g. Kuser *et al.* 1997; Lin & Cheng 1997; Schierup 1998). The problem of estimating effective paternity number and the effective number of alleles is mathematically identical, and we will therefore treat these two problems as a common problem of estimating the number of *types* (alleles or paternity number).

Our purpose here is to explore different statistical methods that have been used to estimate the number of types. Furthermore, we derive a bias correction to the most commonly used estimator of the effective number of types, which greatly improves its statistical properties.

Estimators of actual paternity number

Estimating the number of discrete 'types' or 'classes' in a population from a finite sample is a well known statistical problem, known as 'the number of species problem'

(reviewed by Bunge & Fitzpatrick 1993). One common approach, often used in studies of social insects, is to obtain an estimate of k ($\hat{k}_m$), by solving the following equation iteratively

$$k_0 = \hat{k}_m - \left[\hat{k}_m \left(1 - \frac{1}{\hat{k}_m} \right)^n \right] \qquad (2)$$

(Cornuet & Aries 1980; Oldroyd *et al.* 1997), where n is the number of samples and k_0 is the observed number of different types within the sample. This is a method of moments estimate of k based on the statistic k_0, assuming multinomial sampling from an infinite population and equal contribution of all types (i.e. $p_i = 1/k$ for all i).

Many other estimators of the number of species (types) have been presented in the statistical literature, and we will just mention a few of the more promising estimators here. The maximum likelihood estimator assuming equal contribution of all types, $\hat{k}_{mle}$, is approximately found as the solution to (Good 1950)

$$k_0 = \hat{k}_{mle}(1 - e^{n/\hat{k}_{mle}}) \qquad (3)$$

for which the variance is approximately (Lewontin & Prout 1956)

$$V[\hat{k}_{mle}] \approx \frac{\hat{k}_{mle}}{e^{n/\hat{k}_{mle}} + (1 + n/\hat{k}_{mle})}. \qquad (4)$$

$\hat{k}_{cov}$ (Good & Toulmin 1956; Darroch & Ratcliff 1980) is defined as

$$\hat{k}_{cov} = \frac{k_0}{\hat{u}_{GOOD}}, \qquad (5)$$

where $\hat{u}_{GOOD}$ is an estimator of the coverage defined by $\hat{u}_{GOOD} = 1 - (c_1/n)$ and c_1 is the number of types represented by a single individual (i.e. the number of singletons in the sample). Like $\hat{k}_m$, these estimators are based on the assumption of equal frequency of all types. If this assumption is not met, k will tend to be underestimated.

Many other estimators of the number of species have been proposed including the nonparametric approaches by Chao (1984) and Chao & Lee (1992), which might perform better when all types do not contribute equally, and a number of Bayesian and empirical Bayes methods (e.g. Lewins & Joanes 1984; Leite *et al.* 2000). However, we do not consider them here.

Estimators of effective number of types

A naïve estimator of the effective number of types might be

$$\hat{k}_{e1} = 1 \bigg/ \sum_{i=1}^{k_0} \hat{p}_i^2 \qquad (6)$$

where $\hat{p}_i^2$ is the observed sample frequency of type i (i.e. $\hat{p}_i = x_i/n$), and where x_i is the observed number of type i in the sample. Because x_i is binomially distributed with parameters n and $1/k$, we have

$$E[x_i] = n/k \quad \text{and} \quad E[x_i^2] = \frac{n(n+k-1)}{k^2} \tag{7}$$

so

$$E\left[\sum_{i=1}^{k_0} \hat{p}_i^2\right] = kE[x_i^2/n^2] = \frac{n+k-1}{nk}. \tag{8}$$

This suggests using the following estimator proposed by Pamilo (1993):

$$\hat{k}_{e2} = \frac{n-1}{n\sum_{i=1}^{k_0} \hat{p}_i^2 - 1} \tag{9}$$

This is not an unbiased estimator because it relies on a nonlinear transformation between k and $E[\Sigma \hat{p}_i^2]$. Tarpy & Nielsen (2002) use simulations to show that this estimator can be quite biased, particularly at small sample sizes. It would be helpful therefore to devise an estimator of the effective number of types (mates or alleles) that is approximately unbiased.

Theory and Methods

The aim of this section is to find an approximate expression for the expectation of Equation 9 in order to devise an appropriate bias correction. To do this we will use the method of propagation of error. Expanding $E[\hat{k}_{e2}]$ in a Taylor series about $\sum_{i=1}^{k_0} \hat{p}_i^2$, we get the following second order approximation:

$$E[\hat{k}_{e2}] \approx \frac{n-1}{nE\left[\sum_{i=1}^{k_0} \hat{p}_i^2\right] - 1} + \frac{n^2 V\left[\sum_{i=1}^{k_0} \hat{p}_i^2\right](n-1)}{\left(nE\left[\sum_{i=1}^{k_0} \hat{p}_i^2\right] - 1\right)^3} \tag{10}$$

Assuming $p_i = 1/k$ ($i = 1, 2, \ldots, k$), and after a bit of algebra, we find

$$V(\hat{p}_i^2) = \frac{(k-1)(6 + k^2 + 6k(n-1) - 10n + 4n^2)}{k^4 n^3}, \tag{11}$$

and

$$Cov(\hat{p}_i^2, \hat{p}_j^2) = -\frac{(6 + k^2 + 6k(n-1) - 10n + 4n^2)}{k^4 n^3}$$

$$= -\frac{V(\hat{p}_i^2)}{k-1}, \quad i \neq j, \tag{12}$$

so

$$V\left[\sum_{i=1}^{k_0} \hat{p}_i^2\right] = kV(\hat{p}_i^2) + 2\binom{k}{2} Cov(p_i^2, p_j^2), \quad i \neq j \tag{13}$$

$$= \frac{2(n-1)(k-1)}{k^2 n^3}$$

Inserting Equations 8 and 13 into Equation 10, we find

$$E[\hat{k}_{e2}] \approx \frac{k(2k + (n-2)(n+1))}{n(n-1)}, \tag{14}$$

so the bias for this estimator is approximately $\frac{2k(k-1)}{n(n-1)}$. Similarly, we find

$$V[\hat{k}_{e2}] \approx \frac{n^2 V\left[\sum_{i=1}^{k_0} \hat{p}_i^2\right](n-1)^2}{\left(nE\left[\sum_{i=1}^{k_0} \hat{p}_i^2\right] - 1\right)^4} = \frac{2(k-1)k^2}{n(n-1)}. \tag{15}$$

This suggests the following bias corrected estimator

$$\hat{k}_{e3} \approx \frac{\hat{k}_{e2}n(n-1)}{(2\hat{k}_{e2} + (n-2)(n+1))}$$

$$= \frac{(n-1)^2}{\sum_{i=1}^{k_0} \hat{p}_i^2 (n+1)(n-2) + 3 - n}, \tag{16}$$

which has approximate expectation

$$E[\hat{k}_{e3}] \approx \frac{(n-1)^2}{E\left[\sum_{i=1}^{k_0} \hat{p}_i^2\right](n+1)(n-2) + 3 - n}$$

$$+ \frac{(n+1)^2(n-2)^2 V\left[\sum_{i=1}^{k_0} \hat{p}_i^2\right](n-1)^2}{\left(E\left[\sum_{i=1}^{k_0} \hat{p}_i^2\right](n+1)(n-2) + 3 - n\right)^3} \tag{17}$$

$$= k + \frac{8(k-1)k^3}{(2k + (n+1)(n-2))^3}$$

$$- \frac{8(k-1)k^2}{(2k + (n+1)(n-2))^2}.$$

To the degree that these approximations are accurate, the bias of the new estimator is $O(1/n^4)$ whereas the bias of $\hat{k}_{e2}$ is $O(1/n^2)$. The variance of $\hat{k}_{e3}$ is approximately

$$V[\hat{k}_{e3}] \approx \frac{(n+1)^2(n-2)^2 V\left[\sum_{i=1}^{k_0} \hat{p}_i^2\right](n-1)^4}{\left(E\left[\sum_{i=1}^{k_0} \hat{p}_i^2\right](n+1)(n-2)+3-n\right)^4} \quad (18)$$

$$= \frac{2k^2(k-1)(n-2)^2(n-1)n(n+1)^2}{(2k+(n-2)(n+1))^4}.$$

For $1 < k \le n$, $V[\hat{k}_{e3}] < V[\hat{k}_{e2}]$; the variance of $\hat{k}_{e3}$ is always lower than the variance of $\hat{k}_{e2}$. Clearly, under the assumptions made here, $\hat{k}_{e3}$ is a better estimator of k_e than $\hat{k}_{e2}$. For completeness, we can also find the approximate variance of the naïve estimator as

$$V[\hat{k}_{e1}] \approx \frac{V\left[\sum_{i=1}^{k_0} \hat{p}_i^2\right]}{E\left[\sum_{i=1}^{k_0} \hat{p}_i^2\right]^4} = \frac{2(n-1)(k-1)k^2n}{(n+k-1)^4}. \quad (19)$$

These approximations will, in general, be accurate if the denominators of the estimators (e.g. $n\sum_{i=1}^{k_0} \hat{p}_i^2 - 1$ for $\hat{k}_{e2}$) are much larger than zero. When $k \ge n$, this will not be true, particularly not for $\hat{k}_{e2}$ which is undefined when $k_0 = n$.

Simulations

To verify the results derived in the previous section, and to further examine the statistical properties of the proposed estimators, we performed Monte Carlo simulations similar to those by Tarpy & Nielsen (2002) by sampling from a multinomial distribution with parameters $n, p_1, p_2, \ldots, p_k$. Initially, we assumed $p_i = k^{-1}, i = 1, 2, \ldots, k$, corresponding to assuming equal allele frequencies or equal probability of paternity. 1000 iterations were performed for $k = 5$ and 20 each with $n = 3-50$. For each iteration, we estimated the number of types (k) using $\hat{k}_{mle}$, $\hat{k}_{cov}$, and $\hat{k}_m$, as well as the effective number of types (k_e) by $\hat{k}_{e1}$, $\hat{k}_{e2}$ and $\hat{k}_{e3}$. The mean and variance of an estimator in the 1000 replicates were recorded for each estimator.

When $p_i = k^{-1}$ ($i = 1, 2, \ldots, k$), then k equals k_e. Otherwise, the expected effective number of types will be less than the actual number of types. To determine the effect of unequal values of p_i on estimates of k and k_e, we modified the above simulations to allow unequal values of the p_i's. We used three different simulation schemes, designated by 'low skew', 'moderate skew', or 'high skew'. The three distributions were designed to mimic empirically observed distributions of paternity numbers from the social insects
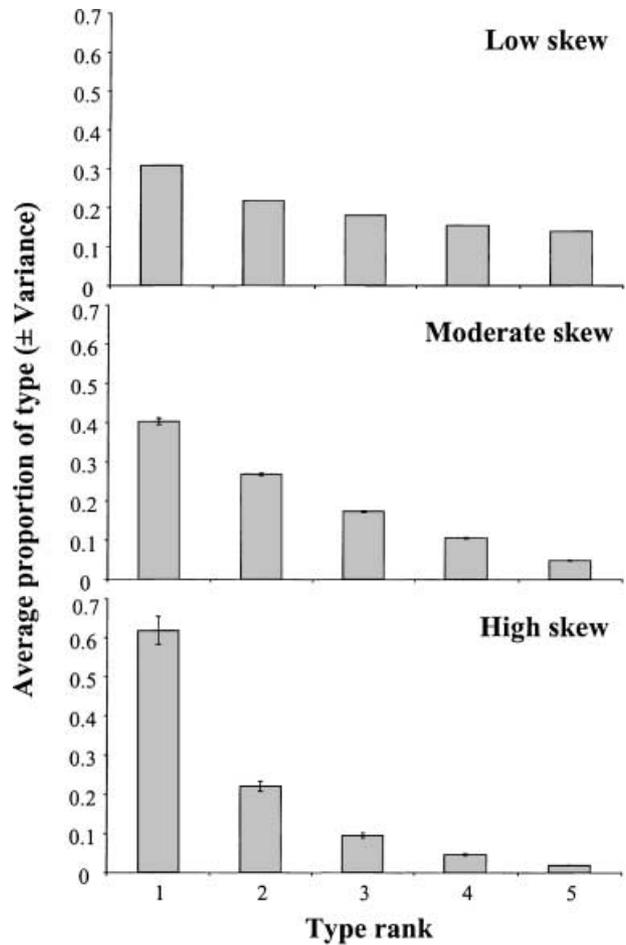

Fig. 1 The rank ordered distributions of p_i for differing levels of skew Error bars are provided for moderate and high skew plots because the distributions were generated using simulations to mimic the variability observed among real data sets.

literature (see Tarpy & Nielsen 2002). Each of these distributions are shown for $k = 5$ in Fig. 1.

Results

Figure 2 illustrates how sample size affects the various estimators of k when $p_i = 1/k, i = 1, 2, \ldots, k$. At low sample sizes, the average estimates tend to be higher than the true value, but all rapidly approach the true value as sample size increases. Moreover, the average estimates produced by all three statistics are very similar to each other, with $\hat{k}_{mle}$ tending to give slightly higher estimates than $\hat{k}_{cov}$, and $\hat{k}_m$ tending to give slightly lower estimates than $\hat{k}_{cov}$. Regardless, each of these estimators is far less biased than the observed number of types in the sample (k_0), particularly at high values of k.

The different estimators of k also have similar variances, and the mean square error (MSE) is therefore also similar
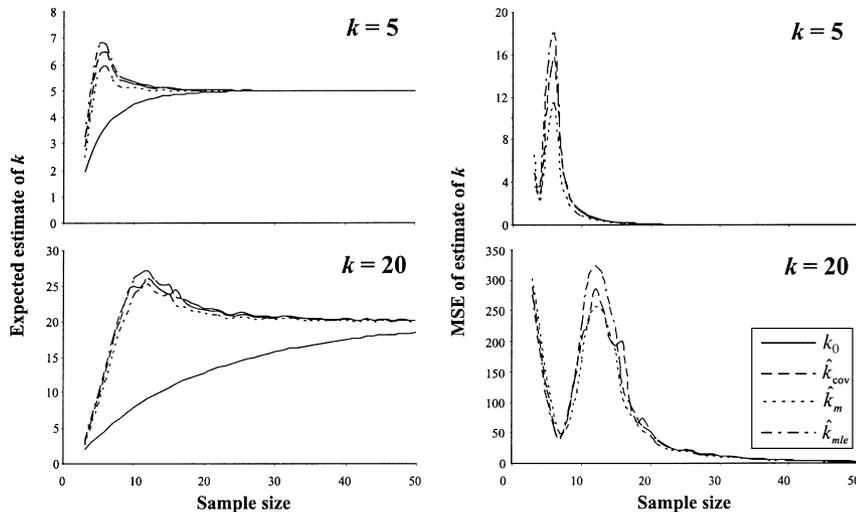
Fig. 2 Effect of n and k on the expected estimate and mean squared error (MSE) of k. $p_i = 1/k$, $i = 1, 2, \ldots, k$ was assumed.
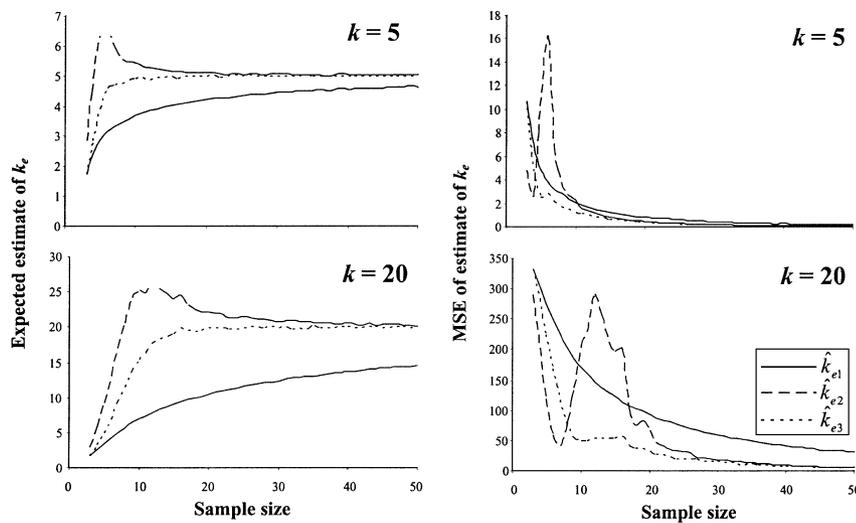

Fig. 3 Effect of n and k on the expected estimates and mean squared error (MSE) of k_e. $p_i = 1/k$, $i = 1, 2, \ldots, k$ was assumed.

for the three estimators (Fig. 2). However, the variance of $\hat{k}_m$ is slightly smaller than the variance of $\hat{k}_{cov}$, which again tends to be slightly smaller than the variance of $\hat{k}_{mle}$. Consequently, $\hat{k}_m$ has the lowest MSE for most parameter values.

Figure 3 compares the various estimators of k_e as a function of k and n. $\hat{k}_{e1}$ is strongly biased towards smaller values and its expectation converges very slowly to the true value. Both $\hat{k}_{e2}$ and $\hat{k}_{e3}$ are far less biased and converge much faster. $\hat{k}_{e3}$, in particular, tends to converge faster than both $\hat{k}_{e1}$ and $\hat{k}_{e2}$. $\hat{k}_{e1}$ has a very low variance (not shown) but it has, nonetheless, a very large MSE because of the strong bias.

The variance of $\hat{k}_{e3}$ is consistently lower than the variance of $\hat{k}_{e2}$. As expected from the analytical results, $\hat{k}_{e3}$ dominates $\hat{k}_{e2}$ as an estimator of k_e because it is less biased and has a lower variance, except when n is very small. The Mean Squared Error (MSE) of $\hat{k}_{e3}$ is lower than the MSE of any of the two other estimators for any value of k explored,

except when n is very small. For very small values of n, $\hat{k}_{e2}$ may have a lower MSE than $\hat{k}_{e3}$.

The approximate variances (Equations 15, 18 and 19) tend to be quite accurate for values of $n > k$, but are not reliable when $k < n$ (not shown). It might also be worth noting that when $p_i = 1/k$ ($i = 1, 2, \ldots, k$), $\hat{k}_{e3}$ has a much smaller MSE than either $\hat{k}_{cov}$, $\hat{k}_m$ or $\hat{k}_{mle}$. So if equal contribution of all types is expected, then $\hat{k}_{e3}$ is the preferred estimator of k as well as k_e.

The effect of unequal type probabilities (skew) has a significant influence on the estimators of k (Fig. 4). With a strong skew, all estimators tend to be considerably biased towards smaller values. Thus the more the number of types is skewed, the more that the value of k is underestimated.

The variances of the estimators of k are all similarly affected by skew resulting in similar MSEs for all estimators in the presence of a strong skew (Fig. 4). At low sample sizes, the variance is reduced in the presence of a
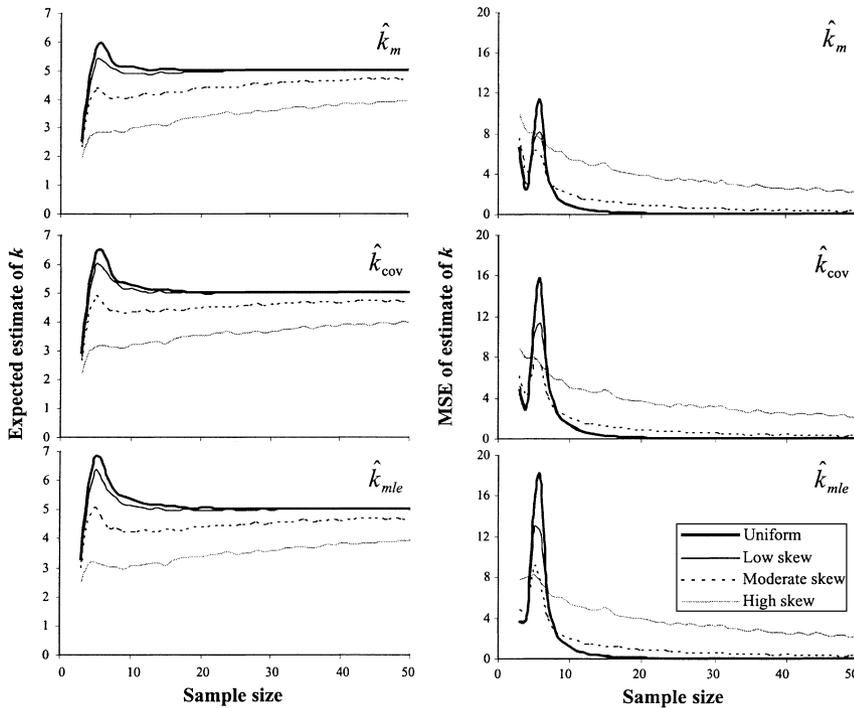
Fig. 4 Effect of a skew in the distribution of the type probabilities on the expected estimates and mean squared error (MSE) of k for $k = 5$.
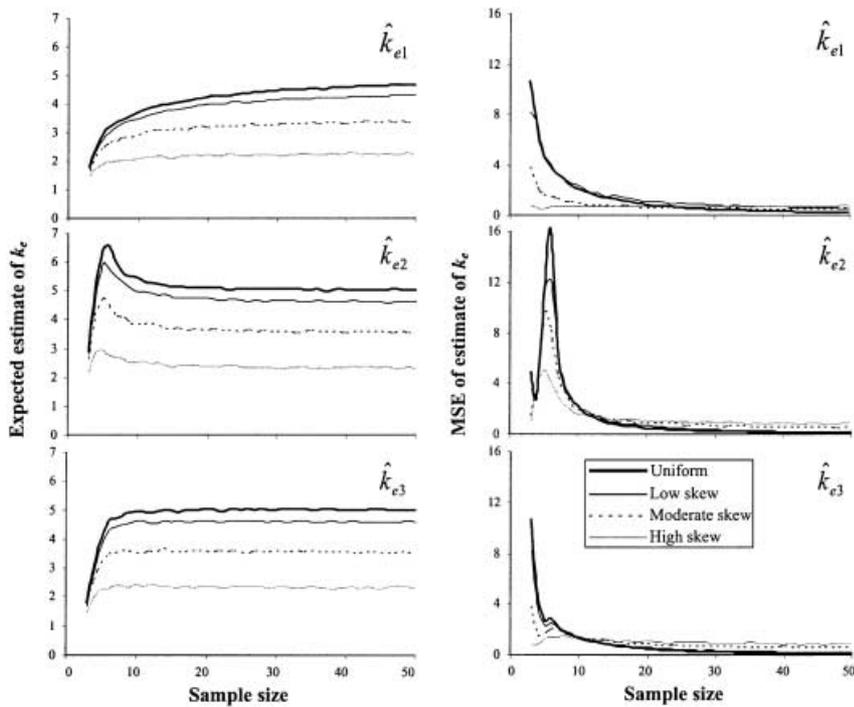

Fig. 5 Effect of skew in the distribution of the type probabilities on the expected estimates and mean squared error (MSE) of of k_e. The true values of k_e are uniform: $k_e = 5.00$, low skew: $k_e = 4.57$, moderate skew: $k_e = 3.59$, and high skew: $k_e = 2.26$.

skew because of the bias in the estimate towards smaller values. However, for high values of n, a skew increases the variance in the estimate of k. In general, the MSE for $\hat{k}_m$ tends to be lower than the MSE for any of the other estimators, but in the presence of a strong skew the performance of the three estimators is almost identical.

The estimators of k_e are less influenced by unequal type probabilities (Fig. 5). They all seem to perform relatively well, with $\hat{k}_{e3}$ converging faster to the true value than $\hat{k}_{e2}$, which again converges faster to the true value than $\hat{k}_{e1}$. However, for a strong skew, the bias of $\hat{k}_{e2}$ is of a similar magnitude as the bias of $\hat{k}_{e1}$. Note that the

expected values are lower because of the skew; that is, $k_e < k$ because of unequal frequencies of the different types. None of the estimators seem to be significantly more robust to violations of the model assumption of equal type probabilities than other of the estimators. In fact, in the presence of a high skew, all the estimators have very similar MSEs, except $\hat{k}_{e2}$ which still has a relatively large MSE when $n \approx k$.

Discussion

Our results demonstrate that the three estimators of k that we compare provide very similar estimates of the number of types in a population. $\hat{k}_m$ has a slightly smaller MSE than the other estimators and might be preferred on this basis when the assumption of $p_i = 1/k$, $i = 1, 2, \dots , k$, may not hold. Otherwise, $\hat{k}_{e3}$ should be used as an estimator of k because of its lower MSE in this special case.

On the contrary, the estimators of effective paternity frequency have very different statistical properties. $\hat{k}_{e3}$ performs in general better than any of the other estimators of k_e. However, when $n < k$, $\hat{k}_{e1}$ and $\hat{k}_{e2}$ may perform better than $\hat{k}_{e3}$ depending on the assumptions regarding the true distribution of type probabilities. In the presence of a strong skew in the type probabilities, all the examined estimators have similar MSEs. For real data analysis one should use $\hat{k}_{e3}$ for all estimates of k_e, except possibly when $k_0 \approx n$, in which case none of the estimators provide very reliable estimates.

For many populations of social insects, the important distinction in paternity frequency is between one and two mates (Pedersen & Boomsma 1999; Strassmann 2001). In such instances, the proposed correction to the sample statistic would provide only a minimal benefit to estimates of effective paternity. However, in cases where: (a) small worker sample sizes are used; or (b) when more extreme mating frequencies occur, the correction may help provide estimates that are significantly more accurate. A more precise estimate of paternity within insect societies will enable researchers to better test many theoretical concepts and improve our understanding of social insect biology.

The effective number of alleles has been used as a measure of genetic variability (e.g. Zouros 1979). Previous estimators have been biased, resulting in underestimation of the effective number of alleles in small samples. It is clearly desirable that measures of genetic variability are not too sensitive to the sample size. The new estimator ($\hat{k}_{e3}$) is very close to being unbiased, even at relatively small sample sizes, and should therefore, be of use in population genetic studies in which genetic variability is summarized in terms of the effective number of alleles.

References

Beshers SN, Fewell JH (2001) Models of division of labor in social insects. *Annual Review of Entomology*, **46**, 413–440.

Boomsma JJ, Ratnieks FL (1996) Paternity in eusocial Hymenoptera. *Philosophical Transactions of the Royal Society of London, B*, **351**, 947–975.

Boomsma JJ, Van der Have TM (1998) Queen mating and paternity variation in the ant *Lasius niger*. *Molecular Ecology*, **7**, 1709–1718.

Bunge J, Fitzpatrick M (1993) Estimating the Number of Species — a Review. *Journal of the American Statistical Association*, **88**, 364–373.

Chao A (1984) Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics, Theory and Applications*, **11**, 265–270.

Chao A, Lee S-M (1992) Estimating the number of classes via sample coverage. *Journal of the American Statistical Association*, **87**, 210–217.

Cornuet J-M, Aries F (1980) Number of sex alleles in a sample of honeybee colonies. *Apidologie*, **11**, 87–93.

Crozier RH, Oldroyd BP, Tay WT *et al.* (1997) Molecular advances in understanding social insect population structure. *Electrophoresis*, **18**, 1672–1675.

Crozier RH, Pamilo P (1996) *Evolution of Social Insect Colonies: Sex Allocation and Kin Selection*. Oxford University Press, New York, NY.

Darroch JN, Ratcliff D (1980) A note on capture-recapture estimation. *Biometrics*, **36**, 149–153.

Devlin B, Ellstrand NC (1990) The development and application of a refined method for estimating gene flow from angiosperm paternity analysis. *Evolution*, **44**, 248–259.

Estoup A, Scholl A, Pouvreau A, Solignac M (1995) Monoandry and polyandry in bumble bees (Hymenoptera; Bombinae) as evidenced by highly variable microsatellites. *Molecular Ecology*, **4**, 89–93.

Estoup A, Solignac M, Cornuet J (1994) Precise assessment of the number of patrilines and of genetic relatedness in honeybee colonies. *Proceedings of the Royal Society of London, B*, **258**, 1–7.

Gonzalez Martinez SC, Gerber S, Cervera MT *et al.* (2002) Seed gene flow and fine-scale structure in a Mediterranean pine (*Pinus pinaster* Ait.) using nuclear microsatellite markers. *Theoretical and Applied Genetics*, **104**, 1290–1297.

Good IJ (1950) *Probability and the Weighting of Evidence*. Charles Griffin, London.

Good IJ, Toulmin GH (1956) The number of new species and the increase in population coverage when a sample is increased. *Biometrika*, **43**, 45–63.

Hamilton WD (1964) The genetical evolution of social behaviour. I and II. *Journal of Theoretical Biology*, **7**, 1–52.

Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. *Genetics*, **49**, 725–738.

Konuma A, Tsumura Y, Lee CT, Lee SL, Okuda T (2000) Estimation of gene flow in the tropical-rainforest tree *Neobalanocarpus*

heimii (Dipterocarpaceae), inferred from paternity analysis. *Molecular Ecology*, **9**, 1843–1852.

Kuser JE, Meagher TR, Sheely DL, White A (1997) Allozyme frequencies in New Jersey and North Carolina populations of Atlantic white-cedar, *Chamaecyparis thyoides* (Cupressaceae). *American Journal of Botany*, **84**, 1536.

Leite JG, Rodrigues J, Milan LA (2000) A Bayesian analysis for estimating the number of species in a population using non-homogeneous Poisson process. *Statistics and Probabilities Letters*, **48**, 153–161.

Lewins WA, Joanes DN (1984) Bayesian estimation of the number of species. *Biometrics*, **40**, 323–328.

Lewontin RC, Prout T (1956) Estimation of the Number of Different Classes in a Population. *Biometrics*, **12**, 211–223.

Lin T-P, Cheng Y-P (1997) Allozyme variation in four geographic areas of *Cinnamomum kanehirae*. *Journal of Heredity*, **88**, 433–438.

Milligan BG, McMurry CK (1993) Dominarnt vs. codominant genetic markers in the estimation of male mating success. *Molecular Ecology*, **2**, 275–283.

Neff BD (2001) Genetic paternity analysis and breeding success in bluegill sunfish (Lepomis macrochiros). *Journal of Heredity*, **92**, 111–119.

Nielsen R, Mattila DK, Clapham PJ, Palsboll PJ (2001) Statistical approaches to paternity analysis in natural populations and applications to the North Atlantic humpback whale. *Genetics*, **157**, 1673–1682.

Nonacs P (2000) Measuring and using skew in the study of social behavior and evolution. *American Naturalist*, **156**, 577–589.

Oldroyd BP, Clifton MJ, Wongsiri W *et al.* (1997) Polyandry in the genus Apis, particularly *Apis andreniformis*. *Behavioral Ecology and Sociobiology*, **40**, 17–26.

Oldroyd BP, Osborne KE (1999) The evolution of worker sterility in honeybees: The genetic basis of failure of worker policing. *Proceedings of the Royal Society Biological Sciences Series B*, **266**, 1335–1339.

Pamilo P (1993) Polyandry and allele frequency differences between the sexes in the ant *Formica aquilonia*. *Heredity*, **70**, 472–480.

Pedersen JS, Boomsma JJ (1999) Multiple paternity in social Hymenoptera: Estimating the effective mate number in single-double mating populations. *Molecular Ecology*, **8**, 577–587.

Peters JM, Queller DC, Imperatriz Fonseca VL, Roubik DW, Strassmann JE (1999) Mate number, kin selection and social conflicts in stingless bees and honeybees. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **266**, 379–384.

Queller DC (1993) Worker control of sex ratios and selection for extreme multiple mating by queens. *American Naturalist*, **142**, 346–351.

Queller DC, Goodnight KF (1989) Estimating relatedness using genetic markers. *Evolution*, **43**, 258–275.

Queller DC, Strassmann JE (1998) Kin selection and social insects. *Bioscience*, **48**, 165–175.

Ratnieks FLW, Reeve HK (1992) Conflict in single-queen hymenopteran societies the structure of conflict and processes that reduce conflict in advanced eusocial species. *Journal of Theoretical Biology*, **158**, 33–65.

Ratnieks FLW, Visscher PK (1989) Worker policing in the honey bee. *Nature*, **342**, 796–797.

Reeve HK, Starks PT, Peters JM, Nonacs P (2000) Genetic support for the evolutionary theory of reproductive transactions in social wasps. *Proceedings of the Royal Society Biological Sciences Series B*, **267**, 75–79.

Rico C, Kuhnlein U, Fitzgerald GJ (1992) Male reproductive tactics in the threespine stickleback — an evaluation by DNA fingerprinting. *Molecular Ecology*, **1**, 79–87.

Ross KG (2001) Molecular ecology of social behaviour: analyses of breeding systems and genetic structure. *Molecular Ecology*, **10**, 265–284.

Schierup MK (1998) The number of self-incompatibility alleles in a finite, subdivided population. *Genetics*, **149**, 1153–1162.

Starr CK (1984) Sperm competition, kinship, and sociality in the Aculeate Hymenoptera. In: *Sperm Competition and the Evolution of Animal Mating Systems*. (ed. Smith RL), pp. 427–464. Academic Press, New York.

Strassmann JE (2001) The rarity of multiple mating by females in the social Hymenoptera. *Insectes Sociaux*, **48**, 1–13.

Tarpy DR, Nielsen DI (2002) Sampling error, effective paternity, and estimating the genetic structure of honey bee colonies (Hymenoptera: Apidae). *Annals of the Entomological Society of America*, **95**, 513–528.

Trivers RL, Hare H (1976) Haplodiploidy and the evolution of the social insects. *Science*, **191**, 249–263.

West-Eberhard MJ (1975) The evolution of social behavior by kin selection. *Quarterly Review of Biology*, **50**, 1–33.

Zouros E (1979) Mutation rates, population sizes, and amounts of electrophoretic variation of enzyme loci in natural populations. *Genetics*, **92**, 623–646.