

Prior specification in Bayesian statistics: Three cautionary tales

Stefan Van Dongen*

Group of Evolutionary Biology, Department of Biology, University of Antwerp, Groenenborgerlaan 171, B-2020 Antwerp, Belgium

Received 4 October 2005; received in revised form 1 February 2006; accepted 6 February 2006

Available online 20 March 2006

Abstract

One of the most important differences between Bayesian and traditional techniques is that the former combines information available beforehand—captured in the prior distribution and reflecting the subjective state of belief before an experiment is carried out—and what the data teach us, as expressed in the likelihood function. Bayesian inference is based on the combination of prior and current information which is reflected in the posterior distribution. The fast growing implementation of Bayesian analysis techniques can be attributed to the development of fast computers and the availability of easy to use software. It has long been established that the specification of prior distributions should receive a lot of attention. Unfortunately, flat distributions are often (inappropriately) used in an automatic fashion in a wide range of types of models. We reiterate that the specification of the prior distribution should be done with great care and support this through three examples. Even in the absence of strong prior information, prior specification should be done at the appropriate scale of biological interest. This often requires incorporation of (weak) prior information based on common biological sense. Very weak and uninformative priors at one scale of the model may result in relatively strong priors at other levels affecting the posterior distribution. We present three different examples intuitively illustrating this phenomenon indicating that this bias can be substantial (especially in small samples) and is widely present. We argue that complete ignorance or absence of prior information may not exist. Because the central theme of the Bayesian paradigm is to combine prior information with current data, authors should be encouraged to publish their raw data such that every scientist is able to perform an analysis incorporating his/her own (subjective) prior distributions.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Bayesian; Prior specification; Subjective; Weak prior; Uninformative prior

1. Introduction

Bayesian statistics are becoming more and more established as a statistical tool in many areas of (bio)sciences (e.g., Congdon, 2003; Ellison, 2004). In methods based on a ‘classical’ statistical philosophy, maximum likelihood estimates and hypothesis tests based on p -values figure prominently. Bayesian methods differ from ‘classic’ frequentist and likelihood approaches in several respects (Carlin and Louis, 2000). Most importantly, Bayesian techniques allow to formally incorporate (subjective) information from outside the available dataset into an analysis in the form of a so-called prior distribution. Bayesian statistical conclusions about a parameter θ or unobserved data are made in terms of probability state-

ments. These statements are conditional on the observed data y , and can be written as $p(\theta|y)$. It is at the level of this conditioning that Bayesian methods depart from the classical approach which is based on the true value of θ (Gelman et al., 2000). Likelihood techniques estimate the probability distribution of the data y being the result of a sampling event from a particular parameter space, denoted as $p(y|\theta)$, and estimates values $\hat{\theta}$ through maximization of the likelihood function (i.e. the so-called maximum likelihood estimates). Accuracy of these maximum likelihood estimates are inferred from the curvature of the likelihood function.

Bayesian inference is based on the probability distributions of the model parameters and therefore does not depend on the assumptions of approximate normality and large sample properties as many likelihood methods do (Gelman et al., 2000). This probability distribution is captured by the so-called *posterior distribution* which takes

*Tel.: +32 3 265 3336; fax: +32 3 265 3474.

E-mail address: stefan.vandongen@ua.ac.be.

the following form:

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)}. \quad (1)$$

Thus, the posterior equals the prior ($p(\theta)$) times the likelihood ($p(y|\theta)$), divided by a normalizing constant ($p(y)$) which is independent of the model parameters. Therefore, the posterior distribution is proportional to the product of prior and likelihood: $p(\theta|y) \propto p(\theta) \times p(y|\theta)$. The posterior distribution represents the combined information of what was known about the parameters beforehand and what the data teach us. These simple expressions form the technical core of Bayesian inference. The primary task of any application is to develop the statistical model $p(\theta, y)$ and perform the required computations to summarize $p(\theta|y)$ (Gelman et al., 2000).

In the past, the general application of Bayesian techniques has been hampered by difficulties to evaluate the posterior analytically. Only in relatively few and simple applications a closed form expression can be obtained (i.e. with so-called conjugate priors), and these limitations have been an important source of criticism in the past. However, with the development of fast computers and readily available software, simulation techniques are more widely used leading to the implementation of Bayesian methods in many different fields (e.g. Congdon, 2003).

While Bayesian statistics has many advantages (e.g. Carlin and Louis, 2000; Gelman et al., 2000), it poses two important additional difficulties, namely: (i) convergence of the MCMC simulation (for which we refer to the review by Cowles and Carlin, 1996); and (ii) prior specification which will be the topic of this paper. Any Bayesian analysis requires both the development of a statistical model [$p(\theta, y)$] and the specification of the prior [$p(\theta)$]. The use of informative priors has been criticized for subjectivity, although it is important to note that there are also non-subjective justifications of the Bayesian approach (O'Hagan and Forster, 2004). In many applications, at least for some parameters very little is known and one would like the posterior distribution to be solely/mainly influenced by the data and not the prior. That is where non-informative priors are being used. In addition, any scientist may include informative priors differently based on subjective beliefs, which may be somewhat disturbing to some statisticians. This subjectivity has been the subject of many debates. Therefore, it has been suggested to describe a situation in which most can be learned from the data which does not incorporate a scientist's personal opinions. This could serve as a basis for comparisons. A scientific report could then display the dependence of the posterior distribution on the different choices of informative priors. There are various justifications and interpretations of non-informative priors. Over the years, invariance, maximum entropy and agreement with classic estimators have been proposed (e.g., Gelman, 2006). We adopt the view of Bernardo (1979) and consider non-informative priors as reference models to be used as standard of comparison against situations where

informative/subjective priors are used. We also view any non-informative prior as provisional (see also Gelman, 2006). After the model has been fit, one should look at both the posterior and the prior and check if they make sense. If not, this means that additional prior knowledge is available that has not been incorporated in the model and contradicts with the prior that has been used (Gelman, 2006). We define a reference prior as a situation where the inference shows a close correspondence with results from the likelihood. These results can then be contrasted with results with informative priors if subjectivity is desired. Throughout this paper we will examine the behavior of presumed uninformative priors with respect to the inference obtained from a classical likelihood approach.

A simple and appealing non-informative prior could be a uniform distribution between $-\infty$ and ∞ . However, as we will show, the uniform is not invariant to transformation and it does not integrate to 1 or a finite value (i.e. *improper* prior). Therefore, bounded uniform distributions, or normal distributions with very high variances have been suggested as *proper* alternatives (Carlin and Louis, 2000). More recently, one of the major recent criticisms on Bayesian analysis is the lack of robustness against alternative (uninformative) priors. It is therefore very important to devote a lot of attention to the specification of the priors instead of routinely assigning flat or wide priors to model parameters at a particular scale (Carlin and Louis, 2000). More specifically, in this paper we explore three examples to illustrate that flat and presumed non-informative priors, may be highly informative at another level of the analysis, having a strong impact on the posterior distribution. This property invalidates their use as reference priors sensu Bernardo (1979, see also above). We show that in our examples—and probably many other cases as well—complete ignorance and lack of any influence of the prior distribution on the posterior, does not exist. This issue is certainly not new and has been emphasized in many classic textbooks (e.g. Carlin and Louis, 2000, p. 29). Nevertheless, a short literature search revealed that often flat priors are used in an automated way (see below for specific references), while its impact on the posterior is ignored.

We chose three examples to illustrate that flat priors are often not uninformative. In a first example, we study the effect of prior specification on the posterior of the variance component of a simple hierarchical model. This example has been treated recently in great detail by Gelman, 2006, but we incorporate it here as well for three reasons: (i) it is a well known and often used example; (ii) hierarchical models are commonly used in biology; and (iii) we also show that misspecification of the prior for the variance also affects the fixed effects part of the model, a fact that has received much less attention in the literature. In a second example we apply a logistic regression model to a small dataset, and show that nearly flat priors are inappropriate, in spite of the fact that they are repeatedly used (e.g., recommendations on p. 388 in Gelman et al., 2000; Mwalili

et al., 2005). As a third example, we investigate the estimation of the mean and variance of a γ -distribution and show that even in relatively large samples ($N \geq 100$), misspecification of priors (as in Van Dongen et al., 2005) may introduce bias.

The aim of this paper is to create renewed awareness of this issue, especially to the non-specialist in the field, and to illustrate the importance of choosing prior distributions at the appropriate scale. We also show that appropriate priors often require some weak biological information on the parameters of interest—which is often known anyway. At this point it is important to note that we neither aim at presenting a full analysis of the example dataset nor want to draw any conclusions about the biology of the data. Instead, we will provide guidance for future users of Bayesian techniques to evaluate the choice of the priors of their analyses. Such choices inherently embody some degree of subjectivity and should often be accompanied by sensitivity analyses.

2. Bayesian analyses using MCMC

All Bayesian analyses presented in this paper were performed in WINBUGS 1.4 (freely available at <http://www.mrc-bsu.cam.ac.uk/bugs/>), using Monte Carlo Markov chain (MCMC) simulations. In each analysis, three independent chains were run for 5000 iterations after excluding a burn-in of 1000 iterations. Convergence was assessed by visual inspection of the chains and by the Gelman and Rubin shrink factor (Gelman et al., 2000). No convergence problems emerged in any of the analyses (data not shown).

To summarize and describe the posterior distribution we present throughout this paper the mean and standard deviation, the median and a 95% credibility interval of the posterior next to a graphical representation for the parameters of interest. It is important to note that, especially in cases where the posterior is skewed, the mean and/or median of the posterior distribution should not necessarily correspond to the maximum likelihood estimator even when the prior did not influence the analysis. Bayesian inference are based on summaries of the entire probability distribution. Maximum likelihood techniques identify the most likely value of a parameter, while with Bayesian methods the posterior is described by a set of characteristics.

3. Three examples

3.1. A “Classic” hierarchical model

We analyse data from 35 house fly wing lengths originating from seven groups ($i = 1 \dots 7$) in a hierarchical model with group as random effect (data in Sokal and Rohlf, 2000, p. 181). The model has three parameters, an intercept β_0 , the between group variance $\sigma_{betw.}^2$ and the residual variance $\sigma_{with.}^2$:

$$y_{ij} \sim N(\mu_i, \sigma_{with.}^2),$$

$$\mu_i = \beta_0 + \beta_1[i],$$

$$\beta_1[i] \sim N(0, \sigma_{betw.}^2). \quad (2)$$

Maximum likelihood estimates as obtained by the Newton–Raphson algorithm in SAS are provided in Table 1. Point estimates correspond to the maximum of the likelihood function, while the standard deviation of this estimate and 95% confidence intervals are based on the curvature of the likelihood function.

In a Bayesian setting, prior distributions need to be specified. Suppose we have no other data to guide us, and want to use uninformative priors to construct a reference model. We will focus on different alternative priors for the variance components and assume a flat prior for $\beta_0 \sim N(0, 10^6)$ (i.e. a normal distribution with zero mean) throughout all analyses. Three different alternative priors for the variance (or standard deviation) in a hierarchical model have often been used in the literature (see also Gelman, 2006): an inverse γ -distribution with small values for shape and rate parameter, a uniform distribution on the variance and a uniform on the standard deviation. We have implemented each of these and for the inverse γ -distribution we used two alternative values for shape and rate, namely 0.001 and 0.1 (Table 1). Fig. 1 shows a graphic representation of prior and posterior distributions. Although each choice should have little effect on the posterior distribution—as an uninformative prior should add nothing to the analysis and thus not influence the posterior—the results are markedly different especially with respect to the estimation of $\sigma_{betw.}^{(2)}$. Relative to the outcome of the maximum likelihood approach, Bayesian inference using the inverse γ -distributions as prior resulted in a lower estimate of between-group variability, while uniform priors, resulted in higher estimates of $\sigma_{betw.}^{(2)}$. In addition, the uniform prior on the standard deviations caused posteriors to exhibit a heavy right tail, by putting some belief on ‘unreasonably high’ variances or standard deviations compared to the likelihood inference (see also Fig. 1). A uniform prior at the level of the variance was not uninformative at the level of the standard deviation (Fig. 1), leading to even higher estimates (Table 1). In conclusion, none of the presumed uninformative priors appeared truly uninformative (see also Gelman, 2006). The priors introduced discordance between Bayesian and likelihood inference and/or resulted in heavy tailed posterior distributions of the between group variability. This difference between both approaches suggests that additional prior information (or rather the lack of it) is available and not captured in the priors explored so far. None of these alternatives appears suitable as reference model. Gelman (2006) also noted that if one would plot the priors on a $\log(\sigma)$ scale, the inverse γ -distribution would look much flatter. Yet, one should not consider it as uninformative since the resulting posteriors remain highly sensitive to the choice of the values of the shape and rate parameters (Table 1 and Gelman, 2006).

Table 1

Parameter estimates and summary of posterior distributions for the hierarchical model analysing house fly wing lengths

Parameter	Point estimate (sd)	95% confidence interval		
<i>Maximum likelihood estimates</i>				
Intercept	45.3 (0.78)	43.4–47.2		
$\sigma^2_{betw.}$	1.03 (2.59)	0–14.4		
$\sigma^2_{with.}$	16.0 (4.28)	10.2–27.3		
Bayesian analyses using different priors for $\sigma^2_{betw.}$				
Parameter	Mean	sd	Median	95% credibility interval
<i>Inverse gamma with shape and rate = 0.001</i>				
Intercept	45.3	0.81	45.3	43.8–46.9
$\sigma^2_{betw.}$	1.12	2.96	0.14	0.001–18.3
$\sigma^2_{with.}$	17.5	4.60	16.8	10.6–28.4
<i>Inverse gamma with shape and rate = 0.1</i>				
Intercept	45.3	0.89	45.3	43.6–47.1
$\sigma^2_{betw.}$	2.03	3.49	0.84	0.06–10.8
$\sigma^2_{with.}$	17.0	4.51	16.4	10.2–27.8
<i>Uniform prior of σ</i>				
Intercept	45.3	1.09	45.3	43.2–47.5
$\sigma^2_{betw.}$	4.67	9.02	2.07	0.007–24.7
$\sigma^2_{with.}$	17.6	4.96	16.72	10.5–29.6
<i>Uniform prior of σ^2</i>				
Intercept	45.3	1.36	45.3	42.6–48.0
$\sigma^2_{betw.}$	10.2	20.5	5.03	0.16–50.7
$\sigma^2_{with.}$	18.1	5.17	17.2	10.5–30.6
<i>Half-Cauchy with prior scale = 5</i>				
Intercept	45.3	0.96	45.3	43.4–47.3
$\sigma^2_{betw.}$	3.22	5.17	1.59	0.005–16.2
$\sigma^2_{with.}$	16.9	4.55	16.2	10.2–27.7

An often overlooked effect of subjective prior specification for the variance is that the mean structure of the model is affected as well. In this example, the mean structure of the model is very simple, the intercept estimates the mean wing length. The precision of the estimation of this intercept was affected by the choice of the prior at the level of the random effect as well. When inverse γ -distributions were applied as prior for the $\sigma_{betw.}^2$, the posterior distribution of the intercept became slightly more narrow compared to the maximum likelihood outcome. Uniform priors at both $\sigma_{betw.}$ and $\sigma_{with.}^2$ on the contrary, resulted in wider 95% CI of the intercept (Table 1). Although quantitatively small, the importance of this effect will depend on the context of the analyses.

The above comparisons of priors which have been used as uninformative or objective priors in many studies, shows that they introduce some subjectivity in the outcomes. Unless such subjectivity is desired, this should be avoided. To overcome these awkward behaviors, Gelman (2006) introduced a half-Cauchy density as reference prior for the standard deviation in a hierarchical model. This prior introduces some very weak information which can be based on common sense and general biological principles. This

weak prior information is captured by the so-called *prior scale* of the half-Cauchy, which we have set equal to 5 in our example. Other alternatives might have been more appropriate here, but as indicated before, we do not intend to provide final analyses, but merely want to illustrate the behaviors of different prior distributions. In order to implement the half-Cauchy prior distribution of the standard deviation of the hierarchical model, some reformulations are required. The hierarchical model given above (Eq. (2)) can be rewritten as

$$y_{ij} \sim N(\mu_i, \sigma_{with.}^2),$$

$$\mu_i = \beta_0 + \zeta \times \eta[i],$$

$$\eta_1[i] \sim N(0, \sigma_\eta^2). \quad (3)$$

This over-parametrization is known to improve MCMC convergence (Gelman, 2006). The parameters $\beta_1[j]$ and the standard deviation $\sigma_{betw.}$ in Eq. (2) correspond to the products $\zeta \times \eta[i]$ and to $|\zeta| \times \sigma_\eta$ respectively, in Eq. (3). Since a Cauchy distribution can be written as the ratio of a normal and the square root of a chi-square distribution with one degree of freedom, a half-Cauchy prior for the

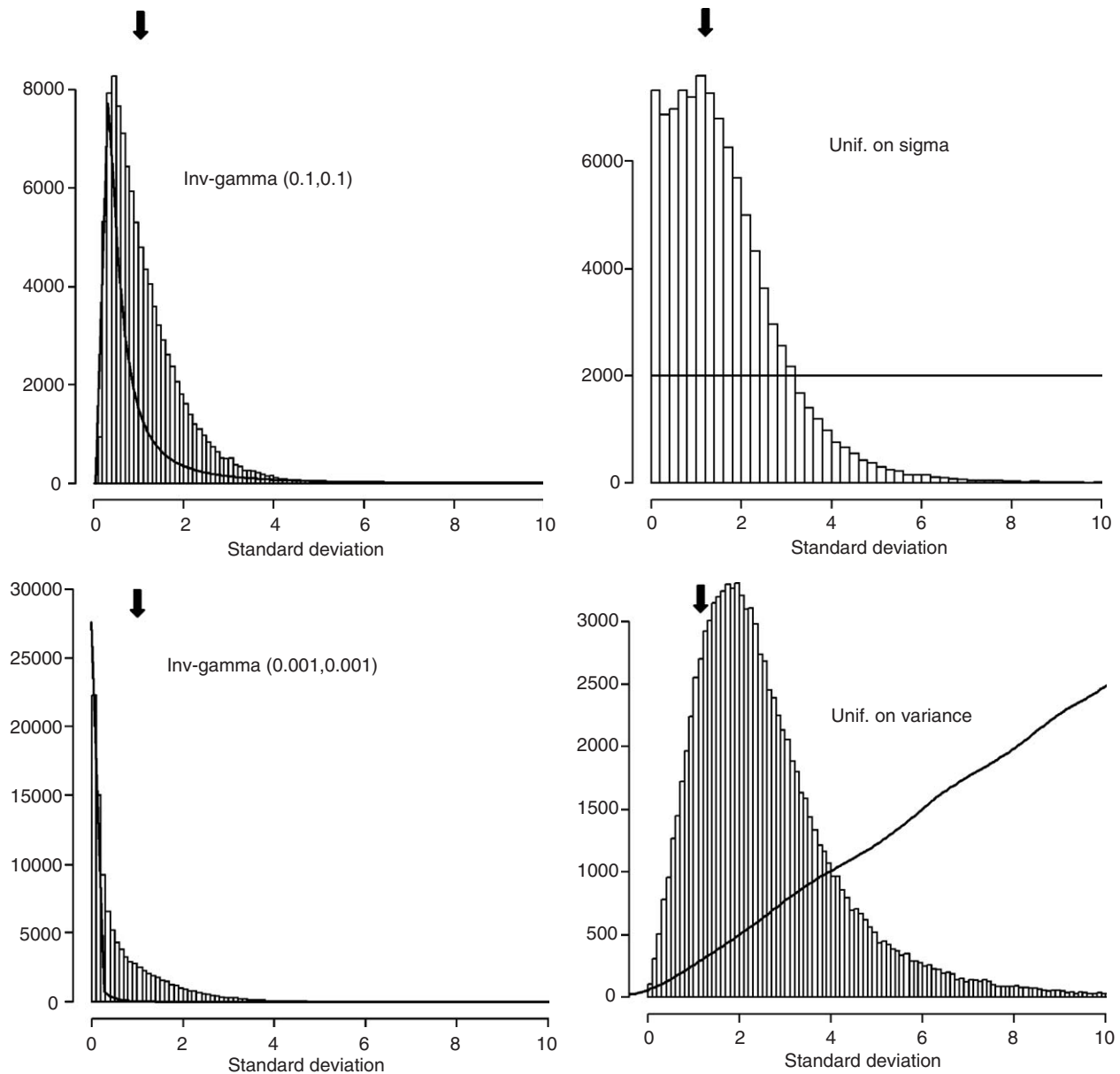


Fig. 1. Prior (line) and posterior distributions (bars) of the standard deviation of between-group variation in house fly wing lengths assuming different prior distributions. Maximum likelihood estimates of the standard deviation of the hierarchical model are indicated by a black arrow.

standard deviation $\sigma_{betw.}$ can be obtained as follows. Use as prior for $\zeta \sim N(0, \text{prior scale})$ and for $(1/\sigma_\eta^2) \sim \chi_1^2$. Setting $\sigma_{betw.} = |\zeta| \times \sqrt{\sigma_\eta^2}$ yields a half-Cauchy prior. Further details of this model are provided by Gelman (2006).

As the half-Cauchy prior showed (slightly) reduced prior belief in high standard deviations (>6), the posterior did not exhibit the heavy tails towards high values (Fig. 2) as was observed with the uniform priors. Also at the level of the mean structure (the intercept of the model in this simple example), the posterior distribution showed results comparable to the maximum likelihood results (Table 1).

In conclusion, weakly informative priors reflecting a slightly lower belief in extremely high between-group

variation resulted in much improved behavior of the posterior distribution of all parameters at the different hierarchical levels of the model, in the sense that inference corresponded more closely to the likelihood results. This emphasizes the importance of correctly specifying the random effects structure of a hierarchical model in order to make valid inferences about the mean structure of the model (Verbeke and Molenberghs, 2000). Our analyses show that inappropriate priors at the random effects level affects inferences at the mean structure of the model. So even when inferences for the random effects are not of primary interest, priors at this level of the model should be specified carefully. Values of the prior scale of the half-Cauchy prior will determine the prior belief in higher

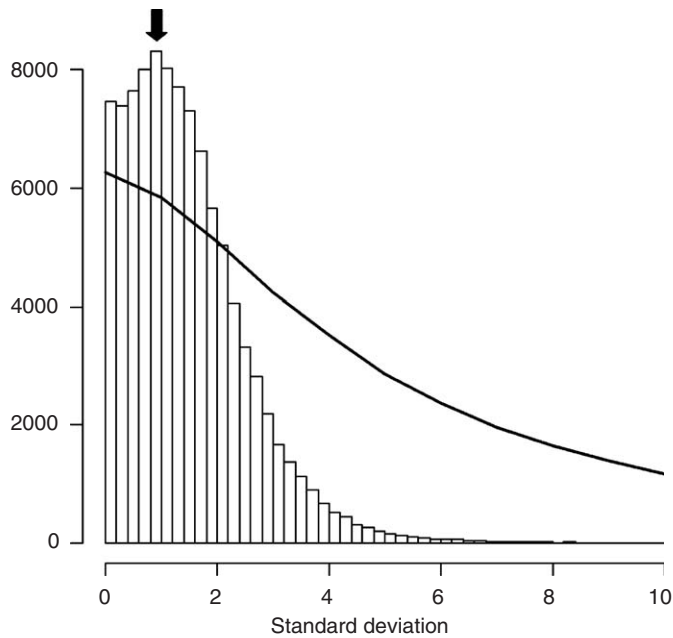


Fig. 2. Prior (line) and posterior distribution (bars) of the standard deviation of between group variation in house fly wings using the half-Cauchy prior with prior scale equal to 5. Maximum likelihood estimate is indicated by a black arrow.

standard deviations, and common biological knowledge of the problem under study should generally allow to determine an adequate value. Additional sensitivity analyses could be carried out to investigate the importance of the value of the prior scale on the posterior distribution (Gelman, 2006).

3.2. Logistic regression

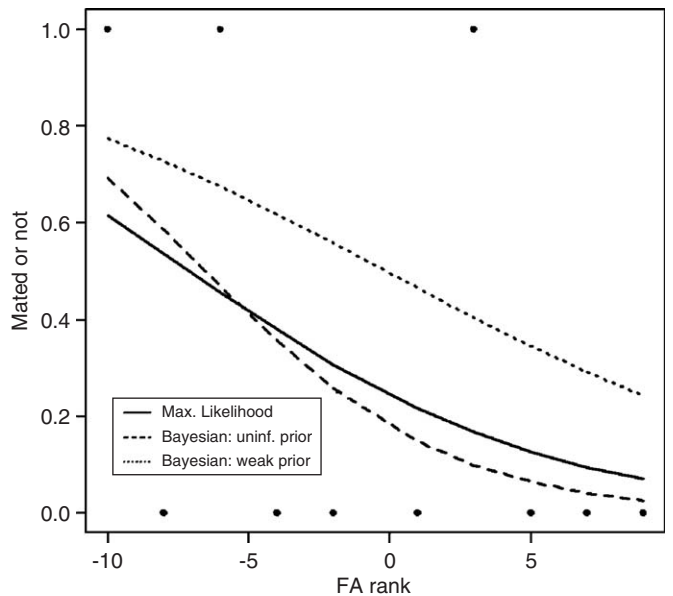
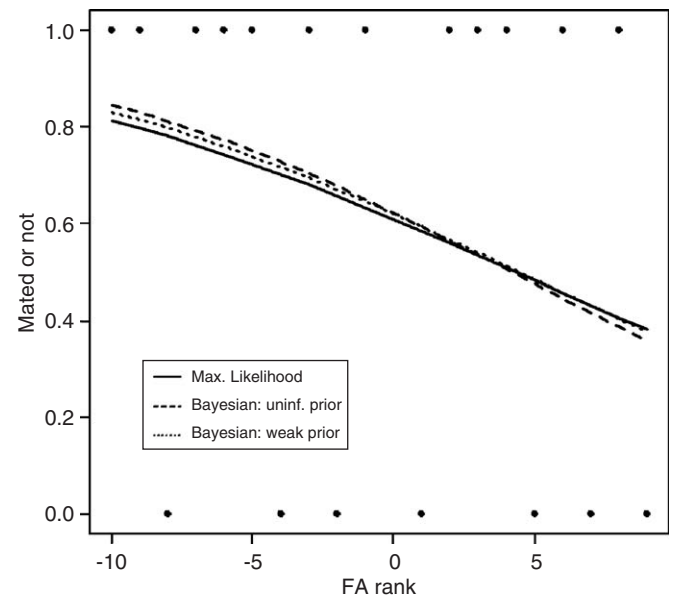
We fit a simple logistic regression model to data from an experiment where male winter moth (*Operophtera brumata* L.) mating success (mated vs. not mated) is analysed in relation to male tarsus asymmetry (Van Dongen unpubl. data). For simplicity, males were ranked from 1 to 20 for their level of asymmetry and then ranks were centered around zero. The intercept of a linear model can then be interpreted as a property of the median individual. The data are provided in Fig. 3. In a second analysis, the number of individuals was halved by using only every second observation starting from the individual with the lowest degree of fluctuating asymmetry (FA) (rank = -10) (Fig. 3). The statistical model can be written as

$$Y[i] \sim \text{bin}(\text{prop}[i]),$$

$$\text{logit}(\text{prop}[i]) = \beta_0 + \beta_1 \times \text{FA}[i]. \quad (4)$$

Suppose the aim of this logistic regression model is to predict the mating probability for an individual with FA rank equal to 8, a relatively highly asymmetrical male winter moth (i.e. 10th percentile).

Flat priors for β_0 and β_1 could be incorporated in the analysis as normal distributions with mean equal to zero



¹Dataset of N=20:

rank:	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1
mating success:	1	1	0	1	1	1	0	1	0	1
rank:	1	2	3	4	5	6	7	7	9	10
mating success:	0	1	1	1	0	1	0	1	0	1

Fig. 3. Estimated logistic regression lines analysing male winter moth mating success in relation to fluctuating asymmetry, using maximum likelihood and Bayesian techniques for both datasets (top: $N = 20^1$; bottom: $N = 10$).

and variance equal to 1000. Posterior distributions of β_0 and β_1 and the predicted mating success differed from the maximum likelihood estimates (Table 2), and the estimated logistic regression lines were steeper (Fig. 3). The effect were strongest for the smallest sample (Table 2, Fig. 3). This suggest that the prior distribution did affect the outcome although it was assumed to be uninformative because of its flat distribution. The informativeness of this

Table 2

Summary of the posterior distributions for the logistic regression model analysing mating success in relation to male fluctuating asymmetry

	Sample size = 20				Sample size = 10			
Parameter	Point estimate (sd)		95% C.I.		Point estimate (sd)		95% C.I.	
<i>Maximum likelihood estimates</i>								
Intercept	0.45 (0.48)		−0.50–1.39		−1.12 (0.85)		−2.80–0.54	
Slope	−0.10 (0.08)		−0.26–0.05		−0.16 (0.14)		−0.44–0.11	
Predicted value (10th percentile)	0.41		0.14–0.75		0.08		0.003–0.68	
Parameter	Mean	sd	Median	95% C.I.	Mean	sd	Median	95% C.I.
<i>Bayesian estimates using uninformative priors at the β level</i>								
Intercept	0.50	0.51	0.49	−0.46–1.54	−1.52	1.05	−1.42	−3.92–0.23
Slope	−0.12	0.09	−0.12	−0.30–0.04	−0.23	0.17	−0.22	−0.62–0.06
Predicted value (10th percentile)	0.40	0.17	0.39	0.11–0.76	0.10	0.13	0.04	0.0002–0.48
<i>Bayesian estimates using weakly informative priors at the β level</i>								
Intercept	0.49	0.35	0.50	−0.20–1.19	−0.02	0.41	−0.01	−0.83–0.78
Slope	−0.11	0.08	−0.11	−0.28–0.04	−0.12	0.11	−0.12	−0.36–0.09
Predicted value (10th percentile)	0.42	0.16	0.41	0.14–0.73	0.30	0.18	0.27	0.04–0.70

Bayesian predictions were performed during the MCMC and thus are averaged with respect to the posterior distribution.

choice of prior becomes apparent at the level of predicting probabilities of mating success. A flat prior at the level of the regression coefficients β_0 and β_1 did not result in a flat prior at the level of the prediction interval but rather a bimodal one for the prediction of mating success (Fig. 4). This is the result of the fact that the prior for the regression coefficients occurs at the logit scale, while prediction is done at the original binary scale. Taking a very wide prior for β_1 looks uninformative at the logit scale, yet reflects a strong prior belief in either a positive or negative association between mating success and FA-rank at the binary scale, hereby influencing the posterior distribution. The tails of the priors at the logit scale reflect an unrealistically strong prior belief in an association. Going from the logit to the binary scale increases the weight of prior belief in the tails of the distribution. Therefore, priors for the regression parameters should be narrowed to obtain a good reference model.

Suppose we have some prior knowledge about the mating success in this experiment. During several previous experiments we have noticed that on average about 50–70% of the males mated successfully. If we are willing to believe that the current experiment will result in an average mating rate comparable to the previous ones, we could use as prior for β_0 a normal distribution with mean equal to 0.5 (since $e^{0.5}/1 + e^{0.5} = 0.62$) and variance equal to 0.25, reflecting a relatively strong belief in an overall mating success between 50 and 70%. As prior for β_1 we suggest a normal distribution with zero mean and variance equal to 0.1, because both theoretical arguments and available data from the literature support relatively weak associations. The use of these informative priors resulted in: (1) much flatter prior distributions at the prediction levels (Fig. 4), and (2) parameter estimates that more closely reflected the maximum likelihood estimates for the

full dataset with 20 individuals (Table 2). The prior did not appear to affect the outcome to a large extent, and therefore appeared less informative compared to the previous approach in spite of including some subjective prior beliefs about the regression parameters. Nevertheless, the prior distribution of the predicted mating success for the relatively asymmetric male was somewhat skewed (Fig. 4), due to our relatively strong prior belief in relatively high average mating success. We refer to [Chen et al. \(1999\)](#) for a more elaborate treatment of informative priors in logistic regression.

For the reduced dataset of size 10, the use of the weakly informative prior had a strong impact on the outcome (Table 2, Figs. 3 and 4), yielding posteriors similar to those from the full dataset. Obviously, the validity of this result depended largely on the relatively informative prior distribution of β_0 , highlighting both the strength and subjectivity of Bayesian analyses.

In conclusion, we have shown in this example that flat and presumed uninformative priors at one scale of the model—i.e., the slope(s) of the logistic regression model at the logit scale—are not uninformative at other scales—i.e. prediction of new observations at the binary scale—hereby introducing bias. This example stresses the need of examining prior as well as posterior distributions at the different levels/scales of the model.

3.3. Estimating mean and variance of a γ -distribution

In this final example, we aim at estimating the mean, variance and coefficient of variation (CV) of a γ -distribution. This γ -distribution reflects between-individual variation in developmental instability. The statistical model is complex and it is beyond the scope of this paper to explain it in detail. We refer to [Van Dongen and Lens \(2001\)](#) and

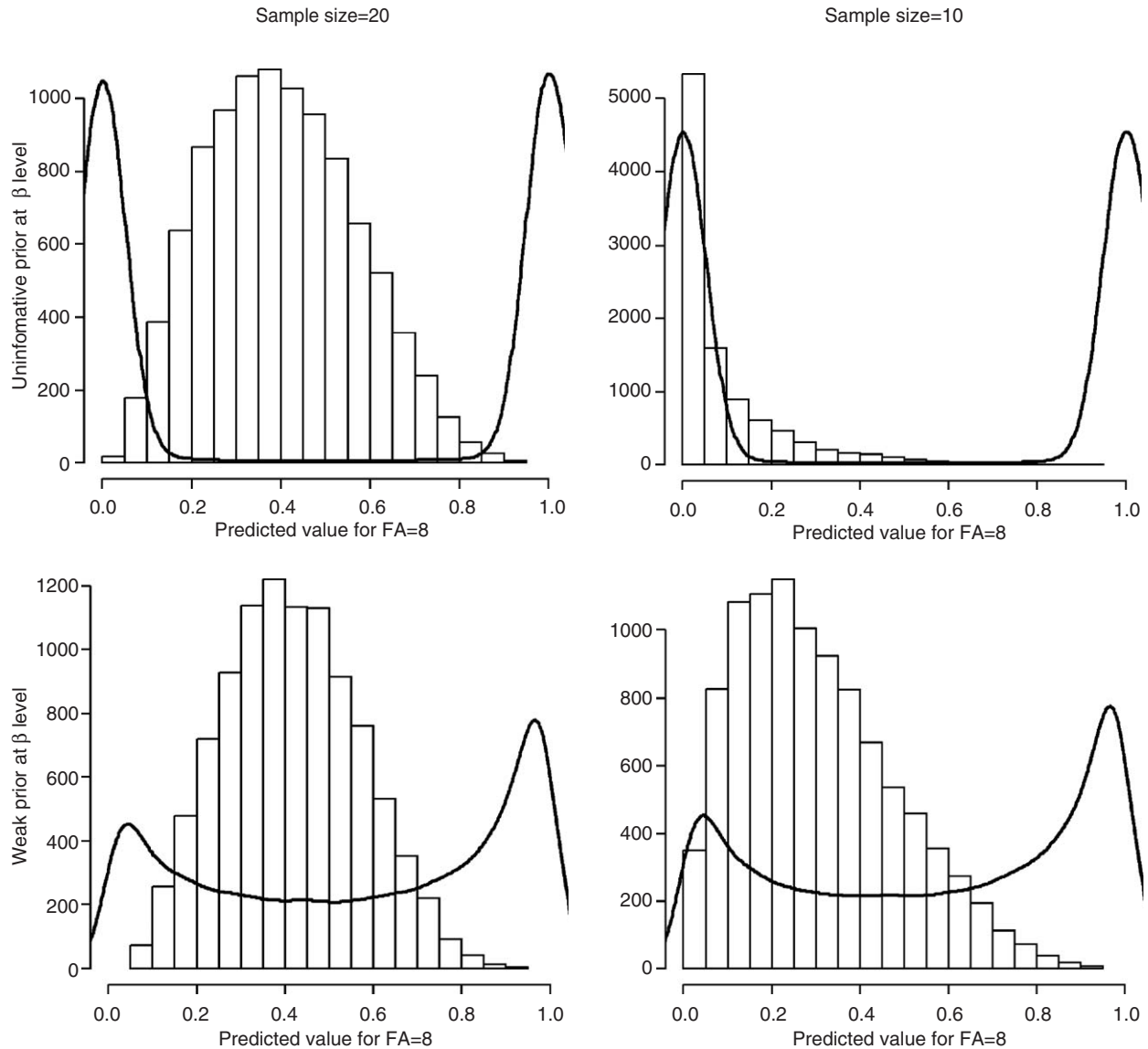


Fig. 4. Prior (line) and posterior distributions (bars) of the predicted mating probability for an individual with degree of asymmetry equal to the 10th percentile.

Van Dongen et al. (2005) for a full development, details and rationale. For the sake of the arguments here, we restrict our description of the statistical model to the estimation of the γ -distribution, being determined by two parameters: shape (α) and rate (β). The shape and rate parameters of the γ -distribution relate to the mean and variance of this distribution as follows:

$$\mu_{\gamma\text{-distribution}} = \frac{\alpha}{\beta},$$

$$\sigma_{\gamma\text{-distribution}}^2 = \frac{\alpha}{\beta^2}. \quad (5)$$

In this example, we analysed part of a dataset ($N = 100$ individuals) from great tit tarsus measurements on the left and right leg (details in Van Dongen et al., 2005). If we

suppose that there is no prior information available, a uniform prior between 0 and 10 000 for both shape and rate could be considered to be uninformative. However, while flat at the level of shape and rate parameters, prior distributions at the level of the mean, variance and CV however, were not at all uninformative (Fig. 5 top row). The prior distribution of the CV was sharply peaked around small values and the prior for the mean of the γ -distribution showed a reduced belief in smaller values (which is in conflict with biological knowledge, see below). It, therefore, seems advisable to specify the priors at the level of the mean, variance and/or CV of the γ -distribution because these are the levels of the model that are of biological interest and can be interpreted accordingly. It is easy to show that the shape and rate parameters of the γ -distribution can be expressed in terms of the mean and

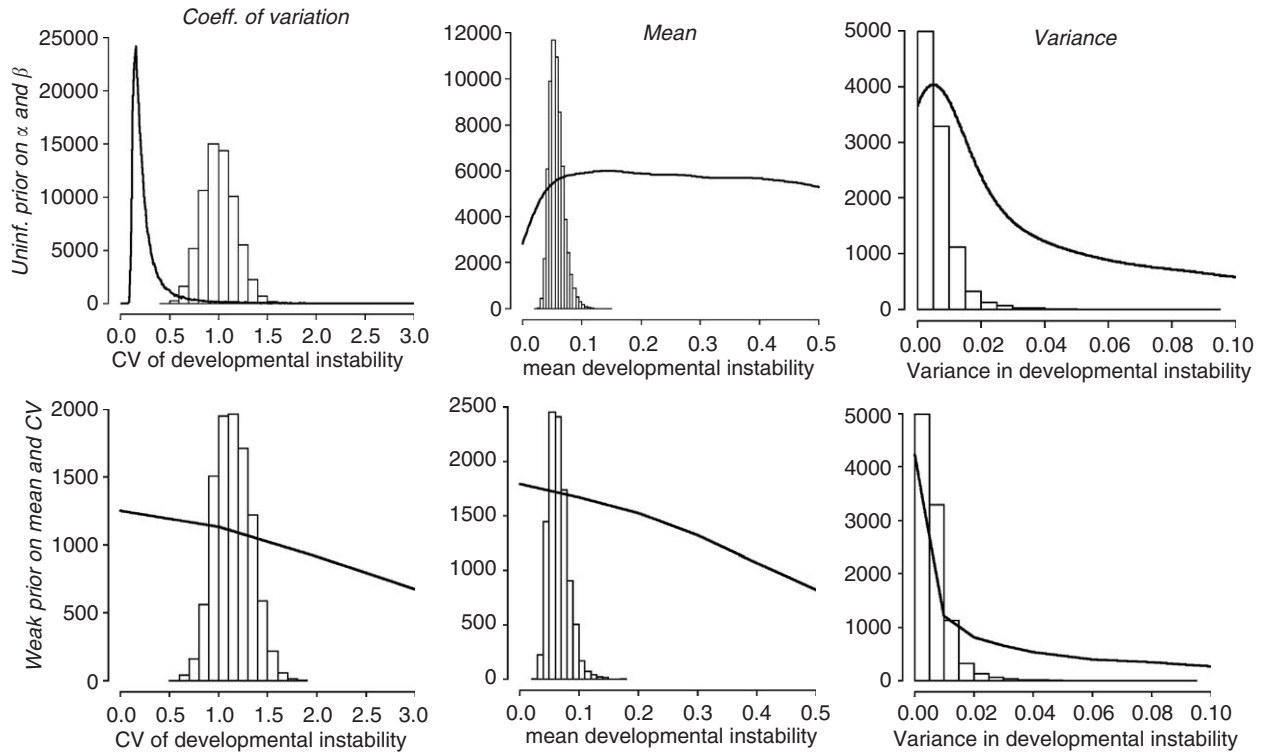


Fig. 5. Prior (line) and posterior distributions (bars) of CV, mean and variance of an estimated γ -distribution using two different priors (see text for details).

variance of the distribution:

$$\alpha = \frac{\mu_{\gamma\text{-distribution}}^2}{\sigma_{\gamma\text{-distribution}}^2},$$

$$\beta = \frac{\mu_{\gamma\text{-distribution}}}{\sigma_{\gamma\text{-distribution}}^2}.$$

So prior distributions at the level of the mean, CV and/or variance of the γ -distribution can be easily transformed into the corresponding priors for the shape and rate parameters. We therefore, applied weak informative information for the mean and CV. From the literature it is well established that levels of asymmetry are relatively small, often in the order of magnitude of 1–2% of the total trait size. Therefore, we used a half-normal distribution for the mean (with mean equal to 0.36 and variance equal to 0.07) level of developmental instability (Fig. 5). With an average tarsus length of 19 mm in this dataset, this prior distribution reflects a relatively strong belief in any degree of developmental instability below 10% of the trait size and reduced beliefs in higher levels.

A CV of between-individual variation in developmental instability has never been observed to be larger than 300%, so we also specified the prior for CV as a half-normal (with mean equal to 2.5 (i.e. 250%) and variance equal to 3.6) to incorporate our reduced belief in extremely and unrealistically large values of the CV of the γ -distribution (Fig. 5). These, only very weak informative priors at the level of the

(6)

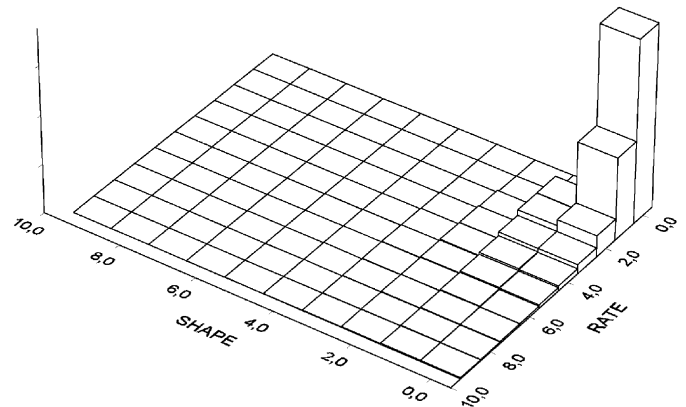


Fig. 6. Joint prior distribution of shape and rate parameters of the γ -distribution after imposing weakly informative priors on the mean and CV (see text and Fig. 5 for details).

mean and CV of the γ -distribution resulted in slightly informative priors at the level of the variance (Fig. 5), and relatively strong priors for the shape and rate parameters (Fig. 6). The resulting posterior distributions also showed considerable shifts compared to the flat and presumed uninformative priors at the level of the shape and rate parameters used before (Table 3). Interestingly, the posterior mean and median of the variance was higher when using weak priors on the mean and CV (Table 3), in spite of the fact that the prior support for higher values was lower (Fig. 5). Again, this example stresses that a relatively flat

Table 3
Summary of posterior distributions reflecting different aspects of the γ -distribution

Parameter	Mean	sd	Median	95% C.I.
<i>Uninformative prior on α and β</i>				
Shape (α)	1.09	0.42	1.00	0.55–2.12
Rate (β)	20.0	9.10	18.2	7.6–42.8
Mean = α/β	0.057	0.013	0.056	0.037–0.086
Variance = α/β^2	0.0037	0.0027	0.0031	0.001–0.011
CV = $\frac{\sqrt{\text{variance}}}{\text{mean}}$	1.00	0.17	1.00	0.69–1.35
<i>Weak informative prior on CV and mean</i>				
Shape (α)	0.80	0.28	0.75	0.42–1.47
Rate (β)	12.8	6.37	11.6	4.16–28.6
Mean = α/β	0.068	0.018	0.065	0.042–0.12
Variance = α/β^2	0.0075	0.0069	0.0056	0.0017–0.027
CV = $\frac{\sqrt{\text{variance}}}{\text{mean}}$	1.17	0.19	1.16	0.82–1.55

prior at one level of the statistical model does not necessarily reflect ignorance at another level. When mean, variance and CV are of interest, this example also stresses that some informativeness will always enter the analyses. In addition, posterior distributions based on relatively large samples (100 individuals here) may be affected by the misspecification of the priors as well.

4. Concluding remarks

Non-informative priors in Bayesian analyses are important for two reasons. First, they may reflect the true state of current knowledge. Second, it may be required to construct reference models against which models that include subjective information can be contrasted. In both cases, the aim is to let the data drive the analyses, which can be evaluated by comparing results with outcomes from likelihood approaches. We aim at alerting practicing biostatisticians that the assignment of prior distributions in Bayesian statistics should be done very carefully. The main conclusion is that the automatic assignment of flat priors to model parameters is inappropriate (see also for example Carlin and Louis, 2000 and Gelman, 2006). We have presented three examples where presumably uninformative flat prior distributions (strongly) influenced the posterior distributions of (some) parameters of the model. In each situation, weak prior information at the level of the model which is of biological interest, generated priors and posteriors with good behavior (i.e. objectivity, which we evaluated through concordance with the outcome of likelihood methods). This process involved the incorporation of at least some weak prior knowledge. Therefore, one might be tempted to doubt the usefulness of Bayesian techniques when no prior information is available at all or the investigator is not willing to include it in the analyses. However, we show that it is often impossible to specify flat and uninformative priors at each scale of the analysis, so one could argue that complete ignorance does not exist. In

most cases though, some prior knowledge is available anyway relying on common sense. Even if the aim is to let the data drive the analysis which we assume throughout this paper, researchers should think carefully about the prior as a distribution that should reflect the biologically plausible parameter space. Weak priors based on common sense and general biological knowledge can be very useful to guide prior specifications and may even facilitate model identifiability.

In (nearly) any case, prior distributions will reflect subjective individual beliefs and our examples have stressed that some weak subjectivity may be required. Sensitivity analyses are very useful and indispensable to investigate the robustness of results against alternative prior beliefs. Since it is not possible to explore all imaginable alternatives, this calls for a more dynamic interaction between authors and readers of manuscripts. Under the ideal scenario, data of Bayesian (if not any) statistical analysis and the used software(-code) should be made available. Every reader should be able to repeat a presented Bayesian analysis using his/her prior beliefs which could alter the outcome of any analysis. The idea of including subjective priors into the analysis forms the basis of the Bayesian paradigm and should therefore be reflected in publications presenting Bayesian analyses. In fact, this is not a new idea nor is it restricted to Bayesian statistics. From a more general point of view, a recent review of the correctness of statistical analyses reported in major journals like for example Nature, concluded that up to 38% of the papers contained mistakes or errors (Garcia-Berthou and Alcaraz, 2004). This prompted the suggestion that authors should make their data available so that anyone can (re-)analyse them and interpret the outcome independently (Garcia-Berthou and Alcaraz, 2004). In many analyses, model construction, checking and validation rely on subjective decisions on which researchers may disagree. In addition, frequentist hypothesis testing depends on subjective decisions about what to test (Phillips, 2004). Prior distributions in a

Bayesian analysis merely add another level of subjectivity to statistical analyses. As subjective ideas, beliefs and decisions are common in any statistical analysis and often differ among scientists, it should ideally be possible to perform any analysis of published material independently from the inherently subjective decisions of the authors.

Acknowledgements

The author thanks Willem Talloen and two anonymous reviewers for their constructive comments on earlier drafts of this paper.

References

- Bernardo, J.M., 1979. Reference posterior distributions for Bayesian inference (with discussion). *J. Roy. Statist. Soc. B* 41, 113–147.
- Carlin, B.P., Louis, T.A., 2000. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall, London.
- Chen, M.H., Ibrahim, J.G., Yiannoutsos, C., 1999. Prior elicitation, variable selection, and Bayesian computation for logistic regression models. *J. Roy. Statist. Soc. Ser. B* 61, 223–242.
- Congdon, P., 2003. *Applied Bayesian Statistical Modelling*. Wiley, New York.
- Cowles, M.K., Carlin, B.P., 1996. Markov chain Monte Carlo convergence diagnostics: a comparative review. *J. Amer. Statist. Assoc.* 91, 883–904.
- Ellison, A.M., 2004. Bayesian inference in ecology. *Ecol. Lett.* 7, 509–520.
- Garcia-Berthou, E., Alcaraz, C., 2004. Incongruence between test statistics and *p* values in medical papers. *BMC Med. Res. Methodol.* 4, 13.
- Gelman, A., 2006. Prior distributions for variance components in hierarchical models. *Bayesian Statist.*
- Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2000. *Bayesian Data Analysis*. Chapman & Hall, London.
- Mwalili, S.M., Lesaffre, E., Declerck, D., 2005. A Bayesian ordinal logistic regression model to correct for interobserver measurement error in a geographical oral health study. *Appl. Statist.* 54, 77–93.
- O'Hagan, A., Forster, J.J., 2004. *Bayesian Inference*, second ed. vol. 2B of Kendall's Advanced Theory of Statistics. Arnold, London.
- Phillips, C.V., 2004. Publication bias in situ. *BMC Med. Res. Methods* 4, 20.
- Sokal, R.R., Rohlf, F.J., 2000. *Biometry*, third ed. Freeman and Co, New York.
- Van Dongen, S., Lens, L., 2001. Modelling developmental instability as the joint action of noise and stability: a Bayesian approach. *BMC Evol. Biol.* 2, 11.
- Van Dongen, S., Talloen, W., Lens, L., 2005. High variation in developmental instability under non-normal developmental error: a Bayesian perspective. *J. Theor. Biol.* 236, 263–275.
- Verbeke, G., Molenberghs, G., 2000. *Linear Mixed Models for Longitudinal Data*. Springer, New York.