MNRAS 000, 1–10 (2015) Preprint 12 December 2017 Compiled using MNRAS LATEX style file v3.0

How proper are Bayesian models in the astronomical literature?

Hyungsuk Tak,[1★] Sujit K. Ghosh,[2] and Justin A. Ellis[3]
[1]Statistical and Applied Mathematical Sciences Institute, Durham, NC 27713, USA
[2]Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA
[3]Department of Physics and Astronomy, West Virginia University, Morgantown, WV 26506, USA

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

The well-known Bayes theorem assumes that a posterior distribution is a probability distribution. However, the posterior distribution may no longer be a probability distribution if an improper prior distribution (non-probability measure) such as an unbounded uniform prior is used. Improper priors are often used in the astronomical literature to reflect on a lack of prior knowledge, but checking whether the resulting posterior is a probability distribution is sometimes neglected. It turns out that 24 articles out of 75 articles (32%) published online in two renowned astronomy journals (*ApJ* and *MNRAS*) between Jan 1, 2017 and Oct 15, 2017 make use of Bayesian analyses without rigorously establishing posterior propriety. A disturbing aspect is that a Gibbs-type Markov chain Monte Carlo (MCMC) method can produce a seemingly reasonable posterior sample even when the posterior is not a probability distribution (Hobert and Casella 1996). In such cases, researchers may erroneously make probabilistic inferences without noticing that the MCMC sample is from a non-existent probability distribution. We review why checking posterior propriety is fundamental in Bayesian analyses when improper priors are used and discuss how we can set up scientifically motivated proper priors to avoid the pitfalls of using improper priors.

Key words: Markov chain Monte Carlo (MCMC) – improper flat prior – vague prior – uniform prior – inverse gamma prior – non-informative prior – scientifically motivated prior

1 INTRODUCTION

A Bayesian model is uniquely determined by two components: (i) a likelihood function of unknown parameters $\boldsymbol{\theta}$ given the data $\boldsymbol{y}$ denoted by $L(\boldsymbol{\theta};\boldsymbol{y})$, which is proportional to a conditional probability density $f(\boldsymbol{y} \mid \boldsymbol{\theta})$ of a sampling distribution, and (ii) a joint prior density, $p(\boldsymbol{\theta})$. Using the fundamental Bayes theorem, we can derive the resulting posterior density of $\boldsymbol{\theta}$ as follows[1]:

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{y}) = \frac{f(\boldsymbol{y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})}{\int f(\boldsymbol{y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} = \frac{L(\boldsymbol{\theta};\boldsymbol{y})p(\boldsymbol{\theta})}{\int L(\boldsymbol{\theta};\boldsymbol{y})p(\boldsymbol{\theta})d\boldsymbol{\theta}}. \quad (1)$$

Even if a joint prior distribution is improper (i.e., $\int p(\boldsymbol{\theta})d\boldsymbol{\theta} = \infty$), the posterior density in Equation (1) can still be a valid probability density as long as the denominator is finite given the data $\boldsymbol{y}$, i.e., $\int L(\boldsymbol{\theta};\boldsymbol{y})p(\boldsymbol{\theta})d\boldsymbol{\theta} < \infty$. The finite integrability of the product $L(\boldsymbol{\theta};\boldsymbol{y})p(\boldsymbol{\theta})$ is called *posterior propriety*. For making posterior inference about $\boldsymbol{\theta}$ using MCMC methods, it is often unnecessary to compute denominator of (1) because the majority of the MCMC sampling algorithms are based only on the posterior kernel function $q(\boldsymbol{\theta} \mid \boldsymbol{y}) \equiv L(\boldsymbol{\theta};\boldsymbol{y})p(\boldsymbol{\theta})$ which is proportional to $\pi(\boldsymbol{\theta} \mid \boldsymbol{y})$ if posterior propriety holds. See Appendix A for more details of the Bayes theorem.

Posterior propriety plays an important role in MCMC methods that have revolutionized Bayesian computation by making it possible to generate a dependent sample from a complex and high-dimensional target distribution (Brooks et al. 2011). A Markov chain converges to a unique stationary distribution if the chain is irreducible, aperiodic, and positive recurrent (Pinsky and Karlin 2011). The irreducibility holds if a jumping rule of an MCMC method enables a Markov chain to move from any state to any other state with positive probability, and posterior propriety guarantees the other two conditions in practice (Gelman et al. 2013).

However, posterior propriety may not necessarily hold

★ E-mail: hyungsuk.tak@gmail.com

[1] Within a finite-dimensional parametric framework, all density functions are formally defined with respect to a common dominating σ-finite measure like Lebesgue measure (or counting measure).

if the prior p is improper. For example, uniform$(0, \infty)$ and uniform$(-\infty, \infty)$ are widely used improper priors. When the posterior is improper, the most serious issue is that a Gibbs-type MCMC method may still appear to work well by producing a seemingly reasonable posterior sample from the path of the Markov chain (Hobert and Casella 1996). Consequently, users may continue making probabilistic inferences without knowing that the posterior sample is in fact drawn from a non-existent posterior distribution. Hobert and Casella (1996) first warned about this insidious feature of posterior impropriety, recommending either proving posterior propriety (analytically) for improper priors before using Bayesian methods or using jointly proper priors. Since then, statisticians have rigorously established posterior propriety using analytical techniques when improper priors are employed (Daniels 1999; Natarajan and Kass 2000; Ghosh 2010; Tak et al. 2017a; Tak and Morris 2017).

Posterior propriety is sometimes neglected in the astronomical literature. Our investigation reveals that 24 articles out of 75 (32%) published online in *ApJ* and *MNRAS* between Jan 1, 2017 and Oct 15, 2017 report Bayesian analyses without rigorously establishing posterior propriety. We hope that the posterior distributions of these 24 articles are actually proper, although it remains an open issue until posterior propriety is analytically established.

The rest of this article is organized as follows. Section 2 introduces a simple but non-trivial example of using an MCMC method on an improper posterior distribution. In Section 3, we describe how we selected 75 articles published online in *ApJ* and *MNRAS* and investigate posterior propriety in these articles. Section 4 discusses several ways to prove posterior propriety, focusing on using scientifically motivated proper priors for posterior propriety.

2 A SIMPLE BUT NON-TRIVIAL EXAMPLE

Here we reproduce a classical example of Hobert and Casella (1996) that handles a Gaussian hierarchical model commonly used in Bayesian analyses. Suppose the observation y_j ($j = 1, \ldots, n$) follows an independent Gaussian distribution given unknown mean μ_j with known measurement variance V_j, where μ_j follows another independent Gaussian distribution with unknown mean θ and unknown variance σ^2:

$$y_j \mid \mu_j \sim \mathrm{N}(\mu_j, V_j) \quad \text{and} \quad \mu_j \mid \theta, \sigma^2 \sim \mathrm{N}(\theta, \sigma^2). \tag{2}$$

We set up a joint prior kernel function of θ and σ^2 as

$$p_1(\theta, \sigma^2) = p_1(\theta)p_1(\sigma^2) \propto \frac{1}{\sigma^2}, \tag{3}$$

which is clearly improper because $\int_0^\infty \int_{-\infty}^\infty p_1(\theta, \sigma^2) d\theta d\sigma^2 = \infty$. The prior on σ^2 in Equation (3) is equivalent to both $d\sigma/\sigma$ and $d\log(\sigma)$, i.e., a commonly used flat uniform prior on a logarithmic scale of σ. The resulting posterior kernel function is

$$q(\boldsymbol{\mu}, \theta, \sigma^2 \mid \boldsymbol{y}) = p_1(\theta, \sigma^2) \prod_{j=1}^{n} \left[f(y_j \mid \mu_j)\ p(\mu_j \mid \theta, \sigma^2) \right], \tag{4}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$, $\boldsymbol{y} = (y_1, \ldots, y_n)$, and density functions f and p are defined by Equations (2). This posterior kernel function is improper due to the prior on σ^2 regardless of the data; see Appendix A for a proof.

Although the posterior kernel function in Equation (4) is not a probability density, we can still derive its MCMC sampling scheme. Let us assume that $\boldsymbol{y} = (-10, 10)$, $n = 2$, and $V_j = 1$ for simplicity, but we keep using the notation V_j, y_j, and n for generality. We use a Gibbs sampler (Geman and Geman 1984) that iteratively samples the following three conditional posterior distributions: For $j = 1, 2$,

$$\begin{aligned} \mu_j \mid \boldsymbol{\mu}_{[-j]}, \theta, \sigma^2, \boldsymbol{y} &\sim \mathrm{N}\left(\frac{\sigma^2 y_j + V_j\theta}{V_j + \sigma^2},\ \frac{\sigma^2}{V_j + \sigma^2} \right), \\ \theta \mid \boldsymbol{\mu}, \sigma^2, \boldsymbol{y} &\sim \mathrm{N}\left(\bar{\mu},\ \frac{\sigma^2}{n} \right), \\ \sigma^2 \mid \boldsymbol{\mu}, \theta, \boldsymbol{y} &\sim \text{inverse-Gamma}\left(\frac{n}{2},\ \frac{\sum_{j=1}^n (\mu_j - \theta)^2}{2} \right), \end{aligned} \tag{5}$$

where $\boldsymbol{\mu}_{[-j]}$ denotes $\boldsymbol{\mu}$ without the jth component, $\bar{\mu} = \sum_{j=1}^n \mu_j/n$, and the inverse-Gamma$(a, b)$ kernel function of x is $x^{-a-1}\exp(-b/x)$. At iteration i, for example, this sampler updates each parameter in a sequence, i.e., $(\boldsymbol{\mu}^{(i)}, \theta^{(i-1)}, \sigma^{2(i-1)})$, $(\boldsymbol{\mu}^{(i)}, \theta^{(i)}, \sigma^{2(i-1)})$, and $(\boldsymbol{\mu}^{(i)}, \theta^{(i)}, \sigma^{2(i)})$. Almost all MCMC schemes for sampling multivariate parameters use such Gibbs-type updates (either parameter-wise or block-wise) at each iteration to form a Markov chain. We set the initial values as $\boldsymbol{\mu}^{(0)} = (-10, 10)$, $\theta^{(0)} = 0$, and $\sigma^{2(0)} = 1$, and draw 10,000 posterior samples of each parameter.

In Figure 1, we display the histogram, trace plot, and auto-correlation function of 10,000 posterior samples of θ on the top and those of $\log(\sigma^2)$ on the bottom. The posterior sample of θ concentrates on zero and that of $\log(\sigma^2)$ also forms a unimodal histogram. The trace plots show that the Markov chain explores the parameter space rapidly and the auto-correlation functions decrease quickly. The effective sample size of θ is 8,662 and that of $\log(\sigma^2)$ is 2,006. Clearly, the Markov chain appears to converge to a certain probability distribution, and thus it makes sense to make a probabilistic inference using this posterior sample. However, if the initial value of σ^2 were close to zero at which the posterior kernel function puts infinite mass, the Markov chain would stay at $\sigma^2 = 0$ permanently without producing such a seemingly reasonable posterior sample; see Hobert and Casella (1996) for more theoretical details.

Such an inappropriate probabilistic inference based on a non-existent probability distribution can actually happen in reality unless posterior propriety is proven in advance. The article of Pihajoki (2017) published in *MNRAS* uses a similar but more complicated Gaussian hierarchical model that can be built upon a marginalized model of Equation (2), that is,

$$y_j \mid \theta, \sigma^2 \sim \mathrm{N}(\theta, V_j + \sigma^2). \tag{6}$$

Pihajoki (2017) replaces θ in Equation (6) with $\alpha + \beta x_j$, where α and β are unknown regression coefficients and x_j is some known covariate information with its known measurement variance V_{x_j}. Also, Pihajoki (2017) replaces V_j in Equation (6) with $\beta^2 V_{x_j} + V_j - 2\beta\rho\sqrt{V_{x_j}V_j}$, multiplies σ^2 in Equation (6) by $(1+\beta^2)$, and adopts an improper prior $d\sigma/\sigma$ that is equivalent to the problematic choice in Equation (3); see equations (33)–(37) of Pihajoki (2017). The resulting posterior is not a probability distribution. This is because

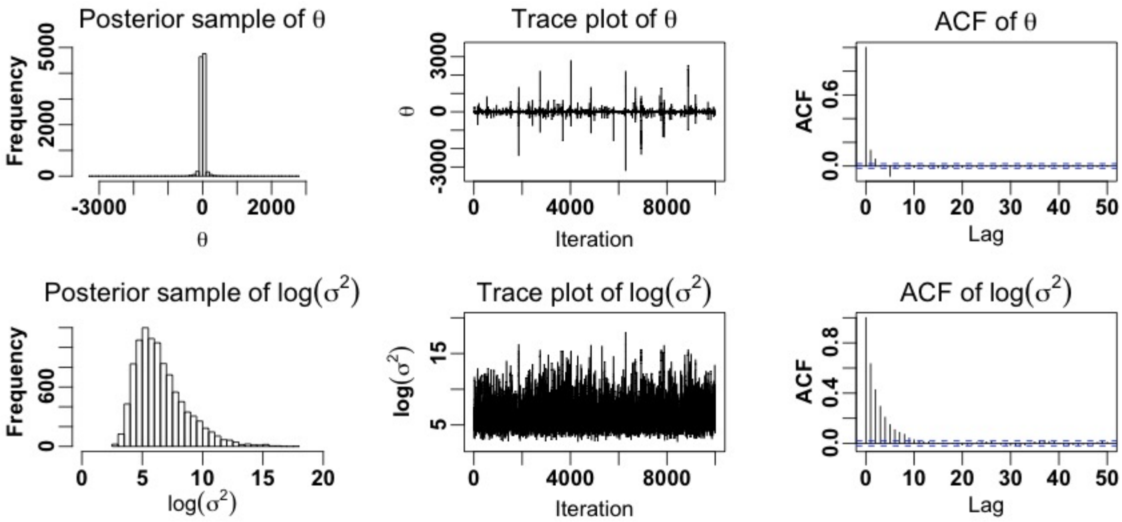

Figure 1. The result of sampling an improper posterior distribution in Equation (4). The histogram, trace plot, and auto-correlation function of 10,000 posterior samples of θ are on the top panels and those of $\log(\sigma^2)$ on the bottom panels. The Markov chain appears to converge to a certain probability distribution, although the target posterior distribution in Equation (4) is not a probability distribution.

when $\beta = 0$, the model of Pihajoki (2017) becomes exactly the same as the one in Equation (6) that is improper with $d\sigma/\sigma$. Therefore, the integral of the posterior kernel function of Pihajoki (2017) is not finite. An MCMC method for this model may not show evidence of posterior impropriety unless a Markov chain starts with the initial value of σ^2 close to zero. Since Pihajoki (2017) does not check posterior propriety before using an MCMC method, the article makes a probabilistic inference using the seemingly reasonable posterior sample drawn from a non-existent posterior distribution. In practice, the inference based on a proper posterior may be similar to the one in Pihajoki (2017) because it is likely that the posterior sample of Pihajoki (2017) resides in a safe (high-likelihood) region without exploring the entire parameter space.

Table 1. Classification of 75 articles published online in *ApJ* and *MNRAS* between Jan 1, 2017 and Oct 15, 2017 according to their prior distributions.

	ApJ	*MNRAS*
(a) Jointly proper priors	18	33
(b) Jointly improper priors	1	3
(c) Unclear priors	11	9
Total	30	45

3 POSTERIOR PROPRIETY IN THE ASTRONOMICAL LITERATURE

We investigated the literature published online in *ApJ* and *MNRAS* between Jan 1, 2017 and Oct 15, 2017. On the webpage of IOPscience[2], we found 33 *ApJ* articles whose titles or abstracts contain a word ‘Bayesian’. We excluded three of them because one is an erratum (Eadie et al. 2017b) and the other two just use Bayesian methods previously developed by other researchers (Abeysekara et al. 2017; Murphy et al. 2017). We also obtained a list of 51 articles that have the word ‘Bayesian’ in their abstracts from the webpage of *MNRAS*[3]. We did not consider six of them because one mentions a Bayesian analysis as a potential application (Watkinson et al. 2017), another uses a Bayesian information criterion for a model selection (Wilkinson et al. 2017), and the other four simply utilize Bayesian methods developed in other articles (Pinamonti et al. 2017; Green et al. 2017; Sampedro et al. 2017; Basak et al. 2017).

None of the remaining 75 articles mention posterior propriety, and thus we check further by classifying them into three categories; (a) priors are jointly proper; (b) priors are jointly improper; and (c) priors are not clearly specified. The last category includes cases where uniform (or flat) prior distributions are used without clearly specified ranges because there are infinitely many uniform prior distributions according to their ranges. Table 1 summarizes the classification; see Appendix B for details. More than half of the articles use jointly proper priors, but there are 24 articles in categories (b) and (c) that need proofs for posterior propriety to assure that their scientific arguments are actually based on their targeted posterior distributions.

The issue of the 20 articles in category (c) is not only posterior propriety but also reproducibility because there are infinitely many possible models that these articles use. Proving posterior propriety can contribute to reproducible science as a by-product because its first step is to clarify a Bayesian model by specifying its likelihood function of unknown parameters and their prior distributions.

[2] http://iopscience.iop.org/
[3] https://academic.oup.com/mnras

4 DISCUSSION: SCIENTIFICALLY MOTIVATED PROPER PRIORS FOR POSTERIOR PROPRIETY

Improper prior distributions are widely used because they are considered non-informative and convenient for modeling[4]. An improper uniform prior on a location parameter, e.g., $p_1(\theta)$ in Equation (3), has an advantage to make the data (likelihood function) speak more about the parameter when prior knowledge is limited. It also results in a proper posterior distribution in many cases.

However, there is a cost to be paid for using improper priors, which is often neglected: *Proving posterior propriety.* There are several ways to prove it. The most rigorous one is to analytically show that the integral of the target posterior kernel function is finite. A justification based on conditional densities (or conditional kernel functions) does not necessarily guarantee the finite integral. For example, suppose we are interested in a kernel function, $q(x, y) \propto \exp(-xy)$ for $x, y > 0$. Its impropriety becomes clear once we analytically integrate out one of the variables, while its conditional kernel functions, $q(x \mid y)$ and $q(y \mid x)$, are finitely integrable without implying any impropriety (Hobert and Casella 1996). However, if the dimension is large and the model is complicated, it is challenging to prove posterior propriety analytically.

Second, we can apply existing theorems about posterior propriety only if a model considered in a theorem is the same as a model to be used. For example, suppose a model to be used has two more parameters than a model whose posterior propriety is proven in a theorem. Posterior propriety of a model to be used holds only if a marginalized model (with the two additional parameters integrated out from a model to be used) is the same as a model considered in the theorem. This is because an unexpected term that is a function of unknown parameters may arise during the integration, which can make seemingly similar models completely different.

Jointly proper priors guarantee posterior propriety based on standard probability theory, and thus a proof for posterior propriety is not needed. A proper prior is sometimes used as a vague and non-informative choice. For a location parameter whose support is a real line, e.g., θ in Equation (3), it is reasonable to adopt a diffuse Gaussian or heavy tailed diffuse Student's t prior distribution with any mean but arbitrarily large variance. In this case, the resulting posterior inference with a proper diffuse prior will be similar to the one with improper flat uniform prior. However, the former does not require users to prove posterior propriety while the latter does. As for a parameter whose support is a positive real line, e.g., σ^2 in Equation (3), a log-Normal, half Normal, and half Student's t centered at zero with relatively large variance are known to be vague choices (Gelman 2006).

Using scientifically motivated priors is one advantage of using Bayesian machinery because it provides a natural way to incorporate scientific knowledge into inference via priors. Proper priors are ideal for this purpose. Tak et al. (2017b), for example, use a uniform$(-30, 30)$ prior for the unknown mean magnitude of a damped random walk process, considering a practical magnitude range from that of the Sun to that of the faintest object identifiable by the Hubble Space Telescope. This prior can be considered weakly informative because the range of the uniform prior is wide enough not to affect the resulting posterior inference. (A bounded uniform prior is not non-informative because its hard bounds completely exclude a certain range of parameter values.) If the range of a uniform prior is narrow and thus it significantly influences the posterior inference, such an informative choice may need further justification.

For an unknown parameter whose support is the positive real line, an inverse-Gamma prior can be used as a scientifically motivated prior because it enables us to set up a soft lower bound of a parameter using scientific knowledge or past studies. The kernel function of x that follows an inverse-Gamma(a, b) distribution is $x^{-a-1}\exp(-b/x)$. Its mode, $b/(a+1)$, plays a role of the soft lower bound, and a small shape parameter a is desirable for a weakly informative prior[5]; an equivalent inverse-χ^2 prior has the degrees of freedom $2a$ and scale b/a. When x goes to infinity, the right tail of this kernel function decreases as a power law, while the left tail exponentially decreases as x approaches zero. Thus x is less likely to take on values much smaller than the mode (soft lower bound) a priori. Modeling quasar variability, for example, Tak et al. (2017b) adopt an inverse-Gamma(1, b) prior for the unknown timescale (in days) of a damped random walk process. The scale parameter b is set to one day so that its soft lower bound, 0.5 day, is much smaller than any timescale estimates of 9,275 quasars in a past study (MacLeod et al. 2010) a priori.

For a second-level variance component in a Gaussian hierarchical model such as σ^2 in Equation (2), Gelman (2006) does not recommend using an inverse-Gamma(a, b) prior as a vague and non-informative choice with arbitrarily small values of a and b. This makes sense because an inverse-Gamma prior always sets up a soft lower bound a priori. When the likelihood puts significant weight at zero but with relatively small data size, it is difficult for the likelihood to dominate the soft lower bound. In this case, the resulting posterior inference becomes sensitive to the location of the soft lower bound. Thus when the data size is small, it is important to construct the soft lower bound carefully, considering scientific knowledge or past studies.

Let us revisit the example in Section 2 to see an impact of adopting jointly proper priors. Instead of the improper choice in Equation (3), we set a diffuse Gaussian prior for θ and a weakly informative inverse-Gamma prior for σ^2 inde-

[4] We do not consider computational convenience including conjugacy. This is because most astronomers are familiar with a generic MCMC sampler such as `emcee` (Foreman-Mackey et al. 2013) and `PyStan` (Carpenter et al. 2017) that automatically samples the target posterior given the likelihood and prior specification.

[5] The relationship between the inverse-Gamma and scaled inverse-χ^2 distributions allows us to interpret the shape parameter of the inverse-Gamma as half the number of pseudo realizations that would carry equivalent information as the prior distribution. (See, e.g., Gelman et al. (2013) for a discussion of the pseudo observation interpretation of prior distributions.) For example, an inverse-Gamma prior with the unit shape parameter, $a = 1$, carries relatively small amount of information from two pseudo observations. If the number of observed data is much larger than two, the likelihood can dominate this inverse-Gamma prior with ease.

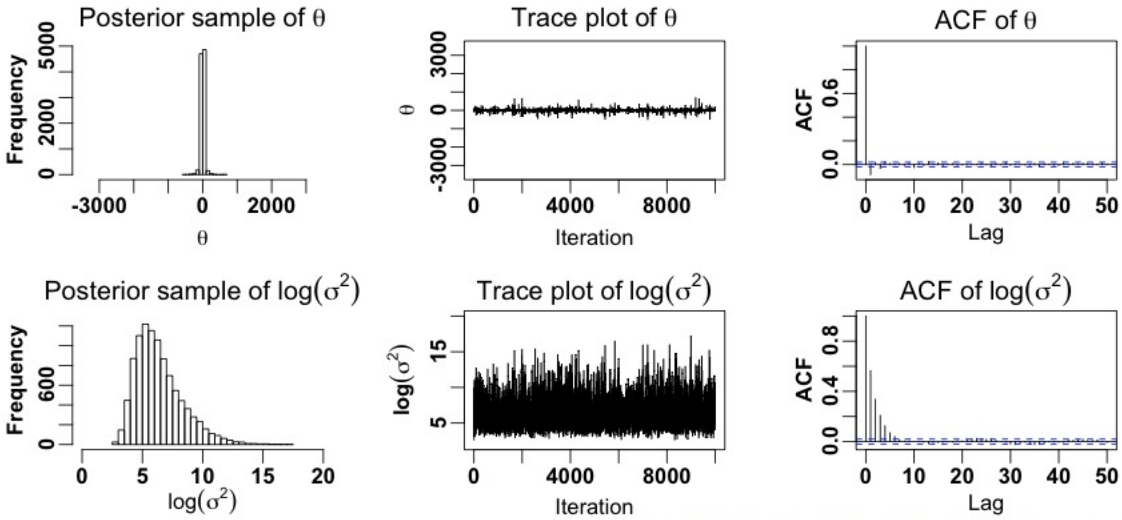

Figure 2. The result of sampling the proper posterior kernel function in Equation (8) that is based on weakly informative and vague proper priors in Equation (7). The histogram, trace plot, and auto-correlation function of 10,000 posterior samples of θ are on the top panels and those of $\log(\sigma^2)$ on the bottom panels. The ranges of vertical and horizontal axes are the same as those in Figure 1. Since priors in Equation (7) are jointly proper, we know that the resulting posterior q^* in Equation (8) is proper and thus the posterior sample is from q^*. Also, the sampling result hardly varies even if the initial value of σ^2 is close to zero. The ensuing Bayesian inference on θ is quite different from that in Section 2 with much shorter tails.

pendently:

$$\theta \sim \text{N}(0,\ 10^5) \quad \text{and} \quad \sigma^2 \sim \text{inverse-Gamma}(0.001, 0.001), \tag{7}$$

where the soft lower bound of σ^2 is set to 0.001/1.001 a priori that is at least significantly smaller than the sample variance of the data, −10 and 10. We denote the joint prior distribution in Equation (7) by $p^*(\theta, \sigma^2)$. The resulting full posterior kernel function is

$$q^*(\boldsymbol{\mu}, \theta, \sigma^2 \mid \mathbf{y}) = p^*(\theta, \sigma^2) \prod_{j=1}^{n} \Big[f(y_j \mid \mu_j)\ p(\mu_j \mid \theta, \sigma^2) \Big], \tag{8}$$

where density functions f and p are defined in Equation (2). The corresponding Gibbs sampler updates each coordinate of $\boldsymbol{\mu}$ by its conditional posterior specified in Equation (5) but updates θ and σ^2 by

$$\begin{aligned} &\theta \mid \mu, \sigma^2, y \sim \text{N}\left(\frac{(n/\sigma^2)\bar{\mu}}{n/\sigma^2 + 1/10^5},\ \frac{1}{n/\sigma^2 + 1/10^5} \right), \\ &\sigma^2 \mid \mu, \theta, y \\ &\ \sim \text{inverse-Gamma}\left(\frac{n}{2} + 0.001,\ \frac{\sum_{j=1}^{n} (\mu_j - \theta)^2}{n} + 0.001 \right). \end{aligned} \tag{9}$$

These two conditional distributions in Equation (9) are similar to those in Equation (5), considering that both 10^{-5} and 0.001 are close to zero. The other simulation configuration is the same.

Figure 2 exhibits the sampling result. The ranges of the horizontal and vertical axes in each panel are the same as those of Figure 1 for a comparison. Because of the jointly proper priors in Equation (7), we know that the resulting posterior kernel function q^* in Equation (8) is proper and thus the posterior sample displayed in Figure 2 represents the target posterior distribution. Also, though not shown here, the MCMC method produces nearly the same sampling result regardless of the initial value of σ^2. The histogram of θ in Figure 2 has much shorter tails than that in Figure 1, although the histogram of $\log(\sigma^2)$ in Figure 2 is similar to that in Figure 1. The soft lower bound of $\log(\sigma^2)$, i.e., $\log(0.001/1.001) = -6.909$, is too small to be displayed. This indicates that this soft lower bound does not affect the posterior inference much even though there are just two data points; the degrees of freedom of an equivalent inverse-χ^2 prior are 0.002. The effective sample size improves greatly; it is 9,974 for θ and 2,808 for $\log(\sigma^2)$. Consequently, the inference on θ becomes quite different from that in Section 2, empirically proving that checking posterior propriety before using MCMC methods can make a significant difference.

It is well understood that any probabilistic inference such as Bayesian framework should be based on a probability distribution. Jointly proper priors lead to a proper posterior distribution and can be either vague or scientifically motivated. However, flat improper priors are often used to represent the lack of prior knowledge. Such improper priors combined with a likelihood function can result in an improper posterior distribution that is not finitely integrable, and the resulting inference may or may not be based on a probability distribution. Therefore, when improper priors are adopted, posterior propriety must be carefully proven before using any MCMC methods. Proving posterior propriety requires specifying a Bayesian model, i.e., a likelihood function of unknown parameters and their prior distributions, which is also desirable for reproducible science. We hope that posterior propriety draws more attention when improper priors are used in the astronomical literature.

ACKNOWLEDGEMENTS

Hyungsuk Tak acknowledges partial support from National Science Foundation grants DMS 1127914 and DMS 1638521 given to the Statistical and Applied Mathematical Sciences Institute. We thank David E. Jones and David C. Stenning for a productive discussion at the International Centre for Theoretical Sciences in Bangalore, India, during a visit when participating in the program 'Time Series Analysis for Synoptic Surveys and Gravitational Wave Astronomy'.

APPENDIX A: THE BAYES THEOREM IN DETAIL

It is well known that a Bayesian statistical model consists of (i) a sampling distribution, $f(\boldsymbol{y} \mid \boldsymbol{\theta})$, denoting the conditional probability density of the data y given unknown parameters $\boldsymbol{\theta}$; and (ii) a prior distribution, $p(\boldsymbol{\theta})$, denoting an unconditional probability density of $\boldsymbol{\theta}$. The resulting joint density of $\boldsymbol{y}$ and $\boldsymbol{\theta}$ is $f(\boldsymbol{y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})$ based on standard probability theory. We can also express this joint density as a product of the unconditional density of the data $h(\boldsymbol{y}) \equiv \int f(\boldsymbol{y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$ and the so-called posterior density of $\boldsymbol{\theta}$ given $\boldsymbol{y}$, i.e.,

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{y}) = \frac{f(\boldsymbol{y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})}{\int f(\boldsymbol{y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}. \tag{A1}$$

All density functions are formally defined with respect to Lebesgue measure (or counting measure). However, in many scientific applications, we may relax the need for the use of a probability measure for the prior distribution by using a kernel function $k(\boldsymbol{\theta}) = c_0 p(\boldsymbol{\theta})$ for some constant $c_0 > 0$ and also write the likelihood function $L(\boldsymbol{\theta}; \boldsymbol{y}) = c(\boldsymbol{y})f(\boldsymbol{y} \mid \boldsymbol{\theta})$ for some function $c(\boldsymbol{y}) > 0$. Then, as illustrated in Ghosh (2010), we can reexpress the equation (A1) as

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{y}) = \frac{f(\boldsymbol{y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})}{\int f(\boldsymbol{y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} = \frac{L(\boldsymbol{\theta}; \boldsymbol{y})k(\boldsymbol{\theta})}{\int L(\boldsymbol{\theta}; \boldsymbol{y})k(\boldsymbol{\theta})d\boldsymbol{\theta}}. \tag{A2}$$

As illustrated in Section 1, even if $\int k(\boldsymbol{\theta})d\boldsymbol{\theta} = \infty$, making the prior distribution improper, the posterior density as given in (A2) is still a valid probability density as long as the denominator $\int L(\boldsymbol{\theta}; \boldsymbol{y})k(\boldsymbol{\theta})d\boldsymbol{\theta} < \infty$ is finitely integrable. However, an improper prior necessarily leads to improper marginal distribution of the data $\boldsymbol{y}$ (and vice versa), i.e., $\int k(\boldsymbol{\theta})d\boldsymbol{\theta} = \infty$ is equivalent to $\int h(\boldsymbol{y})d\boldsymbol{y} = \infty$. This is because

$$\begin{aligned}\int h(\boldsymbol{y})d\boldsymbol{y} &= \int\int f(\boldsymbol{y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}\, d\boldsymbol{y}\\ &= \int\int f(\boldsymbol{y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{y}\, d\boldsymbol{\theta} = \int p(\boldsymbol{\theta})d\boldsymbol{\theta},\end{aligned}$$

where the second equality holds from Fubini's theorem. This aspect is not a concern if $\pi(\boldsymbol{\theta} \mid \boldsymbol{y})$ is a proper probability density. It is well known that, in order to make an inference about $\boldsymbol{\theta}$ (or its function) conditional on the observed data, it is often sufficient to draw samples from a posterior kernel given by $L(\boldsymbol{\theta}; \boldsymbol{y})k(\boldsymbol{\theta})$, i.e., the numerator in (A2) without the need to evaluate the denominator. Unfortunately, Gibbs-type MCMC methods can generate a sample from the posterior kernel which need not correspond to a proper posterior distribution; see Hobert and Casella (1996) for various examples. When a proper prior density $p(\boldsymbol{\theta})$ is used, this is not an issue as a posterior distribution is necessarily proper by standard probability theory. However, when an improper prior kernel is used, then the only option is to verify *analytically* that integral in the denominator of (A2) is finite.

APPENDIX B: PROOF OF POSTERIOR IMPROPRIETY IN SECTION 2

The full posterior kernel function $q(\boldsymbol{\mu}, \theta, \sigma^2 \mid \boldsymbol{y})$ in Equation (4) is improper because the marginal posterior kernel function $q_1(\sigma^2 \mid \boldsymbol{y})$ with $\boldsymbol{\mu}$ and θ integrated out from $q(\boldsymbol{\mu}, \theta, \sigma^2 \mid \boldsymbol{y})$ is improper. We derive the marginal posterior kernel function of θ and σ^2 by integrating out $\boldsymbol{\mu}$ from $q(\boldsymbol{\mu}, \theta, \sigma^2 \mid \boldsymbol{y})$:

$$\begin{aligned}q_2(\theta, \sigma^2 \mid \boldsymbol{y}) &= \int_{\mathbb{R}^n} q(\boldsymbol{\mu}, \theta, \sigma^2 \mid \boldsymbol{y})\, d\boldsymbol{\mu}\\ &= p_1(\theta, \sigma^2)\prod_{j=1}^{n} f_1(y_j \mid \theta, V_j + \sigma^2)\\ &= \frac{1}{\sigma^2}\exp\left(-\sum_{j=1}^{n}\frac{(y_j-\theta)^2}{2(V_j+\sigma^2)}\right)\prod_{j=1}^{n}(V_j+\sigma^2)^{-0.5}\\ &= \frac{1}{\sigma^2}\exp\left(-\sum_{j=1}^{n}\frac{(y_j-\hat{y})^2}{2(V_j+\sigma^2)} - \frac{(\theta-\hat{y})^2}{2V^*}\right)\prod_{j=1}^{n}(V_j+\sigma^2)^{-0.5},\end{aligned} \tag{B1}$$

where density functions p_1 and f_1 are defined in Equations (3) and (6), respectively,

$$\hat{y} \equiv \frac{\sum_{j=1}^{n} y_j/(V_j+\sigma^2)}{\sum_{j=1}^{n} 1/(V_j+\sigma^2)} \quad \text{and} \quad V^* \equiv \frac{1}{\sum_{j=1}^{n} 1/(V_j+\sigma^2)}.$$

Next we marginalize out θ from Equation (B1) as follows:

$$\begin{aligned}q_1(\sigma^2 \mid \boldsymbol{y}) &= \int_{\mathbb{R}} q_2(\theta, \sigma^2 \mid \boldsymbol{y})\, d\theta\\ &= \frac{(V^*)^{0.5}}{\sigma^2}\exp\left(-\sum_{j=1}^{n}\frac{(y_j-\hat{y})^2}{2(V_j+\sigma^2)}\right)\prod_{j=1}^{n}(V_j+\sigma^2)^{-0.5}.\end{aligned} \tag{B2}$$

This marginal posterior kernel function of σ^2 approaches infinity as σ^2 goes to zero due to the prior on σ^2, i.e., $d\sigma^2/\sigma^2$. Therefore, $\int_{\mathbb{R}^+} q_1(\sigma^2 \mid \boldsymbol{y})\, d\sigma^2 = \infty$.

APPENDIX C: CLASSIFICATION OF 75 ARTICLES IN SECTION 3

Among the 30 articles published online in *ApJ*, 18 articles adopt jointly proper priors, and we classify these into category (a); Fogarty et al. (2017); Montes-Solís and Arregui (2017); Zevin et al. (2017); Leung et al. (2017); Farnes et al. (2017); Benson et al. (2017); Sathyanarayana Rao et al. (2017); Sliwa et al. (2017); Park et al. (2017); Khrykin et al. (2017); Budavári et al. (2017); Wang et al. (2017); Scherrer and McKenzie (2017); Tabatabaei et al. (2017); Lund et al. (2017); Eadie et al. (2017a); Martínez-García et al. (2017); Küpper et al. (2017).

Knežević et al. (2017) use an unbounded flat prior on the logarithm of the total flux without proving posterior propriety, and thus we classify this article into category (b).

We cannot check posterior propriety of 11 articles published online in *ApJ* because they do not specify priors clearly, i.e., their Bayesian models are not uniquely determined. We designate them as category (c) which contains cases where Bayesian methods developed by other researchers are used without specifying any details or cases where uniform (or flat) priors are used without clear ranges. Here we list them; Kern et al. (2017) use flat priors over the astrophysical parameters; Bitsakis et al. (2017) say nothing about priors; Raithel et al. (2017) do not specify a joint prior on five pressures; Oyarzún et al. (2017) utilize flat priors on all parameters; Tanaka et al. (2017) make use of uniform priors on $m_{\rm TRGB}$ and a; Mandel et al. (2017) adopt a flat prior on μ_s whose range is unclear; Daylan et al. (2017) adopt uniform priors on many parameters; Warren et al. (2017) utilize an uninformative prior on β; Solá et al. (2017) make use of an uninformative prior on h; Eilers et al. (2017) do not specify priors on γ, σ_C, and σ_{ij}; and Jones et al. (2017) use flat priors on SN Ia distances.

Next, we classify 45 articles published online in *MNRAS* into three categories. Category (a) contains 33 articles whose priors are jointly proper; Ashton et al. (2017); Bainbridge and Webb (2017); Ata et al. (2017); Wagner-Kaiser et al. (2017); Cibirka et al. (2017); Patel et al. (2017); Si et al. (2017); Dwelly et al. (2017); Maund (2017); Davis et al. (2017); Hahn et al. (2017); Burgess (2017); Silburt and Rein (2017); MacDonald and Madhusudhan (2017); Abdurro'uf and Akiyama (2017); Kafle et al. (2017); Aigrain et al. (2017); Henderson et al. (2017); Kimura et al. (2017); Schellenberger and Reiprich (2017); Mejía-Narváez et al. (2017); Köhlinger et al. (2017); Dam et al. (2017); Garnett et al. (2017); Andrews et al. (2017); Kovalenko et al. (2017); McEwen et al. (2017); Oh et al. (2017); Duncan et al. (2017); Galvin et al. (2017); Salvato et al. (2017); Yu and Liu (2017); and Greig and Mesinger (2017).

Three articles published online in *MNRAS* employ improper priors without proving posterior propriety; Sereno and Ettori (2017) use an improper uniform prior on `mu.Z.0` whose upper limit is infinity; Kos (2017) adopts an improper prior on $l_{\rm se}$ without an upper limit; and we prove posterior impropriety of Pihajoki (2017) resulting from the improper prior on σ^2.

We cannot judge posterior propriety of 9 articles published in *MNRAS* because their priors are not clearly specified; Rodrigues et al. (2017) adopt flat priors on metallicity and age; Vallisneri and van Haasteren (2017) do not specify priors on $\sigma_{\rm out}$ and c; Binney and Wong (2017) use uniform priors for the logarithm of scale parameters; Ashworth et al. (2017) use flat priors on α_3 and A_V; Jeffreson et al. (2017) utilize uniform priors on eight parameters (three are on the logarithmic scale); Molino et al. (2017) adopt flat priors on galaxy type and redshift; Accurso et al. (2017) do not specify priors on α and β_j; Günther et al. (2017) adopt uniform priors on all parameters; and Igoshev and Popov (2017) do not clarify a joint prior on $\vec{v}_o$.

REFERENCES

Abdurro'uf and Akiyama, M. (2017). Understanding the Scatter in the Spatially Resolved Star Formation Main Sequence of Local Massive Spiral Galaxies. *Monthly Notices of the Royal Astronomical Society*, 469(3):2806–2820.

Abeysekara, A. U., Albert, A., Alfaro, R., Alvarez, C., Álvarez, J. D., Arceo, R., et al. (2017). Daily Monitoring of TeV Gamma-Ray Emission from Mrk 421, Mrk 501, and the Crab Nebula with HAWC. *The Astrophysical Journal*, 841(2):100.

Accurso, G., Saintonge, A., Catinella, B., Cortese, L., Davé, R., Dunsheath, S. H., et al. (2017). Deriving a Multivariate αCO Conversion Function using the $[C_{II}]$/CO (1–0) Ratio and its Application to Molecular Gas Scaling Relations. *Monthly Notices of the Royal Astronomical Society*, 470(4):4750–4766.

Aigrain, S., Parviainen, H., Roberts, S., Reece, S., and Evans, T. (2017). Robust, Open-Source Removal of Systematics in Kepler Data. *Monthly Notices of the Royal Astronomical Society*, 471(1):759–769.

Andrews, J. J., Chanamé, J., and Agüeros, M. A. (2017). Wide Binaries in Tycho-Gaia: Search Method and the Distribution of Orbital Separations. *Monthly Notices of the Royal Astronomical Society*, 472(1):675–699.

Ashton, G., Jones, D. I., and Prix, R. (2017). On the Free Precession Candidate PSR B1828–11: Evidence for Increasing Deformation. *Monthly Notices of the Royal Astronomical Society*, 467(1):164–178.

Ashworth, G., Fumagalli, M., Krumholz, M. R., Adamo, A., Calzetti, D., Chandar, R., Cignoni, M., et al. (2017). Exploring the IMF of Star Clusters: A Joint SLUG and LEGUS Effort. *Monthly Notices of the Royal Astronomical Society*, 469(2):2464–2480.

Ata, M., Kitaura, F.-S., Chuang, C.-H., Rodríguez-Torres, S., Angulo, R. E., Ferraro, S., et al. (2017). The Clustering of Galaxies in the Completed SDSS-III Baryon Oscillation Spectroscopic Survey: Cosmic Flows and Cosmic Web from Luminous Red Galaxies. *Monthly Notices of the Royal Astronomical Society*, 467(4):3993–4014.

Bainbridge, M. B. and Webb, J. K. (2017). Artificial Intelligence Applied to the Automatic Analysis of Absorption Spectra. Objective Measurement of the Fine Structure Constant. *Monthly Notices of the Royal Astronomical Society*, 468(2):1639–1670.

Basak, R., Iyyani, S., Chand, V., Chattopadhyay, T., Bhattacharya, D., Rao, A. R., and Vadawale, S. V. (2017). Surprise in Simplicity: An Unusual Spectral Evolution of a Single Pulse GRB 151006A. *Monthly Notices of the Royal Astronomical Society*, 472(1):891–903.

Benson, B., Wittman, D. M., Golovich, N., Jee, M. J., van Weeren, R. J., and Dawson, W. A. (2017). MC 2 : A Deeper Look at ZwCl 2341.1+0000 with Bayesian Galaxy Clustering and Weak Lensing Analyses. *The Astrophysical Journal*, 841(1):7.

Binney, J. and Wong, L. K. (2017). Modelling the Milky Way's Globular Cluster System. *Monthly Notices of the Royal Astronomical Society*, 467(2):2446–2457.

Bitsakis, T., Bonfini, P., Gonzáez-Lópezlira, R. A., Ramírez-Siordia, V. H., Bruzual, G., Charlot, S., Maravelias, G., and Zaritsky, D. (2017). A Novel Method to Automatically Detect and Measure the Ages of Star Clusters in Nearby Galaxies: Application to the Large Magellanic Cloud. *The Astrophysical Journal*, 845(1):56.

Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. CRC Press, Boca Raton, FL, USA.

Budavári, T., Szalay, A. S., and Loredo, T. J. (2017). Faint Object Detection in Multi-Epoch Observations via Catalog Data Fusion. *The Astrophysical Journal*, 838(1):52.

Burgess, J. M. (2017). The Rest-Frame Golenetskii Correlation via a Hierarchical Bayesian Analysis. *Monthly Notices of the Royal Astronomical Society*, page stx1159.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., and Li, P. (2017). Rgbp: An R Package for Gaussian, Poisson, and Binomial Random Effects Models, with Frequency Coverage Evaluations. *Journal of Statistical Software*, 71(1):1–32.

Cibirka, N., Cypriano, E. S., Brimioulle, F., Gruen, D., Erben, T., van Waerbeke, L., Miller, L., Finoguenov, A., Kirkpatrick, C., Henry, J. P., Rykoff, E., Rozo, E., Dupke, R., Kneib, J.-P., Shan, H., and Spinelli, P. (2017). CODEX Weak Lensing: Concentration of Galaxy Clusters at z ∼ 0.5. *Monthly Notices of the Royal Astronomical Society*, 468(1):1092–1116.

Dam, L. H., Heinesen, A., and Wiltshire, D. L. (2017). Apparent Cosmic Acceleration from Type Ia Supernovae. *Monthly Notices of the Royal Astronomical Society*, 472(1):835–851.

Daniels, M. J. (1999). A Prior for the Variance in Hierarchical Models. *The Canadian Journal of Statistics*, 27(3):567–578.

Davis, T. A., Bureau, M., Onishi, K., Cappellari, M., Iguchi, S., and Sarzi, M. (2017). WISDOM Project – II. Molecular Gas Measurement of the Supermassive Black Hole Mass in NGC 4697. *Monthly Notices of the Royal Astronomical Society*, 468(4):4675–4690.

Daylan, T., Portillo, S. K. N., and Finkbeiner, D. P. (2017). Inference of Unresolved Point Sources at High Galactic Latitudes Using Probabilistic Catalogs. *The Astrophysical Journal*, 839(1):4.

Duncan, K. J., Brown, M. J. I., Williams, W. L., Best, P. N., Buat, V., Burgarella, D., et al. (2017). Photometric Redshifts for the Next Generation of Deep Radio Continuum Surveys - I: Template Fitting. *Monthly Notices of the Royal Astronomical Society*, page stx2536.

Dwelly, T., Salvato, M., Merloni, A., Brusa, M., Buchner, J., Anderson, S. F., Boller, T., Brandt, W. N., et al. (2017). SPIDERS: Selection of Spectroscopic Targets Using AGN Candidates Detected in All-Sky X-ray Surveys. *Monthly Notices of the Royal Astronomical Society*, 469(1):1065–1095.

Eadie, G. M., Springford, A., and Harris, W. E. (2017a). Bayesian Mass Estimates of the Milky Way: Including Measurement Uncertainties with Hierarchical Bayes. *The Astrophysical Journal*, 835(2):167.

Eadie, G. M., Springford, A., and Harris, W. E. (2017b). Erratum: "Bayesian Mass Estimates of the Milky Way: Including Measurement Uncertainties with Hierarchical Bayes" (2017, ApJ, 835, 167). *The Astrophysical Journal*, 838(1):76.

Eilers, A.-C., Hennawi, J. F., and Lee, K.-G. (2017). Joint Bayesian Estimation of Quasar Continua and the Lyα Forest Flux Probability Distribution Function. *The Astrophysical Journal*, 844(2):136.

Farnes, J. S., Rudnick, L., Gaensler, B. M., Haverkorn, M., O'Sullivan, S. P., and Curran, S. J. (2017). Observed Faraday Effects in Damped Lyα Absorbers and Lyman Limit Systems: The Magnetized Environment of Galactic Building Blocks at Redshift = 2. *The Astrophysical Journal*, 841(2):67.

Fogarty, K., Postman, M., Larson, R., Donahue, M., and Moustakas, J. (2017). The Relationship Between Brightest Cluster Galaxy Star Formation and the Intracluster Medium in CLASH. *The Astrophysical Journal*, 846(2):103.

Foreman-Mackey, D., Hogg, D. W., Lang, D., and Goodman, J. (2013). emcee: The MCMC Hammer. *Publications of the Astronomical Society of the Pacific*, 125(925):306–312.

Galvin, T. J., Seymour, N., Marvil, J., Filipović, M. D., Tothill, N. F. H., McDermid, R. M., et al. (2017). The Spectral Energy Distribution of Powerful Starburst Galaxies I: Modelling the Radio Continuum. *Monthly Notices of the Royal Astronomical Society*, page stx2613.

Garnett, R., Ho, S., Bird, S., and Schneider, J. (2017). Detecting Damped Lyα Absorbers with Gaussian Processes. *Monthly Notices of the Royal Astronomical Society*, 472(2):1850–1865.

Gelman, A. (2006). Prior Distributions for Variance Parameters in Hierarchical Models. *Bayesian Analysis*, 1(3):515–533.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. CRC Press, Boca Raton, FL, USA.

Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):721–741.

Ghosh, S. K. (2010). Basics of Bayesian Methods. In Bang, H., Zhou, X. K., van Epps, H. L., and Mazumdar, M., editors, *Statistical Methods in Molecular Biology*, pages 155–178. Humana Press, Totowa, NJ, USA.

Green, J. A., Breen, S. L., Fuller, G. A., McClure-Griffiths, N. M., Ellingsen, S. P., Voronkov, M. A., Avison, A., et al. (2017). The 6-GHz Multibeam Maser Survey II. Statistical Analysis and Galactic Distribution of 6668-MHz Methanol Masers. *Monthly Notices of the Royal Astronomical Society*, 469(2):1383–1402.

Greig, B. and Mesinger, A. (2017). Simultaneously Constraining the Astrophysics of Reionization and the Epoch of Heating with 21CMMC. *Monthly Notices of the Royal Astronomical Society*, 472:2651–2669.

Günther, M. N., Queloz, D., Gillen, E., McCormac, J., Bayliss, D., Bouchy, F., et al. (2017). Centroid Vetting of Transiting Planet Candidates from the Next Generation Transit Survey. *Monthly Notices of the Royal Astronomical Society*, 472(1):295–307.

Hahn, C., Vakili, M., Walsh, K., Hearin, A. P., Hogg, D. W., and Campbell, D. (2017). Approximate Bayesian Computation in Large-Scale Structure: Constraining the Galaxy–Halo Connection. *Monthly Notices of the Royal Astronomical Society*, 469(3):2791–2805.

Henderson, C. S., Skemer, A. J., Morley, C. V., and Fortney, J. J. (2017). A New Statistical Method for Characterizing the Atmospheres of Extrasolar Planets. *Monthly Notices of the Royal Astronomical Society*, 470(4):4557–4563.

Hobert, J. P. and Casella, G. (1996). The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models. *Journal of the American Statistical Association*, 91(436):1461–1473.

Igoshev, A. P. and Popov, S. B. (2017). How to Make a Mature Accreting Magnetar. *Monthly Notices of the Royal Astronomical Society*, page stx2573.

Jeffreson, S. M. R., Sanders, J. L., Evans, N. W., Williams, A. A., Gilmore, G. F., Bayo, A., et al. (2017). The Gaia-ESO Survey: Dynamical Models of Flattened, Rotating Globular Clusters. *Monthly Notices of the Royal Astronomical Society*, 469(4):4740–4762.

Jones, D. O., Scolnic, D. M., Riess, A. G., Kessler, R., Rest, A., Kirshner, R. P., Berger, E., Ortega, C. A., Foley, R. J., Chornock, R., Challis, P. J., Burgett, W. S., Chambers, K. C., Draper, P. W., Flewelling, H., Huber, M. E., Kaiser, N., Kudritzki, R.-P., Metcalfe, N., Wainscoat, R. J., and Waters, C. (2017). Measuring the Properties of Dark Energy with Photometrically Classified Pan-STARRS Supernovae. I. Systematic Uncertainty from Core-collapse Supernova Contamination. *The Astrophysical Journal*, 843(1):6.

Kafle, P. R., Sharma, S., Robotham, A. S. G., Pradhan, R. K., Guglielmo, M., Davies, L. J. M., and Driver, S. P. (2017). Galactic Googly: The Rotation–Metallicity Bias in the Inner Stellar Halo of the Milky Way. *Monthly Notices of the Royal Astronomical Society*, 470(3):2959–2971.

Kern, N. S., Liu, A., Parsons, A. R., Mesinger, A., and Greig, B. (2017). Emulating Simulations of Cosmic Dawn for 21 cm Power Spectrum Constraints on Cosmology, Reionization, and X-Ray Heating. *The Astrophysical Journal*, 848(1):23.

Khrykin, I. S., Hennawi, J. F., and McQuinn, M. (2017). The Thermal Proximity Effect: A New Probe of the He II Reionization History and Quasar Lifetime. *The Astrophysical Jour-*

nal, 838(2):96.

Kimura, M., Kato, T., Isogai, K., Tak, H., Shidatsu, M., Itoh, H., et al. (2017). Rapid Optical Variations Correlated with X-rays in the 2015 Second Outburst of V404 Cygni (GS 2023+338). *Monthly Notices of the Royal Astronomical Society*, 471(1):373–382.

Knežević, S., Läsker, R., van de Ven, G., Font, J., Raymond, J. C., Bailer-Jones, C. A. L., Beckman, J., Morlino, G., Ghavamian, P., Hughes, J. P., and Heng, K. (2017). Balmer Filaments in Tycho's Supernova Remnant: An Interplay between Cosmic-ray and Broad-neutral Precursors. *The Astrophysical Journal*, 846(2):167.

Köhlinger, F., Viola, M., Joachimi, B., Hoekstra, H., van Uitert, E., Hildebrandt, H., et al. (2017). KiDS-450: the Tomographic Weak Lensing Power Spectrum and Constraints on Cosmological Parameters. *Monthly Notices of the Royal Astronomical Society*, 471(4):4412–4435.

Kos, J. (2017). Spatial Structure of Several Diffuse Interstellar Band Carriers. *Monthly Notices of the Royal Astronomical Society*, 468(4):4255–4272.

Kovalenko, I. D., Stoica, R. S., and Emelyanov, N. V. (2017). Maximum a Posteriori Estimation Through Simulated Annealing for Binary Asteroid Orbit Determination. *Monthly Notices of the Royal Astronomical Society*, 471(4):4637–4647.

Küpper, A. H. W., Johnston, K. V., Mieske, S., Collins, M. L. M., and Tollerud, E. J. (2017). Exploding Satellites—The Tidal Debris of the Ultra-faint Dwarf Galaxy Hercules. *The Astrophysical Journal*, 834(2):112.

Leung, A. S., Acquaviva, V., Gawiser, E., Ciardullo, R., Komatsu, E., Malz, A. I., Zeimann, G. R., Bridge, J. S., Drory, N., Feldmeier, J. J., Finkelstein, S. L., Gebhardt, K., Gronwall, C., Hagen, A., Hill, G. J., and Schneider, D. P. (2017). Bayesian Redshift Classification of Emission-line Galaxies with Photometric Equivalent Widths. *The Astrophysical Journal*, 843(2):130.

Lund, M. N., Aguirre, V. S., Davies, G. R., Chaplin, W. J., Christensen-Dalsgaard, J., Houdek, G., et al. (2017). Standing on the Shoulders of Dwarfs: The Kepler Asteroseismic LEGACY Sample. I. Oscillation Mode Parameters. *The Astrophysical Journal*, 835(2):172.

MacDonald, R. J. and Madhusudhan, N. (2017). HD 209458b in New Light: Evidence of Nitrogen Chemistry, Patchy Clouds and Sub-Solar Water. *Monthly Notices of the Royal Astronomical Society*, 469(2):1979–1996.

MacLeod, C., Ivezić, Ž., Kochanek, C., Kozłowski, S., Kelly, B., Bullock, E., Kimball, A., Sesar, B., Westman, D., Brooks, K., Gibson, R., Becker, A. C., and de Vries, W. H. (2010). Modeling the Time Variability of SDSS Stripe 82 Quasars as a Damped Random Walk. *The Astrophysical Journal*, 721(2):1014.

Mandel, K. S., Scolnic, D. M., Shariff, H., Foley, R. J., and Kirshner, R. P. (2017). The Type Ia Supernova Color–Magnitude Relation and Host Galaxy Dust: A Simple Hierarchical Bayesian Model. *The Astrophysical Journal*, 842(2):93.

Martínez-García, E. E., González-Lópezlira, R. A., Magris C., G., and Bruzual A., G. (2017). Removing Biases in Resolved Stellar Mass Maps of Galaxy Disks through Successive Bayesian Marginalization. *The Astrophysical Journal*, 835(1):93.

Maund, J. R. (2017). The Resolved Stellar Populations Around 12 Type IIP Supernovae. *Monthly Notices of the Royal Astronomical Society*, 469(2):2202–2218.

McEwen, J. D., Feeney, S. M., Peiris, H. V., Wiaux, Y., Ringeval, C., and Bouchet, F. R. (2017). Wavelet-Bayesian Inference of Cosmic Strings Embedded in the Cosmic Microwave Background. *Monthly Notices of the Royal Astronomical Society*, 472(4):4081–4098.

Mejía-Narváez, A., Bruzual, G., C., G. M., Alcaniz, J. S., Benítez, N., Carneiro, S., et al. (2017). Galaxy Properties from J-PAS Narrow-Band Photometry. *Monthly Notices of the Royal Astronomical Society*, 471(4):4722–4746.

Molino, A., Benítez, N., Ascaso, B., Coe, D., Postman, M., Jouvel, S., et al. (2017). CLASH: Accurate Photometric Redshifts with 14 HST Bands in Massive Galaxy Cluster Cores. *Monthly Notices of the Royal Astronomical Society*, 470(1):95–113.

Montes-Solís, M. and Arregui, I. (2017). Comparison of Damping Mechanisms for Transverse Waves in Solar Coronal Loops. *The Astrophysical Journal*, 846(2):89.

Murphy, K. D., Nowak, M. A., and Marshall, H. L. (2017). The Nuclear X-Ray Emission-line Structure in NGC 2992 Revealed by Chandra - HETGS. *The Astrophysical Journal*, 840(2):120.

Natarajan, R. and Kass, R. E. (2000). Reference Bayesian Methods for Generalized Linear Mixed Models. *Journal of the American Statistical Association*, 95(449):227–237.

Oh, S.-H., Staveley-Smith, L., Spekkens, K., Kamphuis, P., and Koribalski, B. S. (2017). 2D Bayesian Automated Tilted-Ring Fitting of Disk Galaxies in Large Hi Galaxy Surveys: 2DBAT. *Monthly Notices of the Royal Astronomical Society*, page stx2304.

Oyarzún, G. A., Blanc, G. A., González, V., Mateo, M., and III, J. I. B. (2017). A Comprehensive Study of Lyα Emission in the High-redshift Galaxy Population. *The Astrophysical Journal*, 843(2):133.

Park, D., Barth, A. J., Woo, J.-H., Malkan, M. A., Treu, T., Bennert, V. N., Assef, R. J., and Pancoast, A. (2017). Extending the Calibration of C IV-based Single-epoch Black Hole Mass Estimators for Active Galactic Nuclei. *The Astrophysical Journal*, 839(2):93.

Patel, E., Besla, G., and Mandel, K. (2017). Orbits of Massive Satellite Galaxies II. Bayesian Estimates of the Milky Way and Andromeda Masses Using High-Precision Astrometry and Cosmological Simulations. *Monthly Notices of the Royal Astronomical Society*, 468(3):3428–3449.

Pihajoki, P. (2017). A Geometric Approach to Non-Linear Correlations with Intrinsic Scatter. *Monthly Notices of the Royal Astronomical Society*, 472(3):3407–3424.

Pinamonti, M., Sozzetti, A., Bonomo, A. S., and Damasso, M. (2017). Searching for Planetary Signals in Doppler Time Series: A Performance Evaluation of Tools for Periodogram Analysis. *Monthly Notices of the Royal Astronomical Society*, 468(4):3775–3784.

Pinsky, M. A. and Karlin, S. (2011). *An Introduction to Stochastic Modeling*. Academic Press, Cambridge, MA, USA.

Raithel, C. A., Özel, F., and Psaltis, D. (2017). From Neutron Star Observables to the Equation of State. II. Bayesian Inference of Equation of State Pressures. *The Astrophysical Journal*, 844(2):156.

Rodrigues, T. S., Bossini, D., Miglio, A., Girardi, L., Montalbán, J., Noels, A., Trabucchi, M., Coelho, H. R., and Marigo, P. (2017). Determining Stellar Parameters of Asteroseismic Targets: Going Beyond the Use of Scaling Relations. *Monthly Notices of the Royal Astronomical Society*, 467(2):1433–1448.

Salvato, M., Buchner, J., Budavári, T., Dwelly, T., Merloni, A., Brusa, M., Rau, A., Fotopoulou, S., and Nandra, K. (2017). Finding Counterparts for All-Sky X-Ray Surveys with Nway: A Bayesian Algorithm for Cross-Matching Multiple Catalogues. *Monthly Notices of the Royal Astronomical Society*, page stx2651.

Sampedro, L., Dias, W. S., Alfaro, E. J., Monteiro, H., and Molino, A. (2017). A Multimembership Catalogue for 1876 Open Clusters Using UCAC4 Data. *Monthly Notices of the Royal Astronomical Society*, 470(4):3937–3945.

Sathyanarayana Rao, M., Subrahmanyan, R., Shankar, N., and Chluba, J. (2017). Modeling the Radio Foreground for Detection of CMB Spectral Distortions from the Cosmic Dawn and the Epoch of Reionization. *The Astrophysical Journal*, 840(1):33.

Schellenberger, G. and Reiprich, T. H. (2017). HICOSMO: Cosmology with a Complete Sample of Galaxy Clusters II. Cosmological Results. *Monthly Notices of the Royal Astronomical Society*, 471(2):1370–1389.

Scherrer, B. and McKenzie, D. (2017). A Bayesian Approach to Period Searching in Solar Coronal Loops. *The Astrophysical Journal*, 837(1):24.

Sereno, M. and Ettori, S. (2017). CoMaLit-V. Mass Forecasting with Proxies: Method and Application to Weak Lensing Calibrated Samples. *Monthly Notices of the Royal Astronomical Society*, 468(3):3322–3341.

Si, S., van Dyk, D. A., von Hippel, T., Robinson, E., Webster, A., and Stenning, D. (2017). A Hierarchical Model for the Ages of Galactic Halo White Dwarfs. *Monthly Notices of the Royal Astronomical Society*, 468(4):4374–4388.

Silburt, A. and Rein, H. (2017). Resonant Structure, Formation and Stability of the Planetary System HD155358. *Monthly Notices of the Royal Astronomical Society*, 469(4):4613–4619.

Sliwa, K., Wilson, C. D., Matsushita, S., Peck, A. B., Petitpas, G. R., Saito, T., and Yun, M. (2017). Luminous Infrared Galaxies with the Submillimeter Array. V. Molecular Gas in Intermediate to Late-stage Mergers. *The Astrophysical Journal*, 840(1):8.

Solá, J., Gómez-Valent, A., and de Cruz Pérez, J. (2017). First Evidence of Running Cosmic Vacuum: Challenging the Concordance Model. *The Astrophysical Journal*, 836(1):43.

Tabatabaei, F. S., Schinnerer, E., Krause, M., Dumas, G., Meidt, S., Damas-Segovia, A., et al. (2017). The Radio Spectral Energy Distribution and Star-formation Rate Calibration in Galaxies. *The Astrophysical Journal*, 836(2):185.

Tak, H., Kelly, J., and Morris, C. N. (2017a). Rgbp: An R Package for Gaussian, Poisson, and Binomial Random Effects Models, with Frequency Coverage Evaluations. *Journal of Statistical Software*, 78(5):1–33.

Tak, H., Mandel, K., van Dyk, D. A., Kashyap, V. L., Meng, X.-L., and Siemiginowska, A. (2017b). Bayesian Estimates of Astronomical Time Delays between Gravitationally Lensed Stochastic Light Curves. *The Annals of Applied Statistics*, 11(3):1309–1348.

Tak, H. and Morris, C. N. (2017). Data-dependent Posterior Propriety of a Bayesian Beta-Binomial-Logit Model. *Bayesian Analysis*, 12(2):533–555.

Tanaka, M., Chiba, M., and Komiyama, Y. (2017). Resolved Stellar Streams around NGC4631 from a Subaru/Hyper Suprime-Cam Survey. *The Astrophysical Journal*, 842(2):127.

Vallisneri, M. and van Haasteren, R. (2017). Taming Outliers in Pulsar-Timing Data Sets with Hierarchical Likelihoods and Hamiltonian Sampling. *Monthly Notices of the Royal Astronomical Society*, 466(4):4954–4959.

Wagner-Kaiser, R., Sarajedini, A., von Hippel, T., Stenning, D. C., van Dyk, D. A., Jeffery, E., Robinson, E., Stein, N., Anderson, J., and Jefferys, W. H. (2017). The ACS Survey of Galactic Globular Clusters XIV. Bayesian Single-Population Analysis of 69 Globular Clusters. *Monthly Notices of the Royal Astronomical Society*, 468(1):1038–1055.

Wang, X., Jones, T. A., Treu, T., Morishita, T., Abramson, L. E., Brammer, G. B., et al. (2017). The Grism Lens-amplified Survey from Space (GLASS). X. Sub-kiloparsec Resolution Gasphase Metallicity Maps at Cosmic Noon behind the Hubble Frontier Fields Cluster MACS1149.6+2223. *The Astrophysical Journal*, 837(1):89.

Warren, H. P., Byers, J. M., and Crump, N. A. (2017). Sparse Bayesian Inference and the Temperature Structure of the Solar Corona. *The Astrophysical Journal*, 836(2):215.

Watkinson, C. A., Majumdar, S., Pritchard, J. R., and Mondal, R. (2017). A Fast Estimator for the Bispectrum and Beyond - A Practical Method for Measuring Non-Gaussianity in 21-cm Maps. *Monthly Notices of the Royal Astronomical Society*, 472(2):2436–2446.

Wilkinson, D. M., Maraston, C., Goddard, D., Thomas, D., and Parikh, T. (2017). firefly (Fitting IteRativEly For Likelihood analYsis): A Full Spectral Fitting Code. *Monthly Notices of the Royal Astronomical Society*, 472(4):4297–4326.

Yu, M. and Liu, Q.-J. (2017). On the Detection Probability of Neutron Star Glitches. *Monthly Notices of the Royal Astronomical Society*, 468(3):3031–3041.

Zevin, M., Pankow, C., Rodriguez, C. L., Sampson, L., Chase, E., Kalogera, V., and Rasio, F. A. (2017). Constraining Formation Models of Binary Black Holes with Gravitational-wave Observations. *The Astrophysical Journal*, 846(1):82.

This paper has been typeset from a TEX/LATEX file prepared by the author.