
Article:
Wilkinson, R.D. orcid.org/0000-0001-7729-7023, Kapranas, A. and Hardy, I.C.W. (2016) Detecting non-binomial sex allocation when developmental mortality operates. Journal of Theoretical Biology, 408. pp. 167-178. ISSN 0022-5193

https://doi.org/10.1016/j.jtbi.2016.08.008

Detecting non-binomial sex allocation when developmental mortality operates

RUNNING TITLE: Detecting non-binomiality

Richard D. Wilkinson[a], Apostolos Kapranas[b,c], Ian C.W. Hardy[b]

[a]*Corresponding author: School of Mathematics and Statistics, University of Sheffield, Sheffield, S3 7RH, UK. r.d.wilkinson@sheffield.ac.uk*

[b]*School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough, LE12 5RD, UK*

[c]*Laboratory for Fundamental and Applied Research in Chemical Ecology, University of Neuchtel, Neuchtel, 2000 Switzerland*

Abstract

Optimal sex allocation theory is one of the most intricately developed areas of evolutionary ecology. Under a range of conditions, particularly under population sub-division, selection favours sex being allocated to offspring non-randomly, generating non-binomial variances of offspring group sex ratios. Detecting non-binomial sex allocation is complicated by stochastic developmental mortality, as offspring sex can often only be identified on maturity with the sex of non-maturing offspring remaining unknown. We show that current approaches for detecting non-binomiality have limited ability to detect non-binomial sex allocation when developmental mortality has occurred. We present a new procedure using an explicit model of sex allocation and mortality and develop a Bayesian model selection approach (available as an R package). We use the double and multiplicative binomial distributions to model over- and under-dispersed sex allocation and show how to calculate Bayes factors for comparing these alternative models to the null hypothesis of binomial sex allocation. The ability to detect non-binomial sex allocation is greatly increased, particularly in cases where mortality is common. The use of Bayesian methods allows for the quantification of the evidence in favour of each hypothesis, and our modelling approach provides an improved descriptive capability over existing approaches. We use a simulation study to demonstrate substantial improvements in power for detecting non-binomial sex allocation in

situations where current methods fail, and we illustrate the approach in real scenarios using empirically obtained datasets on the sexual composition of groups of gregarious parasitoid wasps.

Key words: Sex ratio; under-dispersion; Bayes factor; Markov chain Monte Carlo

1. Introduction

The null model of sex allocation theory is the Düshing-Fisher theory of equal investment (West, 2009). When populations are both large and have unbiased sex ratios, selection for variance in the sexual composition of offspring groups is predicted to be absent (Kolman, 1960). Under these conditions mothers will not be selectively penalized if they randomly allocate sex to offspring, with fixed probability of $p = 0.5$ that the offspring is male, independently of the sex of previous offspring. Thus, the number of males in each offspring group would have binomial variance, i.e., $np(1-p)$, where n is the number of offspring. In smaller populations and under sex ratio bias ($p \neq 0.5$), stabilizing selection for low sex ratio variance is predicted, i.e., variance less than $np(1-p)$ (Verner, 1965; West, 2009). Selection on sex ratio variance is likely to be strong when populations are structured into small reproductive subgroups within which offspring mate with each other on maturity and prior to the dispersal of the daughters (local mate competition; Hamilton, 1967); here, selection favours the evolution of low sex ratio variance, especially when one or a very few mothers contribute offspring to the locally mating group (Green et al., 1982; Hardy, 1992; Nagelkerke and Hardy, 1994; Nagelkerke, 1996; West and Herre, 1998). This is because low variance maximizes the production of mated daughters, a close correlate of maternal fitness. If one male is sufficient to mate successfully with all females within a group and all offspring in the group are progeny of one mother, then the optimal sexual composition is one male and the remainder of the group being females (Green et al., 1982). Similar arguments predict low variance under local resource competition (a generalization of local mate competition) and its converse, local resource enhancement (Lambin, 1994). Variance in the number of males among groups lower than expected under binomial sex allocation is known as under-dispersion, and sex allocation is then termed *precise* (Green et al., 1982; Lambin, 1994; Nagelkerke, 1996).

Control of sex allocation can be detected in some organisms by direct observation of sexually differential aspects of individual offspring production, such as maternal movements during egg laying, or the placement of offspring, or by non-random production sequences (Cole, 1981; Hardy, 1992; Heinsohn et al., 1997; Krackow et al., 2002; Khidr et al., 2013; Ambrosini et al., 2014) but such evidence is not often available. Empiricists must more frequently rely on the statistical analysis of offspring group sex ratios to detect whether sex allocation is being controlled or whether it is, for instance, binomial, as might be the null-expectation under several chromosomal mechanisms of sex-determination (Avilés et al., 2000; Krackow et al., 2002; Ewen et al., 2003; Macdonald and Johnson, 2008; Postma et al., 2011). Furthermore, empirical evaluations of sex ratio variance can provide tests of explicit predictions of sex ratio theory (e.g., Lambin, 1994; Morgan and Cook, 1994; Hardy and Cook, 1995; Hardy et al., 1998; Nagelkerke and Sabelis, 1998; West and Herre, 1998; Kapranas et al., 2011; Khidr et al., 2013; Bowers et al., 2013).

One practical problem often faced by investigations of sex ratios and sex ratio variance is that information on the sexual compositions of offspring is available at maturity but not at the time of sex allocation, and it is not uncommon for some offspring to die before maturity, (e.g., Hardy et al., 1998; Dyrcz et al., 2004; Ewen et al., 2004; Forsyth et al., 2004; Dietrich-Bischoff et al., 2006; Øigarden and Lifjeld, 2013). Provided it has a stochastic component, developmental mortality will act to increase the variance of observed sex ratios, making initially under-dispersed data appear closer to binomial. This effect is expected on logical grounds (Section 3) and has been shown empirically both within and across several species of organisms with group structured mating (Hardy et al. 1998; Kapranas et al. 2011; Khidr et al. 2013; see also Dyrcz et al. 2004 and Dietrich-Bischoff et al. (2006)). Current statistical approaches to assessing sex ratio variance (Krackow et al., 2002) are, however, based on the implicit assumption that developmental mortality does not operate, and they consequently lack power to detect non-binomiality, unless mortality rates are low.

Our aim is to show that by introducing a model that represents the biological processes that generated the data (sex allocation followed by mortality), we can substantially improve our ability to detect underlying biological behaviours. We also demonstrate the advantage of using more descriptive statistical approaches such as estimating effect sizes (with measures

of confidence), rather than relying on null-hypothesis significance testing, where the small dataset sizes mean we often fail to clear an arbitrary significance hurdle (usually $\alpha = 0.05$) even when the data indicate phenomena of interest. We begin by evaluating the performance, under developmental mortality, of the statistical methods commonly used to detect non-binomial sex ratio variance. We find that the power of these methods is adversely affected by developmental mortality. We then develop an alternative approach that explicitly models the mortality process. This has much improved power for detecting non-binomial sex allocation, particularly when there is high mortality or datasets are small.

2. Terms and notation

We define some terms and notation before describing current approaches and their limitations, and then introduce our new approach for detecting non-binomial sex allocation. A summary of the notation is provided in Table 1. The methods developed are general, but are likely to most readily be applied to egg-laying organisms such as birds, parasitoid wasps, fig wasps and phytoseiid mites (Hardy, 1992; Nagelkerke and Sabelis, 1998; West and Herre, 1998; West, 2009; Bowers et al., 2013), and this is reflected in the terminology we adopt (for a mammalian example see Macdonald and Johnson, 2008). Assume that we have a dataset containing data on C different clutches of eggs, all of which were laid in comparable environmental conditions. Offspring group size is called *clutch size* at the time of production (egg-laying) and *brood size* at the time of offspring maturity: brood size is less than clutch size when developmental mortality occurs.

A *primary* dataset consists of counts of the number of eggs and their sex for each clutch. Let N_i denote the number of eggs laid in the ith clutch, and M_i be the number of those N_i eggs that are male. A primary dataset is the collection $\{(N_i, M_i)\}_{i=1}^{C}$. However, for most empirical investigations M_i is not observed, as the sex of an offspring cannot be easily determined from the eggs: it is usual to wait until the eggs hatch and develop to the point at which offspring sex can be discriminated (e.g., Dietrich-Bischoff et al., 2006; Khidr et al., 2013). It is also usual that a proportion of the eggs fail to mature, due to some form of developmental mortality, and consequently their sex cannot be recorded.

A *secondary* dataset consists of counts of n_i, the number of offspring that reach maturity

(brood size) and m_i, the number of those offspring that are male, with the complete secondary dataset denoted $\{(n_i, m_i)\}_{i=1}^{C}$. Although a small number of experiments have been conducted where primary datasets are obtained, either directly from genetic characteristics of eggs (Dijkstra, 1986; Hardy et al., 1998; Nagelkerke and Sabelis, 1998; Khidr et al., 2013) or through selective statistical procedures (Dyrcz et al., 2004; Kapranas et al., 2011), the vast majority of analyses have been conducted using secondary datasets (e.g., Hardy, 1992; West and Herre, 1998; Nagelkerke and Sabelis, 1998; Mackauer and Völkl, 2002; Dietrich-Bischoff et al., 2006; Kapranas et al., 2008).

Our null hypothesis about sex allocation, H_0, is that there is a sex ratio p (the proportion of offspring that are male), and that each egg is male with probability p independently of all other eggs in the clutch, i.e., that the distribution of sex ratios is binomial

$$M_i \sim \text{Bin}(N_i, p). \tag{1}$$

The alternative hypothesis, H_1, is that the number of males is non-binomially distributed, that is, either over- or under-dispersed when compared to the binomial distribution. Note that these are hypotheses about primary sex ratios, not secondary sex ratios.

3. Current approaches for detecting non-binomial sex allocation

Several methods have been used for the statistical analysis of sex ratio variances (James, 1975; Green et al., 1982; Nagelkerke and Sabelis, 1998; West and Herre, 1998; Krackow et al., 2002). Whilst these methods can work well when applied to primary sex ratio data, this is not usually available, and so these methods are instead applied to secondary data, effectively treating them as if they were primary data. Not considering or ignoring that mortality has occurred thus violates the assumptions behind each approach; this results in a lack of statistical power, often leading to incorrect conclusions.

The first method for detecting departures from the binomial distribution, is a formal statistical test derived by E. Meelis (Nagelkerke and Sabelis, 1998), which we refer to as the *Meelis test* (Krackow et al., 2002). The test is a comparison of the estimated variance with the variance under the assumption of a binomial distribution, and is derived by calculating the distribution (under the null hypothesis) of $\sum m_i^2$ conditional on $\sum m_i$. A test statistic

Symbol	Definition
C	Number of clutches in the dataset
N	Number of eggs laid (primary)
M	Number of eggs laid that are male (primary)
n	Number of offspring that reach maturity (secondary)
m	Number of males that reach maturity (secondary)
D	The complete observed dataset, i.e., $D = \{(n_i, m_i)\}_{i=1}^{C}$
p	Sex ratio† (proportion of eggs that are male)
ψ	Dispersion parameter
λ	Average clutch size
d	Mortality rate
$H_0,\ H_1$	Null and alternative hypotheses
U	Test statistic for the Meelis' test
R	Descriptive ratio contrasting observed and expected variance
s^2	McCullagh's dispersion estimator
$\mathcal{S}$	Clutch sizes observed in the data, i.e., $\{k : n_j = k \text{ for some } j\}$
v_k	Number of clutches of size k, i.e., $\sum_{i=1}^{C} \mathbb{I}_{n_i=k}$
s_k^2	Empirical variance of the number of males in clutches of size k
B_{01}	Bayes factor for comparing H_0 with H_1

Table 1: Summary of notation used in this article. Letters in **bold** font indicate vector quantities, indices (e.g., n_i) indicate an instance of that variable, and hats (e.g., $\hat{p}$) indicate estimates. †Care needs to be taken with interpretation of p in the multiplicative binomial model as p is no longer the expected sex ratio when $\psi \neq 0$.

U (see supplementary material for details) is defined which can be shown to have a standard normal distribution under H_0, provided C is sufficiently large. Large negative values of U indicate under-dispersion, and large positive values over-dispersion; typically, the test is applied by calculating the p-value $\mathbb{P}(|U| > |u_{obs}|)$, where $\mathbb{P}$ denotes probability, with small values taken to indicate departure from the null hypothesis.

There are several difficulties with applying the Meelis test to the datasets used in empirical studies of sex-allocation. Firstly, the test assumes that the binomial random variables are observed directly, which is not the case when using secondary data (using m_i instead of M_i). Secondly, the test is derived for use on random variables from a binomial distribution with fixed size ($n_i = n$ for all i), whereas for real data, the values of n_i vary between broods, with datasets typically consisting of a range of brood sizes. It is common practice to collect all the broods of a certain size (e.g., all m_i such that $n_i = j$), then calculate the U-statistic, denoted U_j for those broods, before combining them using

$$U = \frac{\sum U_j}{\sqrt{|\mathcal{S}|}}$$

to give a single statistic U, where $\mathcal{S} = \{k : n_j = k \text{ for some } j\}$ is the collection of clutch sizes observed in the dataset. If each $U_j \sim N(0,1)$, then $U \sim N(0,1)$. However, the Meelis test was derived for large sample sizes. In practice, there may only be a small number of clutches with $n_i = j$, and so each U_j may not be well approximated by a standard normal distribution and hence, U may not have a $N(0,1)$ distribution either.

James' test (James, 1975) is an alternative to the Meelis test that is often used for analysing datasets containing small clutches of unequal sizes. It involves calculation of a test statistic (Krackow et al., 2002, give details), which is known to be approximately normally distributed under the assumption of binomial sex ratios (no mortality). Large positive values indicate over-dispersion, and negative values under-dispersion. It is known to be less powerful for a single clutch size than the Meelis test (and suffers from the same difficulties as the Meelis test), but is included in our analysis for completeness.

The descriptive ratio R is also used:

$$R = \frac{\sum_{k\in\mathcal{S}} v_k s_k^2}{\sum_{k\in\mathcal{S}} v_k k\hat{p}_k(1-\hat{p}_k)}$$

where s_k^2 is the empirical variance of the number of males in clutches of size k, i.e., $s_k^2 = \mathbb{V}ar(\{m_i : n_i = k\})$, and $v_k = \sum_{i=1}^{C} \mathbb{I}_{n_i=k}$ is the number of clutches which have size k. The denominator is the sum of the variances if assuming a binomial distribution, where $\hat{p}_k$ is the estimated sex ratio for clutches of size k, i.e.,

$$\hat{p}_k = \frac{1}{kv_k} \sum_{i=1}^{C} m_i \mathbb{I}_{n_i=k}.$$

The rationale for using R, is that it is the observed variance of the number of males, divided by the variance that would occur if the number of males was binomially distributed (Krackow et al., 2002). We expect to observe $R \approx 1$ if the data are binomially distributed, with $R < 1$ for under-dispersed data. McCullagh and Nelder (1989) introduce a further estimator of dispersion, which is a sum of ratios rather than a ratio of sums

$$s^2 = \frac{1}{C-1} \sum_{i=1}^{C} \frac{(m_i - \hat{p}n_i)^2}{n_i \hat{p}(1-\hat{p})} \quad \text{where} \quad \hat{p} = \frac{\sum m_i}{\sum n_i},$$

and should be interpreted in the same way as the R statistic.

The effect of mortality is to make the data appear less under-dispersed (more binomial), as mortality has the effect of increasing the variance of the number of males. To see this, imagine a species which has perfect precision, with each mother laying the same number of male and female eggs every time, so that the sex ratio variance is zero. Stochastic mortality would introduce an element of randomness to the sexual composition of the offspring groups, such that secondary datasets may even resemble binomial random variables under sufficiently high rates of mortality (see Section 5.3).

3.1. Evaluation of current approaches when developmental mortality occurs

To illustrate the limitations of current approaches, we simulate synthetic under-dispersed primary datasets, and then simulate the mortality process to produce synthetic secondary datasets. By applying the approaches described above, and repeating the process numerous times, we can examine their performance under varying levels of mortality.

We simulated sample experimental datasets as follows: for $i = 1, \ldots, C$,

1. Simulate the clutch size from a Poisson distribution: $N_i \sim Po(\lambda)$, where λ is the average clutch size.

2. Simulate the number of males in the i^{th} clutch, M_i, from an under-dispersed multiplicative binomial distribution (Section 4).

3. Simulate the secondary dataset by assuming each of the N_i eggs has probability d of not reaching maturity, and count the number of females and males that survive.

We used a primary dataset on the parasitoid wasp *Goniozus legneri* (Khidr et al., 2013), a species known to produce a strongly under-dispersed primary sex ratio, to estimate parameter values for the synthetic data model, and used these estimates fixed throughout the simulation study ($\lambda = 10.0, p = 0.00278$, and $\psi = 0.445$, where p and ψ are parameters in the multiplicative binomial distribution, which is an under-dispersed distribution - see Section 4.1). We varied the size of the simulated experiment C, and the mortality rate d, and for each pair of values we simulated 10,000 synthetic datasets, and averaged the test statistics found across the replicates. This allows the effectiveness of all the procedures to be examined across a range of dataset sizes, C, and mortality rates d.

The performance of a hypothesis test can be measured by its power for a given significance level, where power is the probability of detecting non-binomial sex allocation when it occurs (i.e., power $= 1 - \mathbb{P}(\text{Type II error}) = \mathbb{P}(\text{reject } H_0 \mid H_1)$). Contour plots of the power of the Meelis and James tests (at significance level 0.05) as a function of the number of clutches in the dataset and the mortality rate show that the test lacks power if the number of clutches used is small or if the mortality rate is moderate-to-large (Fig. 1a,b). For example, for a dataset containing 50 clutches with a mortality rate of 10% there is only a 35% probability of correctly detecting under-dispersion. The power of the test is lower still if lesser degrees of under-dispersion are assumed as it becomes harder to detect (we used reasonably large under-dispersion of $\psi = 0.445$).

Fig. 1c shows the effect of mortality on R. The expected value of R increases towards 1 as the mortality rate increases, so that species with a high mortality rate will have R values consistent with binomial sex allocation, even if their primary sex ratios are under-dispersed. Fig. 1d shows the same information for McCullagh's s^2. This can be seen to be less affected by mortality and so its use should be preferred to R. The number of clutches in the experiment has only a minor effect on the expected value of both statistics. However,

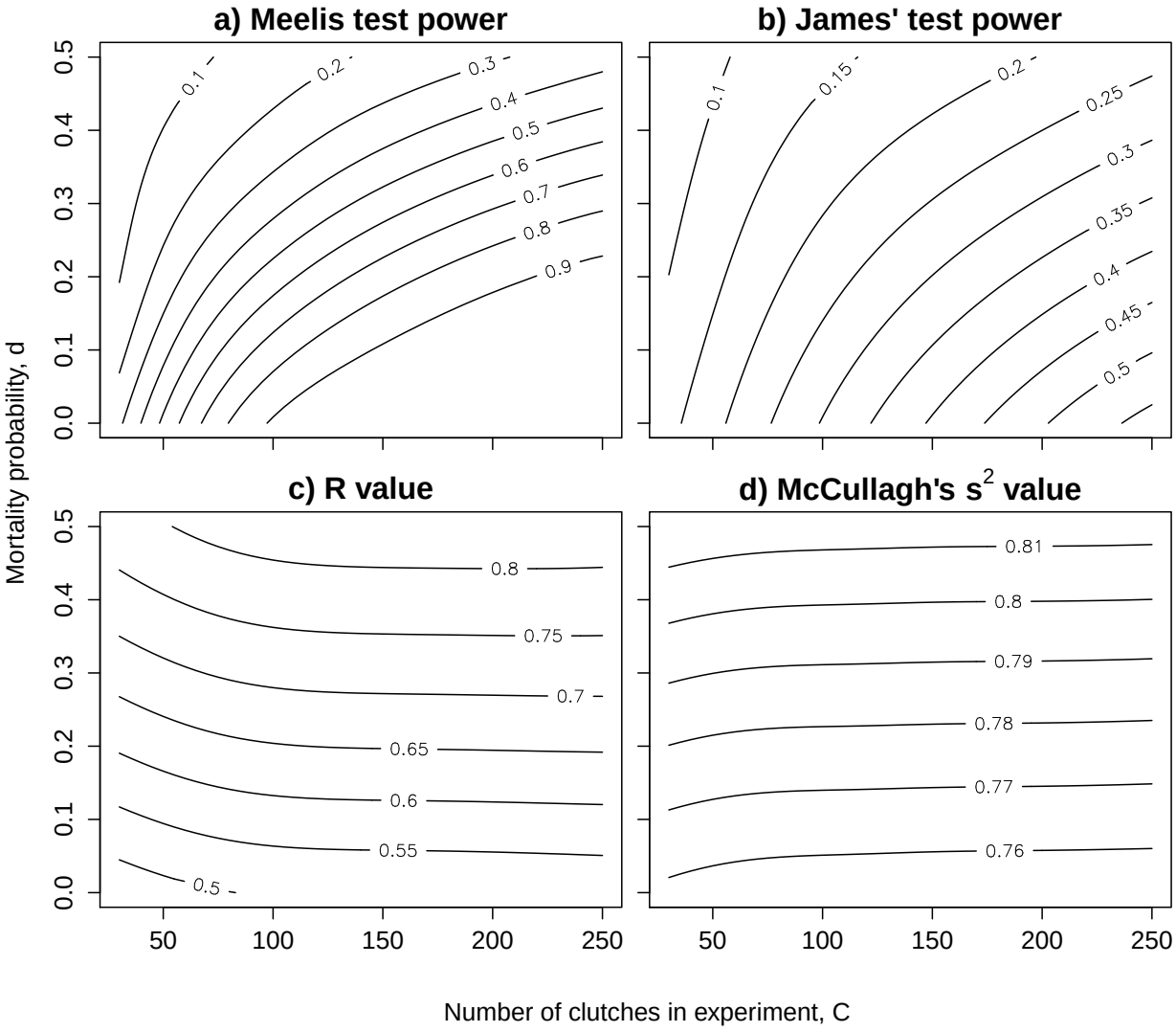


Figure 1: a) and b) show contour plots of the power of the two-sided Meelis and James tests; c) and d) are contour plots of the values of descriptive statistics R and McCullagh's s^2, all as a function of C and d. The values were estimated using 10,000 randomly generated datasets, using parameter values estimated from data on *G. legneri* primary sex ratios.

it strongly affects the variance of the estimate (not shown), and for smaller experiments the observed values can vary widely, and so without appropriate confidence intervals for both statistics, they have little value.

There are two (related) reasons for the lack of power in these approaches. The first is that mortality increases the variance of the secondary values (n_i, m_i) compared to the primary values (N_i, M_i) making under-dispersion harder to detect. The second is that the tests do not take into account the fact that mortality has occurred, and consequently the additional variance is incorrectly interpreted as being consistent with binomial sex ratios.

4. A new test for detecting non-binomial sex allocation

By explicitly modelling mortality we develop a test that has improved statistical power as well as an increased descriptive capability. Our null hypothesis is a binomial model of sex allocation, which we compare to two different generalisations of the binomial distribution, the multiplicative binomial and the double binomial distributions, both of which can model over- and under-dispersion. Our model for the data then consists of a mortality model applied to the output of the sex allocation model. We use Bayesian model selection to determine which model is best supported by the data. The more intricate computational details are given in the supplementary material; here we focus on the broad outline of the approach.

4.1. A model of secondary data

We assume we have data on C different broods from comparable environmental conditions, so that they can be considered to be statistically exchangeable. Note that the unobserved primary counts N_i and M_i, and the corresponding secondary values after mortality has occurred, n_i and m_i, must satisfy the inequalities

$$N_i \geq n_i, \quad M_i \geq m_i \text{ and } N_i - n_i \geq M_i - m_i. \tag{2}$$

We consider three models for the data, which differ only in the distribution of the sex allocation, i.e., the distribution of M_i given N_i. The first is the binomial model, with $M_i|N_i, p \sim \text{Bin}(N_i, p)$, which corresponds to the null hypothesis in Section 2. The second is

the multiplicative binomial distribution introduced by Altham (1978):

$$\mathbb{P}(M|N,p,\psi) = c(p,\psi)\binom{N}{M}p^M(1-p)^{N-M}\mathrm{e}^{\psi M(N-M)}, \tag{3}$$

where $c(p,\psi)$ is an intractable normalising constant. The two parameters are a probability p, and a dispersion parameter ψ. The third, introduced by Efron (1986), is the double binomial model

$$\mathbb{P}(M|N,p,\psi) = c(p,\psi)\binom{N}{M}\frac{N^{N\psi}p^{M(\psi+1)}(1-p)^{(N-M)(\psi+1)}}{M^{M\psi}(N-M)^{(N-M)\psi}} \tag{4}$$

where $c(p,\psi)$ is again an intractable normalising constant. Note that when $\psi = 0$, both the multiplicative and double binomial distributions reduce to the binomial distribution. These models are the key part of our procedure, corresponding to the alternative hypothesis in Section 2, as they both model the three cases of interest:

(i) binomial sex allocation when $\psi = 0$

(ii) under-dispersed sex allocation when $\psi > 0$

(iii) over-dispersed sex allocation when $\psi < 0$.

Unfortunately neither of these two distributions arises from a simple physical mechanism. Familarity does allow an intuition to develop about the meaning of ψ, but our usage here does not require any interpretation beyond that given above, and that larger values of ψ indicate more under-dispersion than small values etc. Care also needs to be taken with interpretation of p, as the expected value of M is no longer Np for the multiplicative binomial distribution, except when $\psi = 0$, and so p can no longer be considered to be the sex ratio (the expected sex ratio, $\mathbb{E}\left(\frac{M}{N}\right)$, can be determined by Monte Carlo integration). We include both models as alternatives, as different datasets fit different models better, and this makes the detection of under-dispersion more likely.

We use the same model of mortality in each hypothesis and assume that each egg has probability d of dying before maturity, and thus of not being counted in the secondary dataset, independently of its sex and the other eggs in the clutch, i.e., we assume mortality is binomially distributed:

$$n_i|N_i,d \sim \mathrm{Bin}(N_i,d). \tag{5}$$

The distribution of m_i can then be shown, by a label permuting argument, to be

$$\mathbb{P}(m|M,N,n,d) = \frac{\binom{M}{M-m}\binom{N-M}{N-n-M+m}}{\binom{N}{N-m}}. \tag{6}$$

We use two complimentary approaches for detecting departures from binomial sex allocation, the first based on estimation of effect size, and the second on hypothesis testing (Nakagawa and Cuthill, 2007). The simpler approach is to estimate the effect size, measured by the dispersion parameter ψ, by finding its posterior distribution $\pi(\psi|D)$. This parameter indicates whether sex allocation is binomial, over-, or under- dispersed, as well as how strong the effect is. Posterior credibility intervals for ψ can be used to assess the precision of the estimates and indicate informally whether the data are consistent with H_0 ($\psi = 0$). We describe methodology to do this below, the code is provided in the `precision` R package, and applications are described in Section 5.

While various authors recommend estimation over hypothesis testing (Robert, 2001; Gelman et al., 2003; Nakagawa and Cuthill, 2007), relying solely on estimation of ψ does not always provide the clarity required. For example, if the posterior distribution contains some support for $\psi = 0$, but the posterior mode is not close to 0, it can be difficult to judge whether or not data are under-dispersed using only the posterior distributions (Section 5.2). Instead, we wish to obtain the probability that sex allocation is under-dispersed, i.e., the posterior probability that each of the three models M_0, M_1 and M_2 are true conditional upon the data: $\mathbb{P}(M_0|D)$, $\mathbb{P}(M_1|D)$, and $\mathbb{P}(M_2|D)$. These probabilities only make sense in a Bayesian setting, although note that p-values obtained from classical hypothesis tests, such as the Meelis test, are often incorrectly interpreted in this way (Goodman, 2008).

Bayesian model selection requires calculation of the Bayes factor (Jeffreys, 1939; Kass and Raftery, 1995), which is defined as the ratio of the evidence for two different hypotheses (or models)

$$B_{01} = \frac{\pi(D|H_1)}{\pi(D|H_0)}. \tag{7}$$

Values of B_{01} greater than 1 indicate evidence in favour of H_1 (over H_0) and values less than 1 indicate evidence for H_0 (over H_1). Jeffreys (1939) suggested interpretation of the strength of evidence in favour of a hypothesis according to the magnitude of the Bayes factor is shown

B_{01} range	$\mathbb{P}(H_1\|D)$ range	Interpretation
1–3	0.5-0.75	Barely worth mentioning
3–10	0.75 - 0.91	Substantial
10–30	0.91-0.97	Strong
30–100	0.97- 0.99	Very strong
> 100	0.99-1	Decisive

Table 2: Jeffreys' suggested interpretation of the Bayes factor for strength of evidence in favour of H_1 over H_0. Values of $1/B_{01} = B_{10}$ give the strength of evidence for H_0 over H_1. Also shown are the corresponding ranges of the posterior probability for H_1 given the data, in the case where we assign equal prior probability to both hypotheses.

in Table 2. The Bayes factor $B_{ij} = \mathbb{P}(D|H_j)/\mathbb{P}(D|H_i)$ can also be interpreted by noting that it is the ratio between the posterior and prior odds in favour of H_j over H_i

$$\frac{\mathbb{P}(H_j|D)}{\mathbb{P}(H_i|D)} = B_{ij}\frac{\pi_j}{\pi_i}$$

where π_j is the prior probability of H_j. Table 2 contains the posterior probabilities of H_1 being true for various Bayes factor ranges when we assume the hypotheses are equally likely *a priori*.

Bayes factors provide a powerful alternative to frequentist hypothesis tests, and have several advantages over classical methods. The first is that they provide a way to evaluate the evidence in favour of a hypothesis, in contrast to the classical approach which only rejects or accepts the null hypothesis for a particular error rate. This is particularly useful in datasets where the effect size or the sample size are small, or where the mortality rate is high, as we can quantify the strength of the evidence for under-dispersion in the data, even if there is not enough evidence to formally reject the null hypothesis. For instance, for analysis of data on *Goniozus thailandensis* (Section 5.2), the Meelis test finds $p > 0.05$ and thus concludes that there is no evidence to reject the null hypothesis, whereas the Bayesian approach reports that the posterior probability of the double binomial model being the true model is 0.79, with the probability of the binomial model being true only 0.14. When combined with the posterior distribution of ψ, which is concentrated on values greater than 0, this strongly suggests that this species produces under-dispersed sex ratios, a message that is lost if we only report the decision from the Meelis test.

4.2. Parameter estimation

We now describe how to find the posterior distribution of the parameters $\theta = (\psi, p, d, \lambda)$ given the data $D = \{(n_i, m_i)\}_{i=1}^{C}$, which we denote $\pi(\theta|D)$. This distribution represents our beliefs about the parameters after training the model to take the observed experimental data into account. The posterior distribution cannot be found analytically, and so we use Markov Chain Monte Carlo (MCMC) methods (e.g., Gilks et al., 1996) to obtain an approximation. We describe the case where only the n and m values, the number of eggs that reached maturity, have been recorded. The simpler situation where N_i is observed is a special case and follows immediately.

We introduce prior distributions for all unknowns. We assume the number of eggs laid in each clutch follows a Poisson distribution with mean λ

$$N_i \sim Po(\lambda) \text{ for } i = 1, \ldots, C \tag{8}$$

and for the fixed parameters we assume that

$$\begin{aligned} p &\sim U[0,1] & \psi &\sim N(0, \sigma^2) \\ \lambda &\sim \Gamma(a, b) & d &\sim \text{Beta}(a', b'). \end{aligned} \tag{9}$$

The distribution of p, λ and d are conjugate to the likelihood, allowing a Gibbs sampler to be used. Informative prior distributions are usually available for λ and d, as scientists often have information about mortality rates and average clutch sizes for the species of interest, although simulation suggests that the quantities of interest (the Bayes factor and the posterior of ψ), are robust to the choice of priors for λ and d. The key parameter is the dispersion parameter ψ, which we assign a zero mean normal distribution, so that under- and over-dispersion are equally likely *a priori*. We use an uninformative prior distribution for p, so that the posterior distribution is determined solely by the data.

To sample from the posterior distribution $\pi(\theta|D)$, we use a Metropolis-Hastings within Gibbs sampler (Metropolis et al., 1953; Geman and Geman, 1984). We introduce vectors of unobserved N_i and M_i values, denoted $\mathbf{N}$ and $\mathbf{M}$, as auxiliary variables, and sample across the chain $\pi(\theta, \mathbf{N}, \mathbf{M}|D)$, which is a $(4 + 2C)$ dimensional Markov chain. The distribution of interest, $\pi(\psi|D)$, is then found by taking the marginal distribution of ψ. Details of the

MCMC algorithm used are provided in the supplementary material, and the algorithm is implemented in the accompanying `precision` R package for each of the three models.

4.3. Bayes factor estimation

To calculate the Bayes factors (Equation 7) we must first calculate the evidence for each model

$$\pi(\mathbf{n}, \mathbf{m}) = \int \pi(\mathbf{n}, \mathbf{m}|\theta)\pi(\theta)\mathrm{d}\theta,$$

where $\mathbf{n}$ and $\mathbf{m}$ are the vectors of the observed n_i and m_i values, which is analytically intractable for the models considered. We use the approach described in Chib (1995) and Chib and Jeliazkov (2001) to estimate the evidence for each model, which relies upon the identity

$$\pi(\mathbf{n}, \mathbf{m}) = \frac{\pi(\mathbf{n}, \mathbf{m}|\theta^*)\pi(\theta^*)}{\pi(\theta^*|\mathbf{n}, \mathbf{m})}. \tag{10}$$

Calculation of both the numerator and denominator is challenging, but can be done with additional samples from an MCMC sampler. The derivation and details of the algorithm are technical, and are presented in the supplementary material. An implementation of these algorithms is available as the `precision` R package, available on github. The next section demonstrates the power of our approach.

5. Results

We illustrate our approach using data on four species of wasp: The strength of evidence for under-dispersion from secondary sex ratio data in these species varies from weak (*Colpoclypeus florus*) to overwhelming (*Metapycus luteolus*), and the mortality rate varies from low (*Goniozus legneri*) to high (*C. florus*). We also present the results from a simulation study which conclusively demonstrates the increased power of our approach.

The Bayesian approach requires prior distributions for all unknown parameters. Simulation studies have shown that the model and data are strongly informative about p and ψ, so that any information in the prior distribution is overwhelmed by the information in the data. In all our analyses we give p an uninformative prior distribution uniform on $[0, 1]$ and ψ a vague prior distribution for both the double and multiplicative binomial models:

$$p \sim U[0, 1], \qquad \psi \sim N(0, 1). \tag{11}$$

The prior for ψ can be justified by examining the degree of under-dispersion for various levels of ψ. If $M \sim \text{DoubleBinom}(n = 10, p = 0.1, \psi)$, then $\mathbb{P}(M = 1) = 0.38$ if $\psi = 0$ (the binomial case), whereas for $\psi = 3$, $\mathbb{P}(M = 1) = 0.85$, indicating strong under-dispersion. The Bayes factors are robust to the choice of priors for λ, d and p (the parameters shared across models), but unsurprisingly, are sensitive to the prior for ψ. More diffuse priors for ψ tend to reduce the evidence for under-dispersion due to an Occam's razor type effect, but for realistic priors for ψ, the conclusion does not usually change significantly (see supplementary material). Fortunately, the posterior distribution for ψ is robust to the choice of prior for ψ, and so this can also be used to indicate whether the data are under-dispersed.

The data typically contain only limited information about the parameters λ and d, but with the two posterior distributions strongly correlated, as large average clutch size and high mortality, or small average clutch size and lower mortality rate, leads to similar datasets. Prior information about λ and d is often available, which we can use to choose prior distributions for these two parameters on a species by species basis. Experimentation has shown that the Bayes factor and the posterior distribution of p and ψ (the primary parameter of interest) are robust to these choices.

5.1. Goniozus legneri: Large dataset, low mortality

We begin by considering data on *G. legneri*, a gregarious parasitoid wasp in which offspring groups are produced by single mothers and sex ratios are female biased due to local mate competition. Khidr et al. (2013) provide both a primary dataset, consisting of pre-mortality counts on 47 clutches obtained using DNA microsatellite markers to identify the sex of eggs, and a secondary dataset containing post-mortality counts of male and female adults in 113 clutches. Both the Meelis and James tests lead to rejection of the null hypothesis of binomial sex allocation (Table 3) with p-values of 0.0041 and 0.0027 respectively for the secondary data. Furthermore, we find $R = 0.572$, which when combined with the negative value of U in the two tests ($U = -2.38$ and $U = -1.98$ for Meelis and James respectively), lead us to conclude, in common with previous studies (Hardy et al., 1998; Khidr et al., 2013), that *G. legneri* has under-dispersed sex ratios.

Khidr et al. (2013) reported that the proportion of offspring that died before maturity was 7.6%, which agrees with previous *G. legneri* mortality estimates (5-12%, Hardy et al., 1998).

		Species						
		G. legneri		*G. thailandensis*	*C. florus*		*M. luteolus*	
Proceedure	Instance	Primary Value	Secondary Value	Secondary Value	Primary Value	Secondary Value	Primary Value	Secondary Value
James	U	-1.98	-3.00	-2.01	-0.89	2.7	-6.7	-7.8
	p	0.047	0.0027	0.045	0.37	0.0068	1.5×10^{-11}	7.0×10^{-15}
Meelis	U	-2.38	-2.87	-0.73	-3.24	-0.97	-7.9	-7.4
	p	0.017	0.0041	0.46	0.0012	0.33	2.6×10^{-15}	1.3×10^{-13}
R		0.44	0.57	0.68	0.13	0.75	0.093	0.44
s^2		0.57	0.61	0.74	0.51	1.18	0.20	0.58
BF	double:binomial	45.1	213.6	5.65	3830	0.27	1.1×10^{29}	9.8×10^{23}
	multiplicative:binomial	9430	31.3	0.54	0.36	0.36	7.0×10^{5}	2.0×10^{6}
	double:multiplicative	0.0048	6.8	10.5	10600	0.74	1.6×10^{23}	5.0×10^{17}
Posterior probability	binomial	0.00010	0.004	0.14	0.00026	0.61	0.000	0.000
	multiplicative	0.995	0.127	0.074	0.000094	0.22	0.000	0.000
	double	0.0048	0.869	0.74	0.9996	0.16	1.000	1.000

Table 3: Analysis of four wasp datasets: *G. legneri* primary ($C = 47$) and secondary ($C = 113$) datasets (Khidr et al., 2013); *G. thailandensis* secondary dataset ($C = 60$) (Witethom and Gordh, 1994); *C. florus* primary ($C = 55$) and secondary datasets ($C = 53$) (Dijkstra, 1986; Hardy et al., 1998); *M. luteolus* primary ($C = 127$) and secondary ($C = 371$) datasets (Kapranas et al., 2011). All values estimated using 10^6 MCMC iterations.

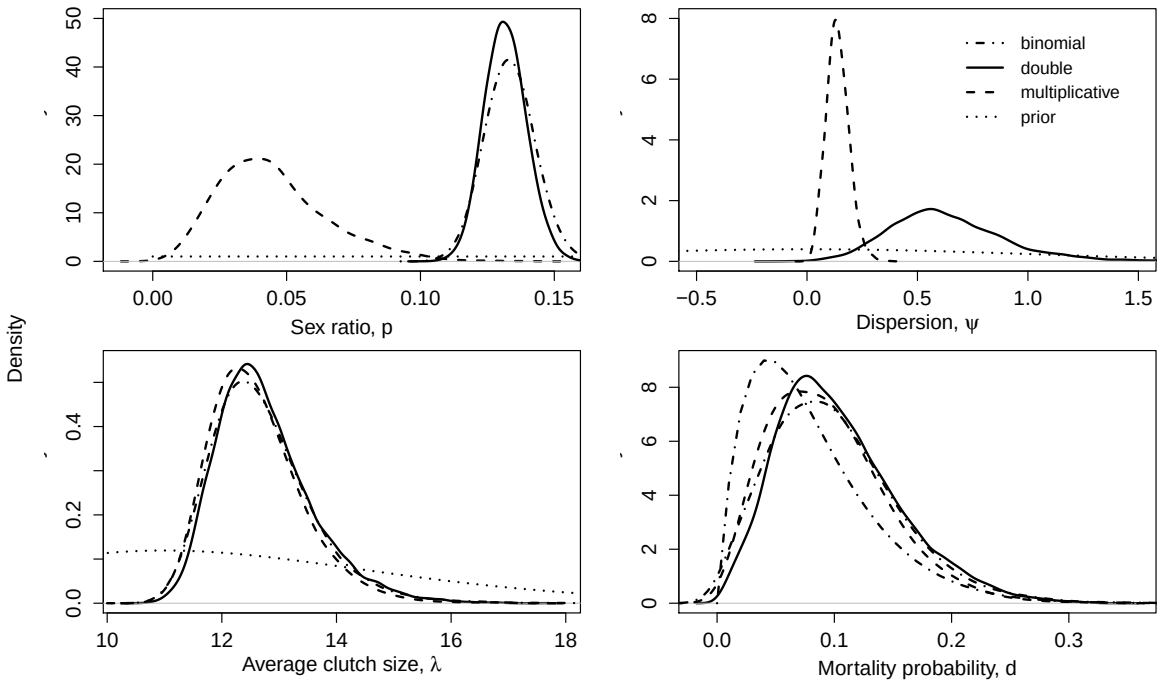


Figure 2: Posterior distributions from the analysis of *G. legneri* secondary data (Khidr et al., 2013), obtained using 5×10^5 MCMC iterations. For each parameter, the prior and posterior distribution are shown for the three alternative models of sex allocation. Note that the binomial model does not have a dispersion parameter (ψ) and that the interpretation of p and ψ is different in each model.

We incorporate this information into the analysis through the use of prior distributions

$$d \sim \text{Beta}(2, 23) \qquad \lambda \sim \text{Gamma}(12, 1).$$

The prior mean for d is thus $2/(23+2) = 8\%$, with values in the range 0-20% all supported *a priori* (Figure 2). The prior for λ was based on an observed secondary clutch size of 11, and the mortality rate of 7.6%, suggesting a prior mean for λ of approximately 12. The Gamma$(12, 1)$ prior distribution has a prior mean of 12/1, and supports prior λ values in a range between 11 and 14 (Figure 2).

Figure 2 shows the posterior distributions of the four parameters for the secondary dataset. Interest lies primarily in the dispersion parameter ψ, with $\psi > 0$ indicating under-dispersion and $\psi < 0$ over-dispersion. We cannot estimate ψ precisely as there is a finite quantity of data, but the posterior distributions show the range of ψ values we believe could feasibly have led to the observed data. The posterior distribution for ψ for both the double

and multiplicative models, suggests that only positive values of ψ are consistent with the data. Equi-tailed 95% credibility intervals for ψ are $[0.047, 0.248]$ for the multiplicative binomial model, and $[0.196, 1.28]$ for the double binomial model, neither of which overlap with 0, leading us to conclude that *G. legneri* has under-dispersed sex allocation.

The Bayes factor (BF) estimates for *G. legneri* are reported in Table 3. We find that the double binomial model is best supported, with a BF of 213.6 in favour of the double binomial over the binomial model, which Jeffreys' scale interprets as decisive evidence. There is also very strong evidence in favour of the multiplicative model over the binomial ($BF = 31.3$), and substantial evidence to suggest the double binomial is better supported than the multiplicative binomial model ($BF = 6.8$). If we are prepared to assign all three models equal prior probability, then the posterior probability that the binomial model is the true model is 0.004, compared to 0.869 for the double binomial model, and 0.127 for the multiplicative binomial model.

For this dataset, the signal from the data is strong ($C = 113$ is a reasonably large sample size), and consequently all the procedures give unambiguous conclusions. However, it is informative to note the difference between the two approaches: the Meelis test strongly rejects H_0, but does not indicate the size of the effect (the R value does indicate the size of the effect, but is unreliable without a measure of uncertainty). The p-value does not give the probability that H_0 is true and should not be interpreted as such. Meanwhile, the Bayesian procedure estimates the probability that H_0 is true, and the posterior distribution for ψ gives the effect size after having accounted for mortality, along with a measure of the uncertainty in the estimate of ψ. For *G. legneri*, Khidr et al. (2013) also provide a primary dataset which we can analyse without modelling mortality (Table 3). The conclusion is the same as for the secondary data, again with strong evidence of under-dispersion. One difference between the primary and secondary analyses is that for the primary data, the multiplicative binomial model is preferred, whereas for the secondary data, the double binomial model is preferred. We believe this is due to differences between the shape of the two distributions. Figure 3 shows the posterior predictive distribution for the number of male eggs laid (in a clutch of 10 eggs) for the six different scenarios (three models on both the primary and secondary data). We can see that for a given sex allocation model, the posterior predictions for the primary

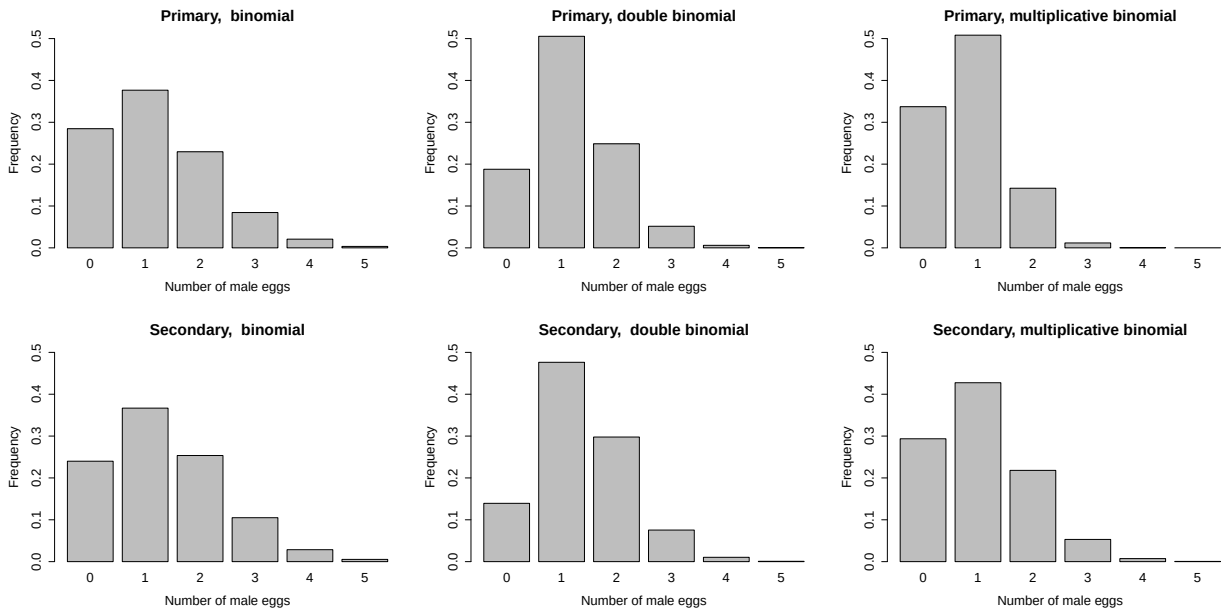


Figure 3: Posterior predictive distributions of the number of male eggs (pre-mortality) in a clutch of 10 eggs for the six different scenarios, namely the primary and secondary data for the three different models of sex allocation. The multiplicative binomial distribution gives the best fit to the primary data, and the double binomial distribution best fits the secondary data.

and secondary data are similar, and that the double and multiplicative distributions both give more concentrated (more precise) predictions than the binomial model. We can also see the difference between the shape of the double and multiplicative distributions, with the multiplicative distribution predicting more clutches with no males than the double binomial. The switch between preferred model for the secondary and primary datasets does not change our conclusion that there is strong evidence of under-dispersion.

Finally, note that the data and model are strongly informative about p and ψ, with the posterior and prior values being markedly different, whereas the posterior value for λ and d are close to the prior distribution. Experimentation (see the supplementary material) has shown that the posterior distributions of λ and d are sensitive to their prior distribution, but that the posterior of p and ψ are not sensitive to these choices.

5.2. Goniozus thailandensis: small dataset, medium mortality

Now we consider a dataset on the parasitoid species *Goniozus thailandensis* collected by Witethom and Gordh (1994). This species has a broadly similar biology to *G. legneri* and has previously been analysed for sex ratio variance by Hardy et al. (1998). The developmental

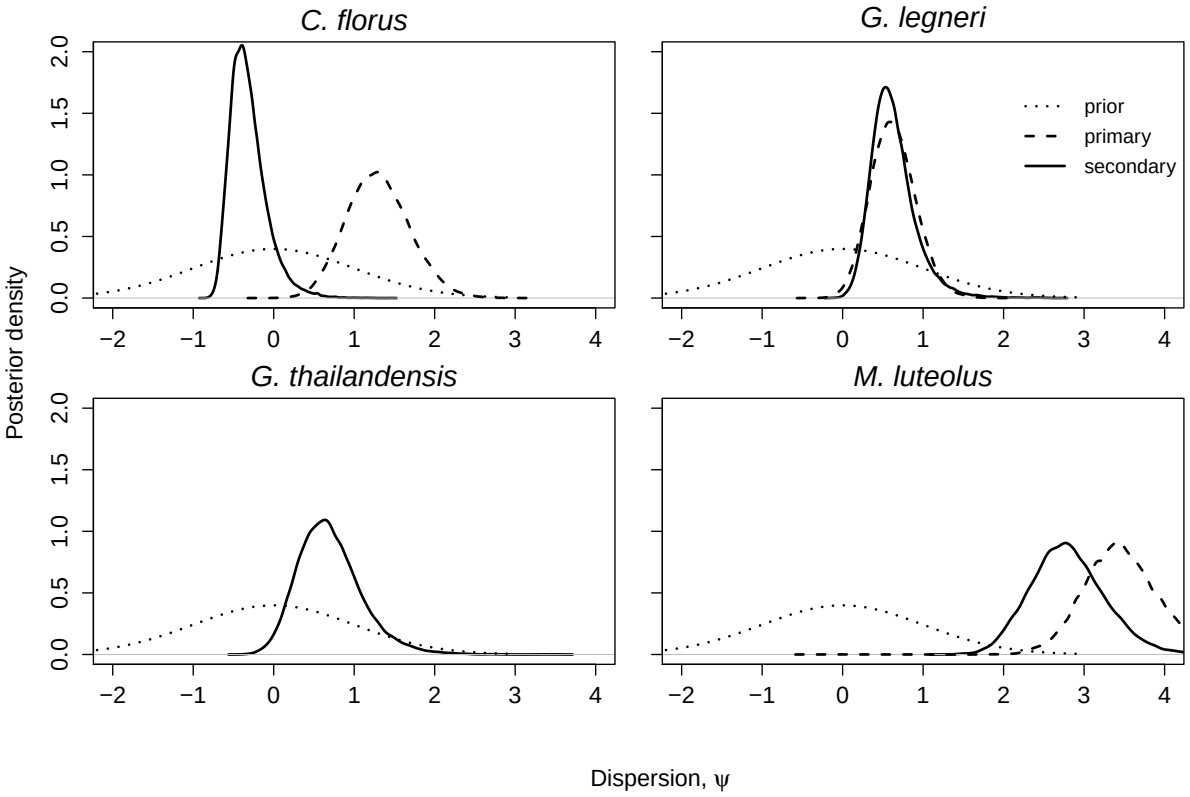


Figure 4: The marginal posterior distribution for ψ for the double binomial model for all four species. The results were obtained using 5×10^5 MCMC iterations. The posterior distributions for p, d and λ are not shown.

mortality rate, 22%, is higher than for *G. legneri* and the dataset is small, thus presenting a more challenging, and possibly more typical, case for analysis. Classical analysis of these data was inconclusive: the Meelis test gave $U = -0.73$ with a p-value of 0.23 and $R = 0.68$, which suggests under-dispersion, but with insufficient evidence to reject H_0 at the 5% significance level. In Section 3 we demonstrated that the Meelis test will lack power on this dataset, as there are only $C = 60$ observations and the probability of developmental mortality is moderate. This leaves us uncertain as to whether this result is due to the limited sample size, the relatively high mortality rate or to sex allocation actually being binomially distributed. The Meelis test only informs us that we cannot reject the null hypothesis due to insufficient evidence; it does not allow us to say that the species has binomially distributed sex allocation.

Carrying out the Bayesian analysis, using the prior distributions

$$d \sim \text{Beta}(5, 20) \qquad \lambda \sim \text{Gamma}(9, 1),$$

(consistent with the observed average clutch size and the mortality rate of 22%) we find the posterior distribution for ψ shown in the bottom left panel of Figure 4 and the Bayes factors given in Table 3. The Bayes factors suggest that there is substantial evidence in favour of the double binomial model over the other two models, and the posterior for ψ shows that under-dispersion is the best explanation of the data (the equi-tailed 95% credibility interval for ψ is $[0.0415, 1.61]$). The posterior distribution does contain a small amount of support for a zero or negative value of ψ (binomiality, or over-dispersion), showing that while this can not conclusively be ruled out, it is unlikely. Assuming equal prior probability for each model, there is a posterior probability of 0.79 that the double binomial model is the true model, and 0.14 that the binomial model (H_0) is true. While this is not conclusive evidence, it has allowed us to state that the data suggest under-dispersion over binomial sex allocation with posterior odds of more than 5 to 1. The posterior for ψ allows us to see the range of possible under-dispersion strengths that are consistent with the data. In comparison, the classical approach only allows us to conclude that there is insignificant evidence to reject H_0.

5.3. Colpoclypeus florus: medium dataset, high mortality

Primary and secondary data on *Colpoclypeus florus* are available from a study by Dijkstra (1986) analysed by Hardy et al. (1998). *C. florus* is a gregarious parasitoid with female biased sex ratios and is the only known member of its genus. The mortality rate was reported to be 57%, which when combined with the average clutch size of 7.4 motivated the prior distributions

$$d \sim \text{Beta}(11, 10) \qquad \lambda \sim \text{Gamma}(16, 1).$$

The results of the analysis of this data are shown in Table 3. These illustrate the tendency of mortality to make data appear less under-dispersed, possibly even over-dispersed. The primary data clearly show that the species has under-dispersed sex allocation, with the Meelis test and Bayes factors agreeing that there is very strong evidence in favour of under-dispersion. Whereas for the secondary data, the Meelis test fails to reject the null hypothesis, and the Bayes factors suggest that the binomial model is the best supported (posterior probability of 0.61, compared to 0.16+0.22=0.38 for the two non-binomial models). The 95% credibility interval for ψ is $[-0.063, 0.019]$ for the multiplicative model, and $[-0.65, 0.24]$ for

the double binomial model, both of which contain 0, showing that the data could be either under- or over-dispersed. The marginal posterior for ψ in Figure 4, shows how the primary data strongly suggest under-dispersion, but that the secondary data (after mortality) suggest over-dispersion, although there is still some support for under-dispersion. While the Meelis test can only lead us to conclude that there is no evidence to reject the hypothesis of binomial sex allocation, the Bayesian test can quantify that evidence and give a posterior probability that indicates that the hypothesis of binomial sex ratios is approximately twice as likely as the hypothesis of non-binomial sex allocation.

5.4. Metaphycus luteolus: large dataset, high mortality

A large secondary dataset on *M. luteolus* was presented in Kapranas et al. (2011). This species is a facultatively gregarious parasitoid which lays eggs inside hosts. Developing offspring may compete within the host, be attacked by the host immune responses, or die of other causes, and the overall mortality rate is approximately 40%. The secondary sex ratio is female biased. Using prior distributions

$$d \sim \text{Beta}(6, 10) \qquad \lambda \sim \text{Gamma}(4, 1)$$

we obtained the results presented in Table 3 and Figure 4. Due to the large sample sizes, and the effect size, all procedures give overwhelming evidence that the data are under-dispersed. By selecting only those clutches that did not experience any mortality, we can obtain an approximation of a primary dataset (this approach is discussed in Khidr et al., 2013). Analysis of this dataset again demonstrates the tendency of mortality to make data appear less under-dispersed.

5.5. Simulation study

We now show that by modelling mortality, we have increased our ability to detect under-dispersion. We analyse the performance of the Meelis test and the Bayes factor approach, using a simulation study in which we apply both procedures to synthetic datasets. The computational expense of the Bayesian approach (typically it takes 2-5 hours of computer time to analyse a single dataset), limited the study to 100 synthetic datasets, but this is sufficient to conclusively demonstrate an improved ability to find evidence against H_0, i.e., statistical power.

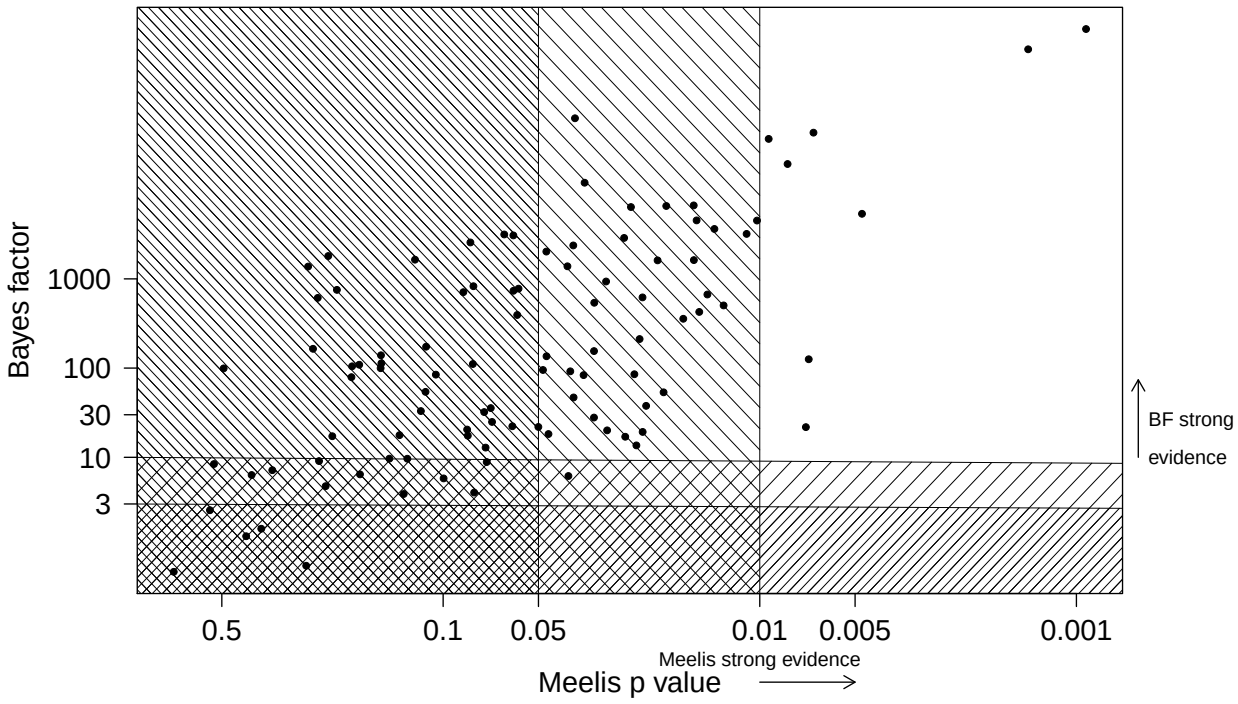


Figure 5: Results of the simulation study. Each point represents the p-value and Bayes factor (BF) for a simulated dataset. The shaded regions indicate p-value or Bayes factor ranges for which we would conclude there was either no, or weak evidence against H_0. The horizontal regions (with 'forwardslash' shading) indicates the Bayes factor is either less than 10 (threshold for strong evidence against H_0), or less than 3 (threshold for substantial evidence). The vertical regions ('backslash' shading) indicate p-values of less than 0.01 or 0.05. Note that the x-axis is reversed.

The synthetic datasets were simulated to each contain 50 clutches using a mortality rate of 30%, moderate values of C and d The model defined by Equations (3), (5) and (8), with $\lambda = 10, p = 0.1$, and $\psi = 0.3$, was used to simulate the datasets, giving a moderate level of under-dispersion comparable to *G. legneri*.

The results of the simulation study are summarised in Figure 5 and Table 4. For each dataset we have plotted the logarithm of the estimated Bayes factor between the multiplicative and binomial models, against the logarithm of the p-value from the Meelis test. The shading shows regions in which one or both of the procedures failed to detect strong evidence of under-dispersion, either because the p-value is greater than 0.05 (or 0.01), and/or because the Bayes factor is less than 3 (or 10). Table 4 summarises each procedure by the percentage of datasets which led to Bayes factors or p-values in a specified range.

Strength of evidence:	insubstantial	substantial	strong	very strong	decisive
Meelis p-value range:	> 0.1	0.05– 0.1	0.01–0.05	0.001– 0.01	< 0.001
% in range:	33	20	39	7	1
BF range:	0 – 3	3 – 10	10–30	30-100	> 100
% in range:	5	13	15	15	52

Table 4: Simulation study results: 100 synthetic datasets, all with moderate levels of under-dispersion ($\psi = 0.3$ in the multiplicative model) and mortality (30%), were analysed and grouped into categories indicating various levels of strength of evidence against H_0. The Bayesian approach can be seen to substantially outperform the Meelis test.

These results clearly demonstrate the improved power of the Bayesian procedure. For example, in more than half of the simulated datasets, the Meelis test returned a p-value greater than 0.05, which would indicate that there was insufficient evidence to reject the null hypothesis of binomial sex allocation. In contrast, 95% of the datasets provided at least substantial evidence against binomial sex ratios according to the Bayesian approach, and over half (52%) of the datasets provided decisive evidence ($BF > 100$). Furthermore, Figure 5 illustrates that every time the Bayesian test failed to detect under-dispersion, the Meelis test also failed, whereas there were 36 datasets where the Bayesian test indicated strong evidence ($BF > 10$) against H_0, but where the Meelis test failed (at the 5% level).

In order to confirm that this increased power is not due to a corresponding increase in the type I error rate (i.e., falsely rejecting H_0), a second simulation study was performed analysing synthetic datasets generated from the binomial model. For 200 simulated datasets, the Meelis test rejected H_0 (at $\alpha = 0.05$) in 3% of cases (i.e., it had approximately the assumed error rate). The Bayes factor gave $\mathbb{P}(H_0|D) \leq 0.05$ (i.e., strong evidence against H_0) in 6% of cases, showing that the increased power of the Bayesian approach is not due to an inflated type I error. The posterior distributions for ψ (available in the supplementary information), ruled out $\psi = 0$ in only one of the 200 simulated datasets.

6. Conclusions

We have shown that the current approaches used to detect under- or over-dispersion in sex allocation lack power when the sample size is small or the mortality rate is moderate to

large. Both are common situations in empirical studies. For example, the Meelis test will usually fail to reject the null hypothesis under these conditions even when sex allocation is strongly non-binomial. We have introduced a new approach to detecting under- or over-dispersion that has much greater power for detecting departures from binomial allocation. The approach gains its power by explicitly modelling mortality, so that the test takes into account that the patterns in the data have occurred through a combination of sex allocation and mortality. The method can be extended further to include non-binomial distributions of mortality (e.g., Hardy et al., 1998; Kapranas et al., 2011). Furthermore, using a Bayesian approach to model selection and parameter estimation increases our descriptive ability: the posterior distribution of the dispersion parameter ψ allows both the size of the effect and the range of possible effects that are consistent with the data to be identified. Using Bayes factors allows us to give the posterior probability that the data derive from a species that has binomially distributed sex allocation, as opposed to p-values, which although commonly interpreted as probabilities, should not be (Goodman, 2008). In situations where the evidence is conclusively in favour of one hypothesis, our test generates the same conclusion as current approaches (but with improved descriptive ability). However, when the evidence is weaker, the additional information provided by the Bayesian approach can allow us to make useful inferences, even if these cannot be conclusive.

7. Coda

The software implementing this approach has been written in R (R Development Core Team, 2008) and is freely available (`https://github.com/rich-d-wilkinson/precision`) as the `precision` R package on github. Details of how to use and install the package are given in the package vignette and in the supplementary material. There are many possible extensions to this approach, primarily through changes and improvements to the model. For example, the binomial mortality model is relatively simple and other more complex models (such as over-dispersion) are possible. These extensions are straightforward to make within the Bayesian testing framework.

The data used in this paper are all available within the `precision` R package (see the package vignette). These datasets, as well as additional data on the sexual compositions

of offspring groups, are available from several previous publications. Secondary sex ratio datasets can be found in Morgan and Cook (1994); Hardy and Cook (1995); Nagelkerke and Sabelis (1998); Mackauer and Völkl (2002); Kapranas et al. (2008, 2009) and Khidr et al. (2013). Primary sex ratios are more difficult to evaluate, but datasets are available in Dijkstra (1986); Avilés et al. (2000), and Khidr et al. (2013).

Acknowledgements: We thank S. K. Khidr, B. Witethom and L. J. Dijkstra for help with data. We thank Andrew Wood for useful advice on the statistical approach taken, and R. F. Green and three anonymous referees for constructive suggestions. Apostolos Kapranas was funded by a Marie Curie Fellowship (FP7-PEOPLE-IEF-273431).

Altham, P. M. E., 1978. Two generalizations of the binomial distribution. Journal of the Royal Statistical Society. Series C (Applied Statistics) 27, 162–167.

Ambrosini, R., Rubolini, D., Saino, N., 2014. Analysis of sex sequences by means of generalized linear mixed models. Behavioral Ecology and Sociobiology 68, 1367–1377.

Avilés, L., McCormack, J., Cutter, A., Bukowski, T., 2000. Precise, highly female–biased sex ratios in a social spider. Proceedings of the Royal Society of London B: Biological Sciences 267, 1445–1449.

Bowers, E. K., Munclinger, P., Bureš, S., Kučerová, L., Nádvorník, P., Krist, M., 2013. Cross-fostering eggs reveals that female collared flycatchers adjust clutch sex ratios according to parental ability to invest in offspring. Molecular Ecology 22, 215–228.

Chib, S., 1995. Marginal likelihood from the Gibbs output. Journal of the American Statistical Association 90, 1313–1321.

Chib, S., Jeliazkov, I., 2001. Marginal likelihood from the metropolis-hastings output. Journal of the American Statistical Association 96, 270–281.

Cole, L. R., 1981. A visible sign of a fertilization action during oviposition by an ichneumonid wasp, *itoplectis maculator*. Animal Behaviour 29, 299–300.

606 Dietrich-Bischoff, V., Schmoll, T., Winkel, W., Krackow, S., Lubjuhn, T., 2006. Extra-pair
607 paternity, offspring mortality and offspring sex ratio in the socially monogamous coal tit
608 (*Parus ater*). Behavioral Ecology and Sociobiology 60, 563–571.

609 Dijkstra, L. J., 1986. Optimal selection and exploitation of hosts in the parasitic wasp
610 *Colpoclypeus florus* (Hym., Eulophidae). Netherlands Journal of Zoology 36, 177–301.

611 Dyrcz, A., Sauer-Gürth, H., Tkadlec, E., Wink, M., 2004. Offspring sex ratio variation
612 in relation to brood size and mortality in a promiscuous species: the Aquatic Warbler
613 *Acrocephalus paludicola*. Ibis 146, 269–280.

614 Efron, B., 1986. Double exponential families and their use in generalized linear regression.
615 Journal of the American Statistical Association 81, 709–721.

616 Ewen, J. G., Cassey, P., King, R. A., Brittingham, M., 2003. Assessment of the randomization
617 test for binomial sex-ratio distributions in birds. The Auk 120, 62–68.

618 Ewen, J. G., Cassey, P., Møller, A. P., 2004. Facultative primary sex ratio variation: a lack
619 of evidence in birds? Proceedings of the Royal Society of London, Series B: Biological
620 Sciences 271, 1277–1282.

621 Forsyth, D. M., Tustin, K. G., Gaillard, J.-M., Loison, A., 2004. Fetal sex ratio variation in
622 the highly polygynous Himalayan tahr: evidence for differential male mortality. Behavioral
623 Ecology 15, 572–578.

624 Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B., 2003. Bayesian data analysis. CRC
625 press.

626 Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian
627 restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6,
628 721–741.

629 Gilks, W., Richardson, S., Spiegelhalter, D., 1996. Markov Chain Monte Carlo in Practice.
630 Interdisciplinary Statistics. Chapman & Hall.

631 Goodman, S., 2008. A dirty dozen: Twelve p-value misconceptions. Seminars in Hematology
632 45, 135–140.

633 Green, R. F., Gordh, G., Hawkins, B. A., 1982. Precise sex ratios in highly inbred parasitic
634 wasps. American Naturalist 120, 653–665.

635 Hamilton, W. D., 1967. Extraordinary sex ratios. Science 156, 477–488.

636 Hardy, I. C., Cook, J. M., 1995. Brood sex ratio variance, developmental mortality and
637 virginity in a gregarious parasitoid wasp. Oecologia 103, 162–169.

638 Hardy, I. C. W., 1992. Non-binomial sex allocation and brood sex ratio variances in the
639 parasitoid Hymenoptera. Oikos 65, 143–158.

640 Hardy, I. C. W., Dijkstra, L. J., Gillis, J. E. M., Luft, P. A., 1998. Patterns of sex ratio,
641 virginity and developmental mortality in gregarious parasitoids. Biological Journal of the
642 Linnean Society 64, 239–270.

643 Heinsohn, R., Legge, S., Barry, S., 1997. Extreme bias in sex allocation in *Eclectus* parrots.
644 Proceedings of the Royal Society of London. Series B: Biological Sciences 264, 1325–1329.

645 James, W. H., 1975. Sex ratio and the sex composition of the existing sibs. Annals of Human
646 Genetics 38, 371–378.

647 Jeffreys, H., 1939. Theory of probability. International series of monographs on physics. The
648 Clarendon press.

649 Kapranas, A., Hardy, I. C. W., Morse, J. G., Luck, R. F., 2011. Parasitoid developmental
650 mortality in the field: patterns, causes and consequences for sex ratio and virginity. Journal
651 of Animal Ecology 80, 192–203.

652 Kapranas, A., Pacheco, P., Forster, L., Morse, J. G., Luck, R. F., 2008. Precise sex alloca-
653 tion manifested by several encyrtid parasitoids of brown soft scale *Coccus hesperidum* L.
654 (Hemiptera: Coccidae). Behavioral Ecology and Sociobiology 62, 901–912.

655 Kapranas, A., Wajnberg, E., Luck, R. F., 2009. Sequences of sex allocation and mortality
656 in clutches of *metaphycus* parasitoids of soft scale insects and the prevalence of all-female
657 broods. Ecological Entomology 34, 652–662.

658 Kass, R. E., Raftery, A. E., 1995. Bayes Factors. Journal of the American Statistical Asso-
659 ciation 90, 773–795.

660 Khidr, S. K., Mayes, S., Hardy, I. C. W., 2013. Primary and secondary sex ratios in a
661 gregarious parasitoid with local mate competition. Behavioral Ecology 24, 435–443.

662 Kolman, W. A., 1960. The mechanism of natural selection for the sex ratio. American Nat-
663 uralist 94, 373–377.

664 Krackow, S., Meelis, E., Hardy, I. C. W., 2002. Analysis of sex ratio variances and sequences
665 of sex allocation. In: Hardy, I. C. W. (Ed.), Sex Ratios: Concepts and Research Methods.
666 Cambridge: Cambridge University Press, pp. 112–131.

667 Lambin, X., 1994. Sex ratio variation in relation to female philopatry in Townsend's voles.
668 Journal of Animal Ecology 63, 945–953.

669 Macdonald, D. W., Johnson, P. J., 2008. Sex ratio variation and mixed pairs in roe deer:
670 evidence for control of sex allocation? Oecologia 158, 361–370.

671 Mackauer, M., Völkl, W., 2002. Brood-size and sex-ratio variation in field populations of
672 three species of solitary aphid parasitoids (Hymenoptera: Braconidae, Aphidiinae). Oe-
673 cologia 131, 296–305.

674 McCullagh, P., Nelder, J. A., 1989. Generalized linear models. Vol. 37. CRC press.

675 Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E., 1953.
676 Equation of State Calculations by Fast Computing Machines. The Journal of Chemical
677 Physics 21, 1087–1092.

678 Morgan, D. J., Cook, J. M., 1994. Extremely precise sex ratios in small clutches of a bethylid
679 wasp. Oikos, 423–430.

Nagelkerke, C. J., 1996. Discrete clutch sizes, local mate competition, and the evolution of precise sex allocation. Theoretical population biology 49, 314–343.

Nagelkerke, C. J., Hardy, I. C. W., 1994. The influence of developmental mortality on optimal sex allocation under local mate competition. Behavioral Ecology 5, 401–411.

Nagelkerke, C. J., Sabelis, M. W., 1998. Precise control of sex allocation in pseudo-arrhenotokous phytoseiid mites. Journal of Evolutionary Biology 11, 649–684.

Nakagawa, S., Cuthill, I. C., 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biological Reviews 82, 591–605.

Øigarden, T., Lifjeld, J. T., 2013. Primary sex ratios vary with clutch size in the size-dimorphic White-throated Dipper *Cinclus cinclus*. Journal of Ornithology 154, 91–97.

Postma, E., Heinrich, F., Koller, U., Sardell, R. J., Reid, J. M., Arcese, P., Keller, L. F., 2011. Disentangling the effect of genes, the environment and chance on sex ratio variation in a wild bird population. Proceedings of the Royal Society B: Biological Sciences 278 (1720), 2996–3002.

R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Robert, C., 2001. The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation. Springer Texts in Statistics. Springer.

Verner, J., 1965. Selection for sex ratio. The American Naturalist 99, 419–421.

West, S. A., 2009. Sex allocation. Vol. 44. Princeton University Press.

West, S. A., Herre, E. A., 1998. Stabilizing selection and variance in fig wasp sex ratios. Evolution 52, 475–485.

Witethom, B., Gordh, G., 1994. Development and life table of *Goniozus thailandensis* Gordh & Witethom (Hymenoptera: Bethylidae), a gregarious ectoparasitoid of a phycitine fruit borer (Lepidoptera: Pyralidae). Journal of the Science Society of Thailand 20, 101–114.

Supplementary information for *Detecting non-binomial sex allocation when developmental mortality operates*

Richard D. Wilkinson, Apostolos Kapranas, Ian C.W. Hardy

1 Simulation studies

A second simulation study was performed to check that the increased power of the Bayesian approach was not due to biased estimates causing the test to have a high type I error rate, i.e., falsely rejecting the null hypothesis when it is true more commonly than the frequentist tests. Using $\lambda = 10$, $C = 100$, and $d = 0.3$, we generated 200 datasets from the binomial sex-allocation model, choosing the sex-ratio parameter p to be uniform in $[0.05, 0.3]$. For each dataset we applied the existing tests and the Bayesian approach described in the main paper.

The posterior distributions obtained in each case are shown in Figure S1. In only one of the 200 datasets did the posterior suggest $\psi = 0$ was not consistent with the data, a result that can easily occur by chance with small datasets and high mortality. The Meelis and James tests also rejected H_0 in this case. The Bayes factor analysis gave $\mathbb{P}(H_0|D) \leq 0.05$ (i.e., strong evidence against the binomial model) in only 6% of cases, whereas the Meelis and James tests both rejected H_0 in 3% of cases. These results suggest that the significance level of both these tests is approximately as claimed, and that the Bayesian approach does not gain its power through having a substantially higher type I error rate.

One weakness of our approach is that only a few datasets gave strong evidence in favour of H_0. This is because the binomial model is a special case of the two alternative models. Note that the frequentist tests are, by definition, incapable of providing evidence in favour of H_0, so the limitation of the Bayesian approach is not a weakness in comparison.

1.1 Prior sensitivity

To test the sensitivity of our results to the choice of prior distributions, we conducted a small simulation study. Using the *G. legneri* data, we repeated the analysis from Section 5.1 of the main paper using the prior distribution $\psi \sim N(0, \sigma^2)$ with σ^2 taking values $0.5, 1, 2$, and 10. Larger values of ψ can be ruled out *a priori*, as discussed in the main

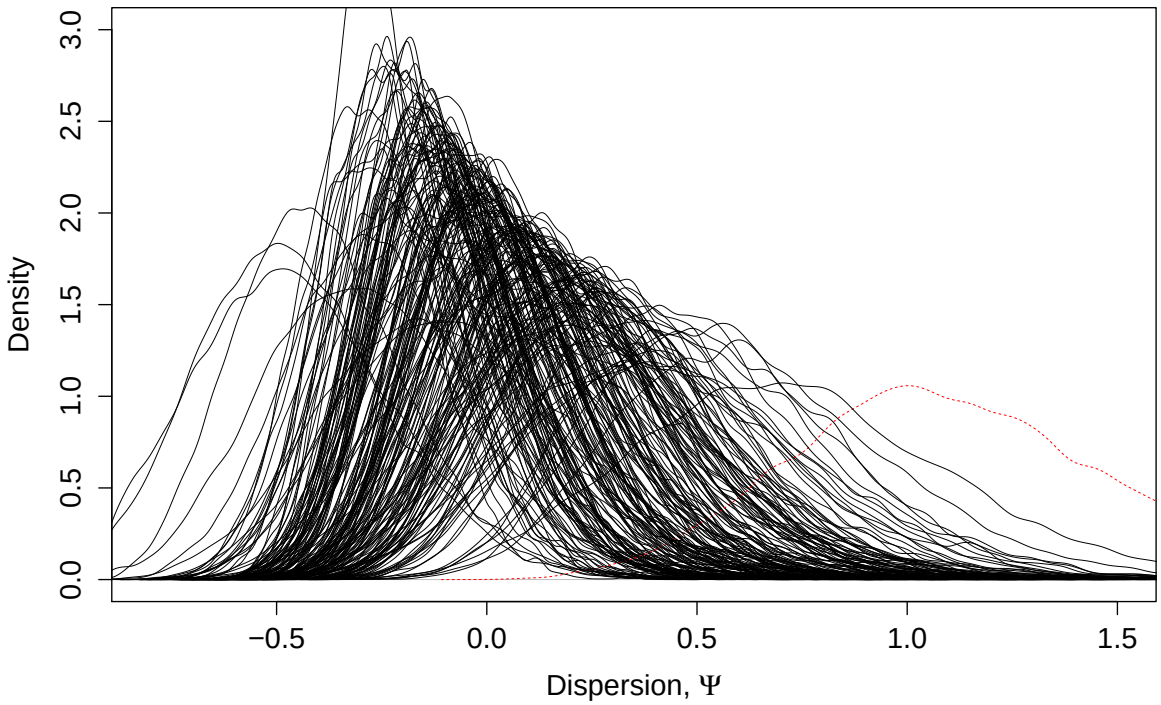


Figure S1: The posterior distributions for ψ in the double binomial model for each of the 200 simulated datasets. Note that only one of these curves (red dashed) conclusively rules out $\psi = 0$, the true value. The plots for the multiplicative binomial model are similar (not shown).

σ^2	0.5	1	2	10
BF	253	214	162	79

Table S1: Bayes factor estimates for comparing the double binomial and binomial models when analysing the *G. legneri* data using a $N(0, \sigma^2)$ prior distribution for ψ.

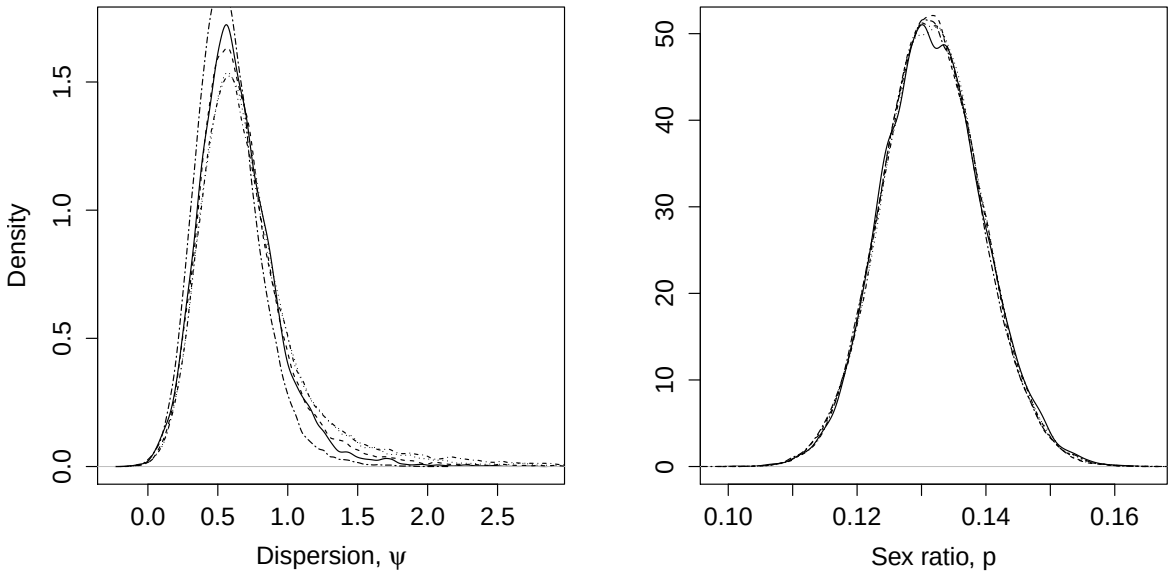


Figure S2: The posterior distributions for the double binomial model when analysing the *G. legneri* secondary data using a $\psi \sim N(0, \sigma^2)$ prior distribution $\sigma^2 = 0.5, 1, 2, 10$.

paper. The posterior distributions for p and ψ for the double binomial model are shown in Figure S2. Note that the posterior distributions are largely unchanged. As expected, the Bayes factor for comparing the double binomial and binomial model do change as the prior distribution changes (Table S1). As the prior for ψ becomes more diffuse, the BF reduces, as is expected (Bernardo and Smith, 2000). Even when using an unrealistic $N(0, 10)$ prior distribution for ψ, the Bayes factor still indicates very strong evidence for the double binomial model over the binomial model.

2 Algorithmic details

Here we provide the technical detail on how to estimate the posterior distributions and the Bayes factors.

2.1 Estimating the posterior distributions

To sample from the posterior distribution of interest $\pi(\psi|D)$, we use a Metropolis-Hastings within Gibbs sampler (Metropolis et al., 1953; Geman and Geman, 1984). We introduce the missing N_i and M_i values, denoted $\mathbf{N}$ and $\mathbf{M}$ as auxiliary variables, and sample from

the distribution $\pi(\psi, p, \lambda, d, \mathbf{N}, \mathbf{M}|D)$, which admits $\pi(\psi|D)$ as a marginal distribution. For the priors in Equation (9), the full conditional distributions of λ and d can be found, with

$$\pi(\lambda|p, \psi, d, \mathbf{N}, \mathbf{M}, D) = \Gamma\left(\lambda; a + \sum N_i, C + b\right)$$

and

$$\pi(d|p, \psi, \lambda, \mathbf{N}, \mathbf{M}, D) = \text{Beta}\left(d; a' + \sum(N_i - n_i), b' + \sum n_i\right)$$

allowing a Gibbs update step to be used for these parameters. For the model in which sex allocation is binomial ($\psi = 0$), there is also a Gibbs sampler for p with

$$\pi(p|\psi, \lambda, d, \mathbf{N}, \mathbf{M}, D) = \text{Beta}\left(p; \alpha + \sum M_i, \beta + \sum(N_i - M_i)\right)$$

if $p \sim \text{Beta}(\alpha, \beta)$ a priori (if $\alpha = \beta = 1$ then $p \sim U[0, 1]$).

In the case where $\psi \neq 0$, we need to use a Metropolis-Hastings update for both p and ψ. We use a symmetric Gaussian random walk on both parameters, working on the logit scale of p to avoid difficulties with finite prior support regions:

$$\psi' = \psi + \sigma_\psi Z, \qquad \text{logit}(p') = \text{logit}(p) + \sigma_p Z'$$

where Z and Z' are independent $N(0, 1)$ random variables. As both proposals are symmetric, they cancel from the Metropolis-Hastings acceptance rate, giving acceptance probability

$$\begin{aligned}\frac{\pi(p', \psi'|\lambda, d, N, M, n, m)}{\pi(p, \psi|\lambda, d, N, M, n, m)} &= \frac{\pi(N, M, n, m|\lambda, d, p', \psi')}{\pi(N, M, n, m|\lambda, d, p, \psi,)} \frac{\pi(p', \psi'|\lambda, d)}{\pi(p, \psi|\lambda, d)} \\ &= \frac{\pi(M|N, p', \psi')}{\pi(M|N, p, \psi)} \frac{\pi(p')\pi(\psi')}{\pi(p)\pi(\psi)} \end{aligned} \tag{S1}$$

with $\pi(M|N, p, \psi)$ given by Equation (3). We have found that using $\sigma_\psi = 0.2$ and $\sigma_p = 0.3$ provides a good compromise between mixing and acceptance rate.

To update the $\mathbf{N}$ and $\mathbf{M}$ values, we update each (N_i, M_i) pair separately. We use the relationship

$$\begin{aligned}\pi(N_i, M_i|n_i, m_i, \lambda, p, \psi, d) \propto \pi(n_i|N_i, d_i)\pi(m_i|M_i, N_i, n_i, d)\pi(M_i|N_i, p, \psi) \\ \cdot \pi(N_i|\lambda)\mathbb{I}_{M_i \geq m_i,\ N_i \geq n_i}\mathbb{I}_{M_i - m_i \leq N_i - n_i}\end{aligned}$$

where each expression in this equation has been calculated previously or is part of the model definition. We use the prior distribution of n and m, conditioned to satisfy the three inequalities in Equation (2), as an independence sampler proposal for (N_i, M_i).

$$q((N_i M_i), (N_i' M_i')) = \pi(M_i'|N_i', p, \psi,)\pi(N_i'|\lambda)\mathbb{I}_{M_i' \geq m_i,\ N_i' \geq n_i}\mathbb{I}_{M_i' - m_i \leq N_i' - n_i}$$

which we can simulate from using the rejection algorithm. It is more efficient to reject infeasible values of N and M at the proposal stage rather than in the MCMC acceptance, as it leads to higher MCMC acceptance rates and thus quicker mixing. This gives the Metropolis-Hastings acceptance probability

$$\alpha((N_i M_i), (N_i' M_i')) = \min\left(1, \frac{\pi(m_i|M_i', N_i', n_i)}{\pi(m_i|M_i, N_i, n_i)} \frac{\pi(n_i|N_i', d)}{\pi(n_i|N_i, d)}\right).$$

The Markov chain sampler then alternates between updating the four parameters and updating each of the C pairs (M_i, N_i).

2.2 Bayes factor estimation

We use the approach described in Chib (1995) and Chib and Jeliazkov (2001) to estimate the Bayes factors. This relies upon the identity

$$\pi(\mathbf{n}, \mathbf{m}) = \frac{\pi(\mathbf{n}, \mathbf{m}|\theta^*)\pi(\theta^*)}{\pi(\theta^*|\mathbf{n}, \mathbf{m})}. \tag{S2}$$

to estimate the evidence for each model. This holds for all θ^*, but works best when θ^* has large posterior support, such as when $\theta^* = \arg\max \pi(\theta|\mathbf{n}, \mathbf{m})$. To evaluate the likelihood contribution, we note that

$$\pi(\mathbf{n}, \mathbf{m}|\theta) = \sum_{\mathbf{N},\mathbf{M}} \pi(\mathbf{n}, \mathbf{m}|\mathbf{N}, \mathbf{M}, \theta)\pi(\mathbf{N}, \mathbf{M}|\theta). \tag{S3}$$

and that

$$\pi(\mathbf{n}, \mathbf{m}|\mathbf{N}, \mathbf{M}, \theta)\pi(\mathbf{N}, \mathbf{M}|\theta) = \prod_{j=1}^{C} \pi(n_j|N_j, \theta)\pi(m_j|M_j, N_j, n_j)\pi(M_j|N_j, \theta)\pi(N_j|\theta).$$

Estimating Equation (S3) using Monte Carlo integration does not work well, due to the extreme variance of the resulting estimator, and so instead we directly calculate each summand, evaluating the M_i sum over the range allowed by the inequalities in Equation (2), and truncating the sum with respect to N_i when each term drops below a value of 10^{-6}, ensuring good accuracy in the estimate.

Estimation of the denominator in Equation (S2) is more difficult, and we need a slightly different approach for the two models. For the binomial model, we note that

$$\pi(\theta|\mathbf{n}, \mathbf{m}) = \sum_{\mathbf{N},\mathbf{M}} \pi(\theta|\mathbf{N}, \mathbf{M}, \mathbf{n}, \mathbf{m})\pi(\mathbf{N}, \mathbf{M}|\mathbf{n}, \mathbf{m}) \tag{S4}$$

which we can estimate using

$$\frac{1}{B}\sum_{i=1}^{B} \pi(\theta^*|\mathbf{N}^{(i)}, \mathbf{M}^{(i)}, \mathbf{n}, \mathbf{m}) \tag{S5}$$

where $\mathbf{N}^{(i)}, \mathbf{M}^{(i)} \sim \pi(\mathbf{N}, \mathbf{M}|\mathbf{n}, \mathbf{m})$, $i = 1, \ldots, B$, are simulated random vectors from the posterior distribution $\pi(\theta, \mathbf{N}, \mathbf{M}|\mathbf{n}, \mathbf{m})$. The summands are explicitly available for the binomial model:

$$\begin{aligned}\pi(p, \lambda, d \mid \mathbf{N}, \mathbf{M}, \mathbf{n}, \mathbf{m}) = \text{Beta}(p; \alpha + \sum_{i=1}^{C} M_i, \beta + \sum_{i=1}^{C}(N_i - M_i)) \times \\ \Gamma(\lambda; a + \sum_{i=1}^{C} N_i, C + b)\text{Beta}(d; a' + \sum_{i=1}^{C}(N_i - n_i), b' + \sum_{i=1}^{C} n_i).\end{aligned} \tag{S6}$$

Estimating the denominator for the multiplicative and double binomial models is more difficult, as the summand in Equation (S5) cannot be explicitly calculated. We instead split the parameter $\theta = (\lambda, p, \psi, d)$ into two blocks $\theta = (\theta_1, \theta_2)$ where $\theta_1 = (p, \psi)$ and $\theta_2 = (\lambda, d)$ and use the identity

$$\pi(\theta_1^*, \theta_2^*|\mathbf{n}, \mathbf{m}) = \pi(\theta_1^*|\mathbf{n}, \mathbf{m})\pi(\theta_2^*|\mathbf{n}, \mathbf{m}, \theta_1^*).$$

The second term on the right is the easier to evaluate, as

$$\pi(\lambda^*, d^*|\mathbf{n}, \mathbf{m}, p^*, \psi^*) = \sum_{\mathbf{N},\mathbf{M}} \pi(\lambda^*, d^*|\mathbf{N}, \mathbf{M}, \mathbf{n}, \mathbf{m}, p^*, \psi^*)\pi(\mathbf{N}, \mathbf{M}|\mathbf{n}, \mathbf{m}, p^*, \psi^*)$$

and

$$\pi(\lambda, d|\mathbf{N}, \mathbf{M}, \mathbf{n}, \mathbf{m}, p^*, \psi^*) = \Gamma(\lambda;\ \alpha + \sum_{i=1}^{C} N_i, C + \beta) \times \\ \text{Beta}(d; a + \sum_{i=1}^{C}(N_i - n_i), b + \sum_{i=1}^{C} n_i)$$

as before. We can simulate $\mathbf{N}^{(i)}, \mathbf{M}^{(i)} \sim \pi(\mathbf{N}, \mathbf{M}|\mathbf{n}, \mathbf{m}, p^*, \psi^*)$ using a Gibbs sampler with p and ψ fixed and use a Monte Carlo estimate of the sum.

The term $\pi(\theta_1^*|\mathbf{n}, \mathbf{m}) \equiv \pi(p^*, \psi^*|\mathbf{n}, \mathbf{m})$ is more difficult to evaluate (Chib and Jeliazkov, 2001). Our approach relies on the fact that the subkernel of the Markov Chain on θ_1 satisfies the detailed balance equations. Consider sampling from $\pi(\theta_1|\mathbf{n}, \mathbf{m}, \mathbf{N}, \mathbf{M}, \theta_2)$ using the Metropolis-Hastings algorithm with proposal $q(\theta_1, \theta_1')$, and acceptance rate

$$r(\theta_1, \theta_1') = \min\left(1, \frac{q(\theta_1', \theta_1)\pi(\theta_1'|\mathbf{n}, \mathbf{m}, \mathbf{N}, \mathbf{M}, \theta_2)}{q(\theta_1, \theta_1')\pi(\theta_1|\mathbf{n}, \mathbf{m}, \mathbf{N}, \mathbf{M}, \theta_2)}\right),$$

which is given by Equation (S1). The subkernel of this Markov chain is

$$p(\theta_1, \theta_1') = q(\theta_1, \theta_1')r(\theta_1, \theta_1').$$

By rearranging the detailed balance equation, we find the identity

$$\pi(\theta_1^*|n, m) = \frac{\mathbb{E}(p(\theta_1, \theta_1^*))}{\mathbb{E}(r(\theta_1^*, \theta_1))} \tag{S7}$$

where q and r will potentially depend upon $\mathbf{N}, \mathbf{M}$ and θ_2 (suppressed in the notation). The expectation in the numerator is with respect to the distribution $\pi(\theta_1, \theta_2, \mathbf{N}, \mathbf{M}|\mathbf{n}, \mathbf{m})$, from which we have generated samples using the full Markov chain. The expectation in the denominator is with respect to $\pi(\theta_2, \mathbf{N}, \mathbf{M}|\mathbf{n}, \mathbf{m}, \theta_1^*)q(\theta_1^*, \theta_1)$ which we can sample from by simulating from $\pi(\lambda, d, \mathbf{N}, \mathbf{M}|\mathbf{n}, \mathbf{m}, p^*, \psi^*)$ using a Gibbs sampler, and then generating (p, ψ) values by simulating from $q(\theta_1^*, \theta_1)$ for each realisation from the first chain. We can then estimate both terms in Equation (S7) using the Monte Carlo sum approximation to the integral.

3 Meelis test details

Suppose $X_1, \ldots, X_C$ are independent identically distributed $\text{Bin}(n, p)$ random variables. Nagelkerke and Sabelis (1998) showed that the test statistic U, defined by

$$U = \frac{\sum_{i=1}^{C} X_i^2 - f}{\sqrt{V}}$$

where

$$
\begin{aligned}
f &= \frac{v(v(n-1)+n(C-1))}{Cn-1} \\
v &= \sum_{i=1}^{C} X_i \\
x^{(j)} &= x(x-1)\dots(x-j+1) \\
V &= \frac{v^{(4)}(n-1)(Cn(n-1)-(4n-6))}{(Cn-1)^{(3)}} + \frac{4v^{(3)}(n-1)^{(2)}}{(Cn-1)^{(2)}} \\
&\quad + \frac{2v^{(2)}(n-1)}{Cn-1} - \frac{v^2(v-1)^2(n-1)^2}{(Cn-1)^2}
\end{aligned}
$$

follows a standard normal distribution provided that C is sufficiently large, and that n is not very small (unless C is large).

4 R package

4.1 Installation

The easiest way to install is to use devtools to install directly from github.

```
devtools::install_github('rich-d-wilkinson/precision')
```

Alternatively, download the package from https://github.com/rich-d-wilkinson/precision and install manually.

4.2 Data

All of the datasets used in the paper are included in the R package.

```
library(precision)
data(package='precision')$results[,c('Item')]
```

```
##  [1] "CflorusPrimary"                  "CflorusSecondary"
##  [3] "GlegneriPrimary"                 "GlegneriSecondary"
##  [5] "GthailandensisSecondary"         "MluteolusPrimary"
##  [7] "MluteolusSecondary"              "hyper (CflorusSecondary)"
##  [9] "hyper (GlegneriSecondary)"       "hyper (GthailandensisSecondary)"
## [11] "hyper (MluteolusSecondary)"
```

For example, the C. florus secondary dataset is

```
data(CflorusSecondary)
tail(CflorusSecondary)
```

```
##        n  m
## [48,] 15  6
## [49,] 19  4
## [50,] 21  3
## [51,] 22  7
## [52,] 23 10
## [53,] 24  5
```

To use your own dataset, specify a $C \times 2$ matrix, with the first column containing the clutch size, and the second the number of males. It is necessary to label your columns as n and m.

```
my_data <- matrix(c(3,2,4,3,5,1,6,2,7,1,7,1), nc=2, byrow=TRUE)
colnames(my_data) <- c('n', 'm')
my_data
```

4.3 Standard Analyses

The pre-existing analysis methods are all built into the R functions `Meelis.test` and `James.test`. For example,

```
(meelis.out <- Meelis.test(CflorusSecondary, TwoSided = TRUE))
```

```
## $vals
##       clutch size no. clutches     p hat binom var   obs var         R
##  [1,]           1            8 0.0000000 0.0000000 0.0000000       NaN
##  [2,]           2            6 0.4166667 0.4861111 0.1666667 0.3428571
##  [3,]           3            3 0.3333333 0.6666667 1.0000000 1.5000000
##  [4,]           4            5 0.2000000 0.6400000 0.2000000 0.3125000
##  [5,]           5            6 0.3000000 1.0500000 0.3000000 0.2857143
##  [6,]           6            2 0.4166667 1.4583333 0.5000000 0.3428571
##  [7,]           7            2 0.1428571 0.8571429 0.0000000 0.0000000
##  [8,]           8            2 0.2500000 1.5000000 2.0000000 1.3333333
##  [9,]           9            4 0.5833333 2.1875000 2.2500000 1.0285714
## [10,]          10            2 0.2500000 1.8750000 0.5000000 0.2666667
## [11,]          12            3 0.4722222 2.9907407 5.3333333 1.7832817
## [12,]          13            1 0.3076923 2.7692308 0.0000000        NA
## [13,]          14            1 0.7142857 2.8571429 0.0000000        NA
## [14,]          15            3 0.2666667 2.9333333 7.0000000 2.3863636
## [15,]          19            1 0.2105263 3.1578947 0.0000000        NA
## [16,]          21            1 0.1428571 2.5714286 0.0000000        NA
## [17,]          22            1 0.3181818 4.7727273 0.0000000        NA
```

```
## [18,]          23            1 0.4347826 5.6521739 0.0000000        NA
## [19,]          24            1 0.2083333 3.9583333 0.0000000        NA
##                M          V          U    p value
##  [1,]   0.000000  0.0000000        NaN        NaN
##  [2,]   6.818182  1.5426997 -1.4638501 0.07161745
##  [3,]   4.500000  1.6071429  0.3944053 0.65335909
##  [4,]   5.894737  2.5028516 -1.1976540 0.11552588
##  [5,]  18.931034  9.5124851 -1.2745587 0.10123273
##  [6,]  14.090909  4.6280992 -0.5070926 0.30604494
##  [7,]   2.923077  0.9940828 -0.9258201 0.17726974
##  [8,]   9.600000  4.3323077  0.1921765 0.57619803
##  [9,] 117.000000 27.7219251  0.0000000 0.50000000
## [10,]  14.473684  6.8437347 -0.5633235 0.28660732
## [11,] 102.485714 35.7355102  0.7551601 0.77492354
## [12,]  16.000000  0.0000000        NaN        NaN
## [13,] 100.000000  0.0000000        NaN        NaN
## [14,]  54.000000 33.4883721  1.3824294 0.91658006
## [15,]  16.000000  0.0000000        NaN        NaN
## [16,]   9.000000  0.0000000        NaN        NaN
## [17,]  49.000000  0.0000000        NaN        NaN
## [18,] 100.000000  0.0000000        NaN        NaN
## [19,]  25.000000  0.0000000        NaN        NaN
##
## $R.av
## [1] 0.7532786
##
## $s2
## [1] 1.181833
##
## $U.av
## [1] -0.9672868
##
## $p.av
## [1] 0.3334007
##
## $exp.table
##
##      0 1 2 3 4 5 6 7 10
##   1  8 0 0 0 0 0 0 0  0
##   2  1 5 0 0 0 0 0 0  0
##   3  1 1 1 0 0 0 0 0  0
##   4  1 4 0 0 0 0 0 0  0
##   5  0 3 3 0 0 0 0 0  0
##   6  0 0 1 1 0 0 0 0  0
##   7  0 2 0 0 0 0 0 0  0
##   8  0 1 0 1 0 0 0 0  0
##   9  0 0 0 0 2 0 1 1  0
##   10 0 0 1 1 0 0 0 0  0
```

```
##   12 0 0 0 1 0 0 0 2  0
##   13 0 0 0 0 1 0 0 0  0
##   14 0 0 0 0 0 0 0 0  1
##   15 0 1 0 0 0 1 1 0  0
##   19 0 0 0 0 1 0 0 0  0
##   21 0 0 0 1 0 0 0 0  0
##   22 0 0 0 0 0 0 0 1  0
##   23 0 0 0 0 0 0 0 0  1
##   24 0 0 0 0 0 1 0 0  0
```

```
(james.out <- James.test(CflorusSecondary, TwoSided = TRUE))
```

```
## $U
## [1] 2.70893
##
## $p.val
## [1] 0.006750059
##
## $exp.table
##
##      0 1 2 3 4 5 6 7 10
##   1  8 0 0 0 0 0 0 0  0
##   2  1 5 0 0 0 0 0 0  0
##   3  1 1 1 0 0 0 0 0  0
##   4  1 4 0 0 0 0 0 0  0
##   5  0 3 3 0 0 0 0 0  0
##   6  0 0 1 1 0 0 0 0  0
##   7  0 2 0 0 0 0 0 0  0
##   8  0 1 0 1 0 0 0 0  0
##   9  0 0 0 0 2 0 1 1  0
##   10 0 0 1 1 0 0 0 0  0
##   12 0 0 0 1 0 0 0 2  0
##   13 0 0 0 0 1 0 0 0  0
##   14 0 0 0 0 0 0 0 0  1
##   15 0 1 0 0 0 1 1 0  0
##   19 0 0 0 0 1 0 0 0  0
##   21 0 0 0 1 0 0 0 0  0
##   22 0 0 0 0 0 0 0 1  0
##   23 0 0 0 0 0 0 0 0  1
##   24 0 0 0 0 0 1 0 0  0
```

From this we can see the test statistics for the Meelis and James' tests, as well as the corresponding p-values. The value of R and McCullagh's s^2 are included in the output from `Meelis.test`.

4.4 Bayesian analysis

The Bayesian analysis consists of two parts. The first is finding the posterior distributions of the parameters. The second optional stage is to go on to estimate the Bayes factors. These calculations require us to run an MCMC sampler, which can be computationally intensive depending on how long it is run for. The longer it is run, the more accurate the calculations are likely to be.

The calculations all require the specification of prior distributions. The family of distributions used for each parameter is hard coded into the package, but the user is free to choose the hyper-parameters that define the mean and variance of the distribution. The priors used are

$$\begin{aligned} p &\sim \text{Beta}(a_p, b_p) \\ \psi &\sim N(\mu, \sigma^2) \\ \lambda &\sim \text{Gamma}(\alpha, \beta) \\ d &\sim \text{Beta}(a_d, b_d). \end{aligned}$$

We specify all of these through a list.

```
hyper<-list()
```

The elements of the list must use the naming convention used below. A reasonable default choice of prior for p and ψ (see paper for the rationale) is to use $p \sim U[0,1]$ and $\psi \sim N(0,1)$, which we can set as follows:

```
hyper$a.p <-  1
hyper$b.p <-  1
hyper$mu.psi <- 0
hyper$sd.psi <- 1
```

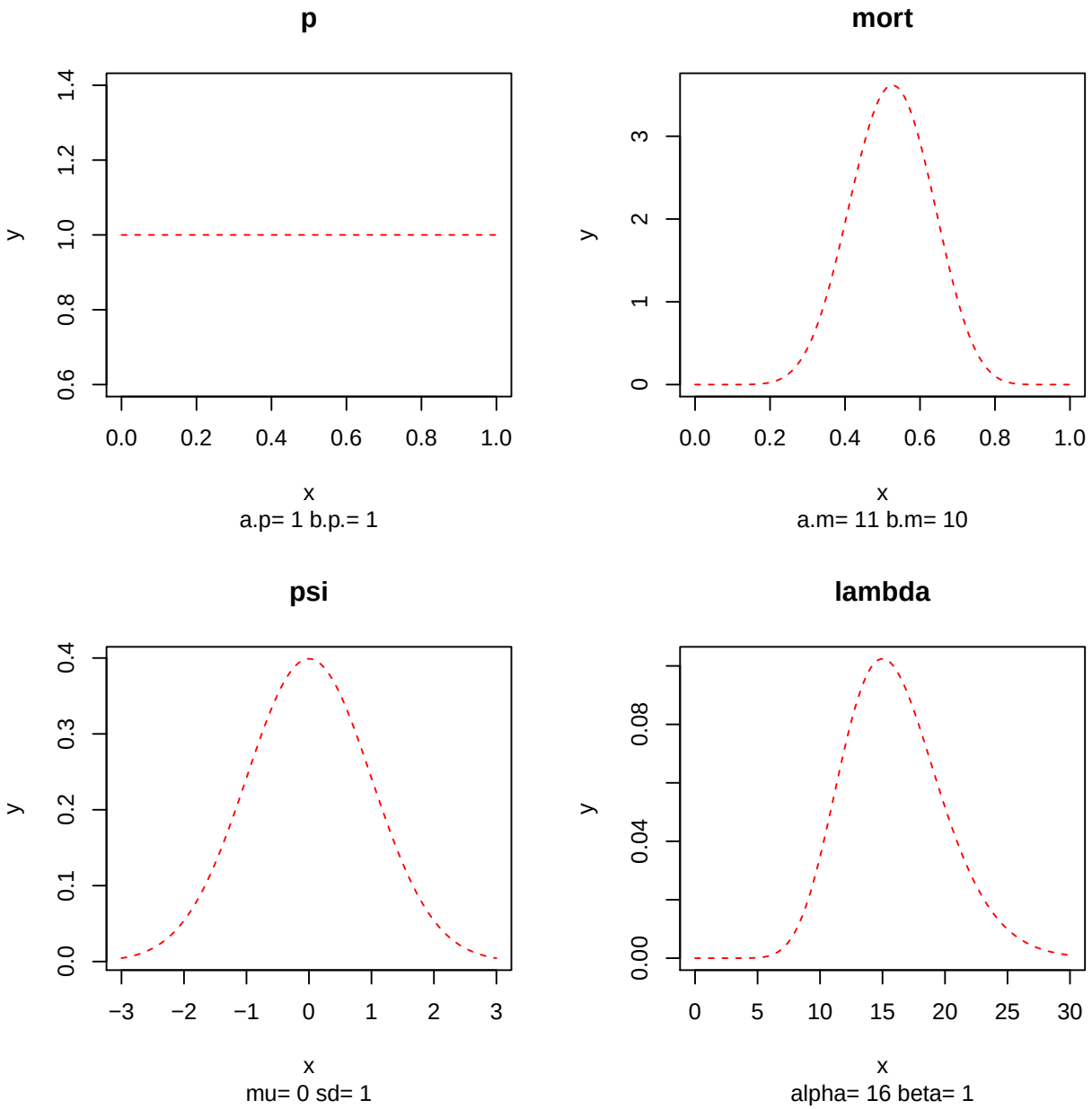
For C. florus, previous work has reported a mortality rate of 57% and an average clutch size of 7.4. Some experimentation with the values, and recalling that the mean of a Gamma(α, β) distribution is α/β and the variance is α/β^2, led us to use

```
hyper$a.m <- 11
hyper$b.m <-  10

hyper$alpha.lambda <- 16
hyper$beta.lambda  <- 1
```

It is a good idea to plot the prior distributions, to check that they agree with prior beliefs. This can be done as follows:

```
plot.prior(hyper=hyper, show=TRUE, family="multbinom")
```




```
## pdf
##   2
```

4.4.1 Posteriors

To calculate the posterior distribution, we have to run an MCMC sampler for a larger number of iterations. The longer we run the sampler, the better the posterior estimates will be. We would suggest a minimum of 10^5 iterations to get a reasonable estimate of the posteriors, and that 10^6 iterations should be more than sufficient. If Bayes factors are to be estimated, we would err towards the higher end of that range. The run time will depend upon both the number of MCMC iterations used, and the number of clutches in the dataset (as the MCMC algorithm samples the unobserved primary counts). To do 10^6 iterations with the C. florus dataset, you should expect to wait about an hour, depending on processor speed, for each set of MCMC results.

```
nbatch <- 10^6
```

Binomial Model The proceedure for fitting each of the three models (binomial, multiplicative binomial, and double binomial) is the same, and each can be done independently (on different cores if possible). To begin with, we choose a start point for the MCMC chain. The chains mix well and so a random value chosen from the prior works well here. It is necessary to label the parameters in the parameter matrix

```
b.theta0 <-c("lambda"=10, "p"=0.1, "mort"=0.5)
```

To run the code, we then just call the `MCMCWithinGibbs` function. Note that you can specify whether to keep the imputed missing primary values (the N and M values).

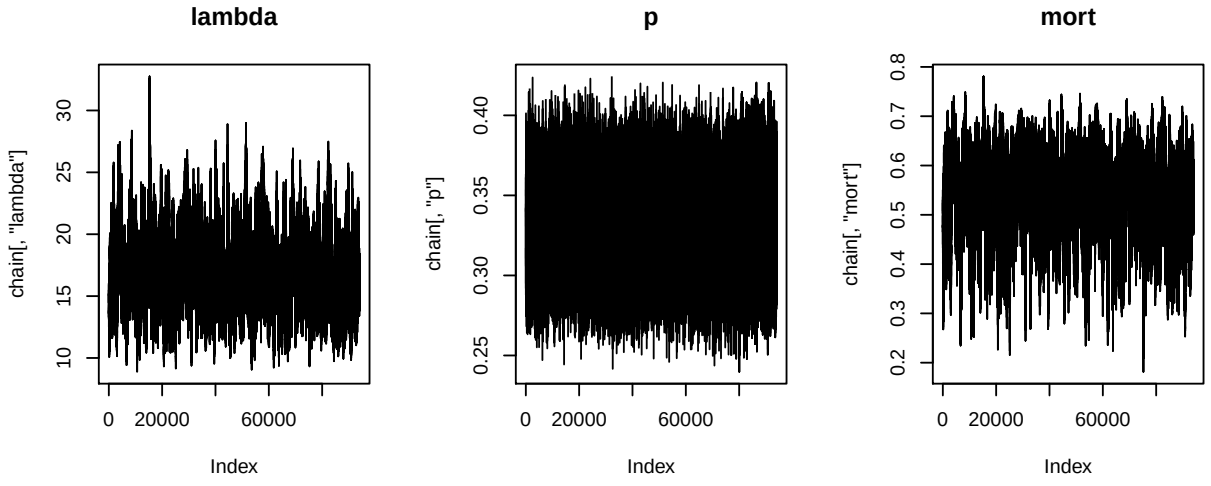
```
b.mcmc.out <- MCMCWithinGibbs(theta0=b.theta0,  data=CflorusSecondary, hyper=hyper,
                              nbatch=nbatch, family="binomial", keepNM=TRUE)
```

Finally, it can often be a good idea to thin the MCMC output (by only keeping every 10th value for example) and to discard an initial 'burn-in' period.

```
b.mcmc.out.t  <- ThinChain(b.mcmc.out, thinby=10, burnin=10^5)
```

The trace plots are useful to ensure that the chains have converged, and that they are mixing well.

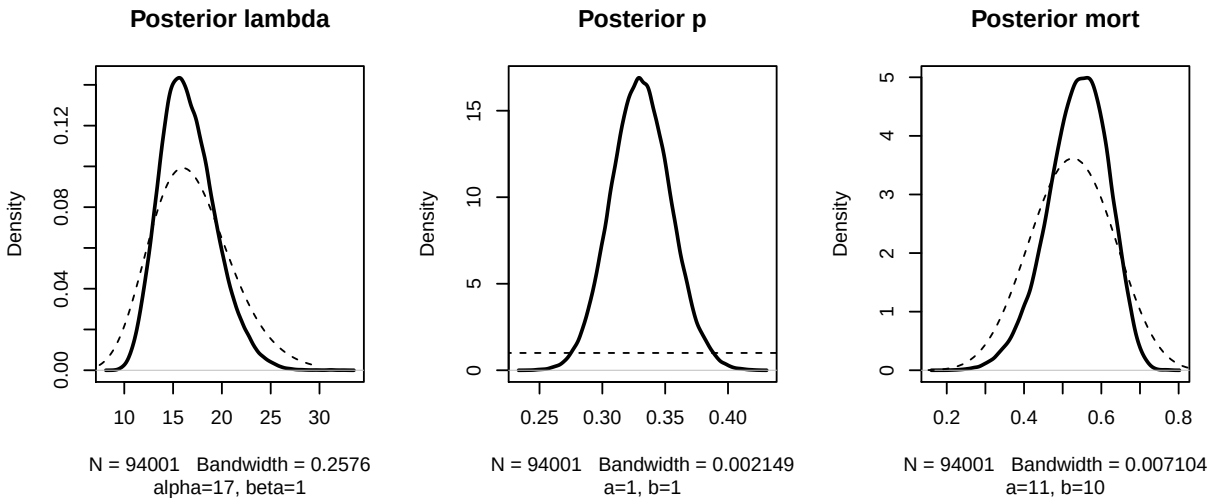
```
plot.trace(chain=b.mcmc.out.t$chain, show=T, family="binomial")
```




```
## pdf
##   2
```

These all look fine, and so we can plot the posteriors and draw conclusions:

```
plot.posterior(chain=b.mcmc.out.t$chain, hyper=hyper, show=T, family="binomial")
```




```
## pdf
##   2
```

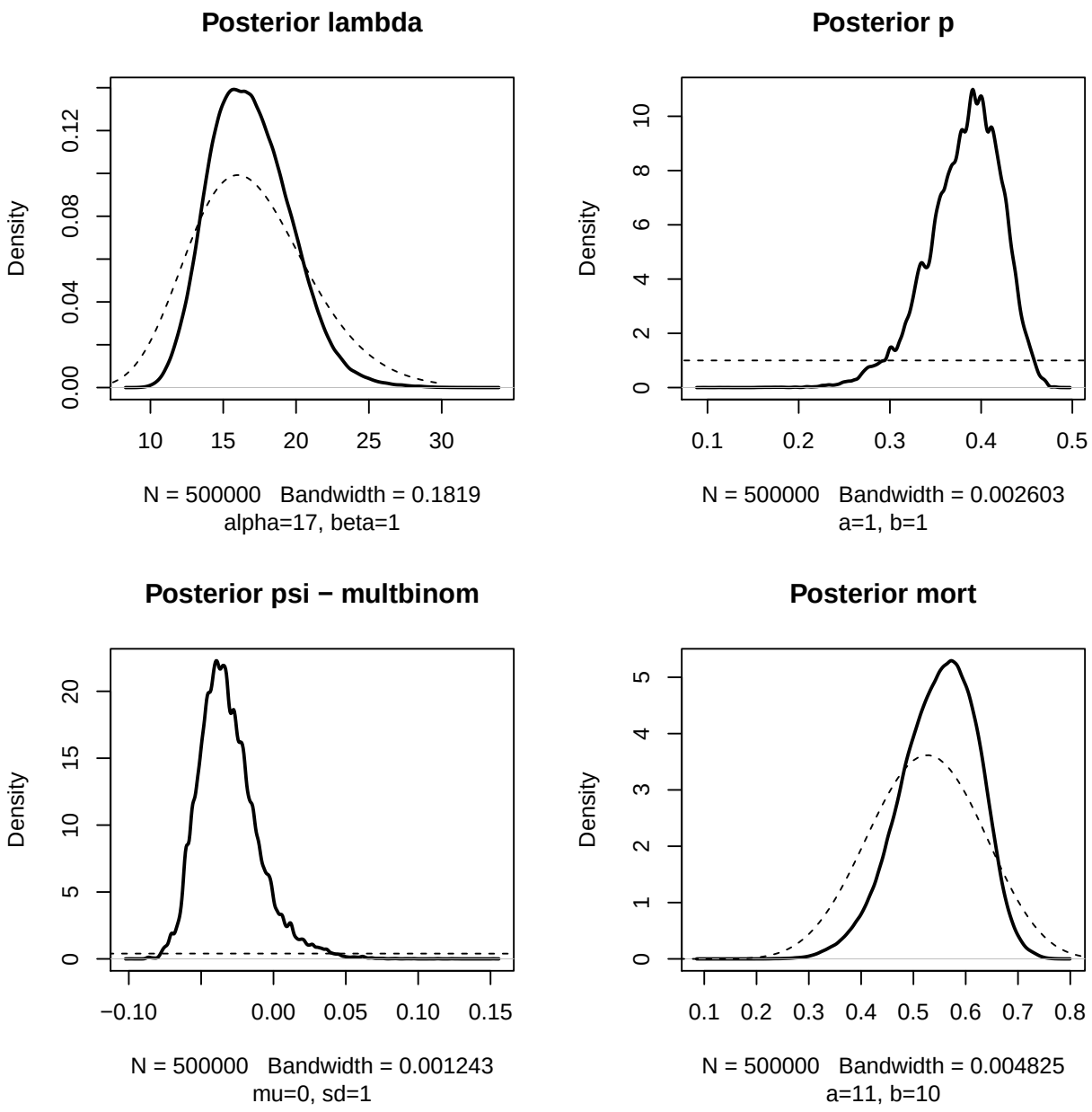
Multiplicative and Double Binomial Models The process for fitting the other models is very similar. However, now we are forced to use Metropolis-Hastings as well as a Gibbs sampler, and so we need to specify the Metropolis-Hastings random walk step size.

```
m.step.size<-c('p.logit'=0.3, 'psi'=0.2)
```

Note again that it is necessary to name the elements in this vector to avoid ambiguity. The rest of the code is the same as for the binomial model:

```
m.theta0 <-c('lambda'=10, 'p'=0.1,'psi'=0, 'mort'=0.1)
m.mcmc.out <- MCMCWithinGibbs( theta0=m.theta0,  data=GlegneriSecondary,
    hyper=hyper, nbatch=nbatch,
    family="multbinom", step.size=m.step.size,
    keepNM=TRUE)
m.mcmc.out.t  <- ThinChain(m.mcmc.out, thinby=10, burnin=10^5)
```

```
plot.posterior(chain=m.mcmc.out.t$chain, hyper=hyper, show=T,   family="multbinom")
```

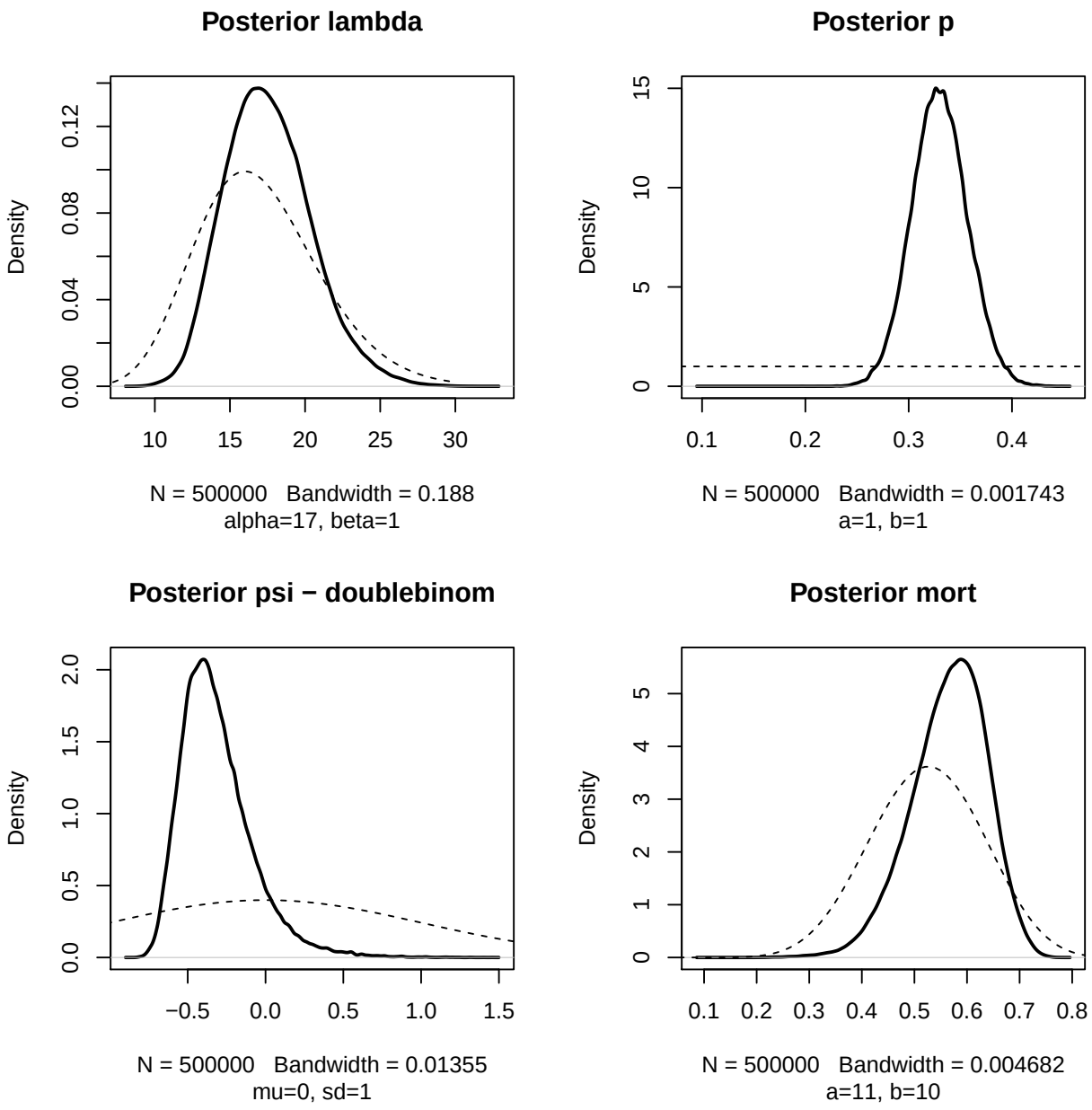



```
## pdf
##   2
```

```
#plot.trace(chain=m.mcmc.out.t$chain, show=T, family="multbinom")
```

```
d.step.size<-c('p.logit'=0.3, 'psi'=0.2)
d.theta0 <-c('lambda'=10, 'p'=0.1,'psi'=0, 'mort'=0.1)
d.mcmc.out <- MCMCWithinGibbs( theta0=d.theta0,  data=GlegneriSecondary,
  hyper=hyper, nbatch=nbatch,  family="doublebinom", step.size=d.step.size,
  keepNM=TRUE)
d.mcmc.out.t  <- ThinChain(d.mcmc.full, thinby=1, burnin=10^4)
```

```
plot.posterior(chain=d.mcmc.out.t$chain, hyper=hyper, show=TRUE,
      family="doublebinom")
```




```
## pdf
##   2
```

```
#plot.trace(chain=d.mcmc.out.t$chain, show=TRUE, family="doublebinom")
```

4.4.2 Bayes factors

The posterior distributions give most of the information about how much under-dispersion there is in the data. However, often we will want to also calculate the Bayes factor to see how strongly the data support one model over the others. To do this, we have to

run additional MCMC chains fixing some of the parameters (and so this step is also computationally costly).

```
b.log.evidence <- CalculateEvidence(mcmc.out=b.mcmc.out.t, data=GlegneriSecondary,
            hyper=hyper, family="binomial")
m.log.evidence <- CalculateEvidence(mcmc.out=m.mcmc.out.t, data=GlegneriSecondary,
            hyper=hyper, family="multbinom" nbatch=nbatch, step.size=m.step.size)
d.log.evidence <- CalculateEvidence(mcmc.out=d.mcmc.out.t, data=GlegneriSecondary,
             hyper=hyper, family="doublebinom", sd=FALSE, nbatch=nbatch,
             step.size=d.step.size)
```

Finally, we can put all the information together in a nice format as follows:

```
log.evidence <- c(b.log.evidence, m.log.evidence, d.log.evidence)
BF<-CalcBF(log.evidence)
chib.out <- list(BF=BF$BF, probH0 = BF$probH0 ,
 ProbPosPsi = c(
 "multbinom"=sum((m.mcmc.out.t$chain[,"psi"]>0))/length(m.mcmc.out.t$chain[,"psi"]),
 "doublebinom"=sum((d.mcmc.out.t$chain[,"psi"]>0))/length(d.mcmc.out.t$chain[,"psi"])
 log.BF=log(BF$BF), log.evidence=log.evidence,
 R= c("R"=meelis.out$R.av),
 s2 = c(meelis.out$s2),
 meelis = c("U"=meelis.out$U.av, "p"=meelis.out$p.av,  "conclusion"=ifelse(meelis.out
 james = c("U"=james.out$U, "p"=james.out$p.val,
 "conclusion" = ifelse(james.out$p.val<0.05, "RejectH0", "AcceptH0") ) )
print(chib.out)
```

```
## $BF
##        mb        db        dm
## 0.3620485 0.2687031 0.7421744
##
## $probH0
##    binomial   multbinom doublebinom
##   0.6132142   0.2220133   0.1647726
##
## $ProbPosPsi
##   multbinom doublebinom
##    0.075386    0.089588
##
## $log.BF
##        mb        db        dm
## -1.015977 -1.314148 -0.298171
##
## $log.evidence
## [1] -307.7081 -308.7241 -309.0222
##
## $R
```

```
##         R
## 0.7532786
##
## $s2
## [1] 1.181833
##
## $meelis
##                     U                     p           conclusion
## "-0.967286848797403"  "0.333400656210468"           "AcceptH0"
##
## $james
##                      U                      p            conclusion
##    "2.70892995669601" "0.00675005884011858"            "RejectH0"
```

From this we can read off the Bayes factors (which show that the binomial model is slightly favoured here), the posterior probabilities of each model, the posterior probability that ψ is positive (which is only 0.075 and 0.090 for the multiplicative and double binomial models respectively), as well as the other descriptive statistics previously used.

References

Bernardo, J. M., Smith, A. F., 2000. Bayesian theory. John Wiley & Sons.

Chib, S., 1995. Marginal likelihood from the Gibbs output. Journal of the American Statistical Association 90, 1313–1321.

Chib, S., Jeliazkov, I., 2001. Marginal likelihood from the metropolis-hastings output. Journal of the American Statistical Association 96, 270–281.

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6, 721–741.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E., 1953. Equation of State Calculations by Fast Computing Machines. The Journal of Chemical Physics 21, 1087–1092.

Nagelkerke, C. J., Sabelis, M. W., 1998. Precise control of sex allocation in pseudo-arrhenotokous phytoseiid mites. Journal of Evolutionary Biology 11, 649–684.