THE SOCIETY OF POPULATION ECOLOGY

CrossMark

Bayes estimates as an approximation to maximum likelihood estimates

Kohji Yamamura[1]

Abstract Ronald A. Fisher, who is the founder of maximum likelihood estimation (ML estimation), criticized the Bayes estimation of using a uniform prior distribution, because we can create estimates arbitrarily if we use Bayes estimation by changing the transformation used before the analysis. Thus, the Bayes estimates lack the scientific objectivity, especially when the amount of data is small. However, we can use the Bayes estimates as an approximation to the objective ML estimates if we use an appropriate transformation that makes the posterior distribution close to a normal distribution. One-to-one correspondence exists between a uniform prior distribution under a transformed scale and a non-uniform prior distribution under the original scale. For this reason, the Bayes estimation of ML estimates is essentially identical to the estimation using Jeffreys prior.

Keywords Empirical Jeffreys prior · Posterior distribution · Sika deer population · Skewness · State-space model · Transformation

✉ Kohji Yamamura
yamamura@affrc.go.jp

[1] National Institute for Agro-Environmental Sciences, 3-1-3 Kannondai, Tsukuba 305-8604, Japan

Introduction

Bayesian estimation contains a logical problem that is a subject of debate between Bayesian and non-Bayesian scientists. Recent articles on Bayesian inference sometimes portray the debate as 'Bayesian versus classical approach' (e.g., Ellison 2004; Clark 2005, 2007), although Bayesian inference that originates from Bayes (1763) is historically older than the classical approach. Fisher (1922) criticized the logical problem of Bayes estimation, and proposed the maximum likelihood estimation (ML estimation) that should replace Bayes estimation. Neyman (1934) later developed the concept of confidence intervals. The approach of Neyman-Pearson, which is called 'frequentist', is much different from the approach of Fisher, which is called 'Fisherian' (Efron 1998; Salsburg 2001; Hubbard and Bayarri 2003; Shibamura 2004). We can use the term 'non-Bayesian' to refer to both Fisherian and frequentist. Although many researchers collectively use the term frequentist to refer to all scientists that use the approach of Fisher and Neyman-Pearson (e.g., Dennis 2004), I prefer to use the terminology of Efron (1998) in this paper, because I will use the approach of Fisher while rejecting the approach of Neyman-Pearson in defining the confidence limits later in this paper. Neyman-Pearson proposed the statistical test to control the frequency of false decisions in a repeated sampling, but Fisher strongly criticized the Neyman-Pearson approach (e.g., Fisher 1973, p44, p95, p104) because such an endless series of repeated sampling will never take place actually except for industrial samplings and because the strength of the evidence is not to be measured by such frequency.

We should first strictly define the meanings of Bayes estimates. Several people call the estimates 'Bayes estimates' when they used Bayes theorem in their estimation.

However, this usage is causing a serious confusion because Bayes theorem is also commonly used in calculating ML estimates. The debate between Bayesian and non-Bayesian scientists mostly lies on the use of uniform prior distributions for parameters. Such uniform prior distributions were first used by Bayes (1763). Hence, I call the procedure of estimation that uses uniform prior distributions (or flat prior distributions) as 'Bayes (1763) estimation' in this paper to avoid the confusion.

Fisher (1922) criticized the Bayes (1763) estimation of using uniform prior distributions, because we can create estimates arbitrarily by changing the transformation before the analysis. The Bayes (1763) estimation lacks the scientific objectivity, especially when the amount of data is small. However, we can utilize the Bayes (1763) estimation as an approximation to the objective ML estimation if we use an appropriate transformation that makes the posterior distribution close to a normal distribution. In this paper, I illustrate it by the example of ML estimation of sika deer populations. I show that we can create various estimates by changing the transformation arbitrarily in the Bayes (1763) estimation, but that only one transformation is appropriate as an approximation to ML estimation. One-to-one correspondence exists between a uniform prior distribution under a transformed scale and a non-uniform prior distribution under the original scale. I show that the Bayes (1763) estimation for obtaining approximate ML estimates is essentially identical to the estimation that uses Jeffreys (1946) prior.

Fisher's criticism to Bayes (1763) estimates

Bayes (1763) considered a binomial sampling procedure in which a sample of size n is drawn from an infinite population of which a proportion (p) is of a certain kind such as 'successes' while the remaining proportion ($1-p$) is of another kind such as 'failures'. Bayes (1763) then calculated the posterior distribution of p by assuming a uniform distribution for the prior distribution of p. Fisher (1922, p325) says "it is, however, evidently extremely arbitrary" because "it is not even a unique solution." As an example, Fisher introduced another parameter, θ, where θ is defined by $\theta = \sin^{-1}(2p - 1)$. Fisher (1922, p325) says that "the quantity θ measures the degree of probability, just as well as p, and is even, for some purposes, the more suitable variable". If we assume a uniform distribution for θ, the distribution of p becomes non-uniform. If we assume a uniform distribution for p, on the other hand, the distribution of θ becomes non-uniform. Consequently, the Bayes (1763) estimate becomes different, depending on whether we assumed a uniform distribution for p or θ. We can generally create any estimates we like by changing the

transformation before the analysis. In contrast, if we use ML estimation that was proposed by Fisher (1922), the estimates are the same; ML estimates are invariant to the transformation such as $\theta = \sin^{-1}(2p - 1)$.

It should be noted that Fisher strongly recommended the use of prior knowledge, although he recommended the use of fiducial argument in the absence of prior knowledge. Fisher mentioned that "my method was appropriate to the absence of knowledge a priori" (Fisher 1973, p105) and "the application of the fiducial argument can only be made in the absence of such information a priori" (Fisher 1973, p54).

Example of arbitrary creation of estimates by Bayes (1763) estimation

I next estimate the abundance of sika deer *Cervus nippon* Temminck to illustrate how Bayes (1763) estimation can correctly create arbitrary estimates. The estimation of the number of sika deer is one of the major concerns in Hokkaido prefecture, Japan (Yamamura et al. 2008); these deer are causing severe damage to agriculture, forestry and natural vegetation. I estimate the number of sika deer in Sorachi subprefecture in Hokkaido by using the data of spotlight survey from 2007 to 2012; I use this small data set for the purpose of illustration. Spotlight survey has been conducted by Hokkaido prefecture annually between late October and early November, before the start of the hunting season. (Uno et al. 2006). Each survey route was about 10 km in length. Two observers holding spotlights (Q-Beam 160,000 candle-power; Brinkmann, Dallas, TX) searched for deer on both sides of the survey route from 19:00 to 20:00, riding in a vehicle (at a speed ranging from 20 to 40 km/h). The total number of deer detected along the routes was recorded.

I perform the harvest-based estimation using a simple state-space model. I use the univariate model of Yamamura et al. (2008). Let N_t be the population at t year. I assume the following state model.

$$N_t = \rho N_{t-1} - K_{t-1}, \tag{1}$$

where ρ is the average natural rate of increase per individual per year, and K_t is the number of harvested (killed) at t year. I do not consider the process error in this paper, because the preliminary analyses by Yamamura et al. (2008) indicated that the process error is sufficiently small in these data.

Spotlight survey of sika deer provides only the information about the relative abundance of sika deer, and hence Hokkaido prefecture is using 'population index' to describe the change in abundance. I use the population

index I_t defined by N_t/N_{2007} for $t = 2008, 2009,\ldots, 2012$. It is given by a linear form in a logarithmic scale.

$$\log_e(I_t) = \log_e(N_t) - \log_e(N_{2007}). \tag{2}$$

I first estimate the logarithmic population index $\log_e(I_t)$ by a generalized linear mixed model (GLMM) in which the log link and Poisson errors are used (Yamamura et al. 2008). Let $\log_e(\hat{\boldsymbol{I_t}})$ be the vector of the estimates of logarithmic population indices. Let $\log_e(\boldsymbol{I_t})$ be the vector of true logarithmic population indices. I consider the following observation model.

$$\log_e(\hat{\boldsymbol{I_t}}) = \log_e(\boldsymbol{I_t}) + \boldsymbol{e_t}, \quad \boldsymbol{e_t} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}), \tag{3}$$

where $\boldsymbol{e_t}$ is the vector of observation errors which follow a multivariate normal distribution with mean $\boldsymbol{0}$ and variance–covariance matrix $\boldsymbol{\Sigma}$. I assume that the observation errors in $\log_e(N_t)$ are mutually independent. Then, the estimates of $\log_e(I_t)$ have a correlation coefficient of 0.5 mutually, because the estimate of $\log_e(N_{2007})$ is repeatedly used in estimating $\log_e(I_t)$ for $t = 2008\text{–}2012$ by the form of Eq. 2. Hence, we should use the following variance–covariance matrix.

$$\boldsymbol{\Sigma} = \begin{bmatrix} 2\sigma^2 & \sigma^2 & \ldots & \sigma^2 \\ \sigma^2 & 2\sigma^2 & \ldots & \sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2 & \sigma^2 & \ldots & 2\sigma^2 \end{bmatrix}. \tag{4}$$

Estimability is one of the troublesome problems in constructing a complicated model (Lele et al. 2010). The parameter ρ, which is the average natural rate of increase per individual, is not estimable in this case. The spotlight survey indicates only the relative abundance of sika deer, I_t ($= N_t/N_{2007}$). The state model (Eq. 1) can be expressed by

$$(I_t/I_{t-1}) = \rho - (K_{t-1}/N_{t-1}). \tag{5}$$

For a given set of (I_t/I_{t-1}) and K_{t-1}, we have an infinite number of combinations of ρ and N_{t-1} that satisfy Eq. 5. Hence, we must estimate ρ beforehand from another field data. I adopt $\rho = 1.21$ that was observed in a population in Cape Shiretoko (Kaji et al. 2004).

I use the Box-Cox transformation which has a superior flexibility in transforming a variable defined in the range of $(0, \infty)$.

$$f(x, \lambda) = (x^\lambda - 1)/\lambda, \tag{6}$$

where λ is a fixed exponent. The transformation includes several simpler transformations such as logarithmic transformation ($\lambda = 0$), square root transformation ($\lambda = 0.5$), inverse transformation ($\lambda = -1$), and no transformation ($\lambda = 1$). The function for back-transformation is given by

$$g(u, \lambda) = (\lambda u + 1)^{(1/\lambda)}. \tag{7}$$

I use various exponents of Box-Cox transformation for each parameter. The exponent for N_{2007} is denoted by λ_N while the exponent for σ is denoted by λ_σ.

I obtain Bayes (1763) estimates of two unknown parameters, N_{2007} and σ, by using WinBUGS 1.4.3 which is the standard software for performing estimation using Markov chain Monte Carlo (MCMC) (Spiegelhalter et al. 2003). I set the possible range of N_{2007} to a sufficiently large range: from 1 thousand to 500 thousands. I set the possible range of σ to (0.01, 150). I generate a uniform random number U_N in the range of $[f(1, \lambda_N), f(500, \lambda_N)]$ for a given exponent λ_N. The random number is back-transformed by $g(U_N, \lambda_N)$ to yield a sample of N_{2007}. I generate a uniform random number U_σ in the range of $[f(0.01, \lambda_\sigma), f(150, \lambda_\sigma)]$ for a given exponent λ_σ. The random number is back-transformed by $g(U_\sigma, \lambda_\sigma)$ to yield a sample of σ. Then, I calculate the mean of the posterior distribution for N_{2007} and σ. I perform 100,000 iterations with a burn-in period of 10,000. The R-code is listed in Electronic Supplementary Material (ESM) which is available at Springer website. The latest version of ESM is available at http://cse.niaes.affrc.go.jp/yamamura/Empirical_Jeffreys.html.

Figure 1 indicates the influence of exponent λ_N of Box-Cox transformation on Bayes (1763) estimates of population. The quantity of λ_N was changed by a 0.5 interval. We see that the estimate of N_{2007} increases from 26.2 to 71.5 thousands with changing λ_N from -3 to 1.5. All these estimates are valid estimates of N_{2007} if we permit Bayes (1763) estimation, and hence we can adopt any estimate of
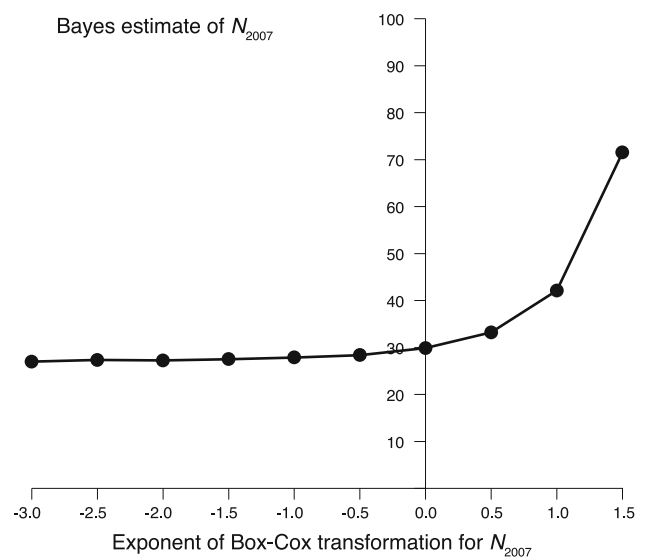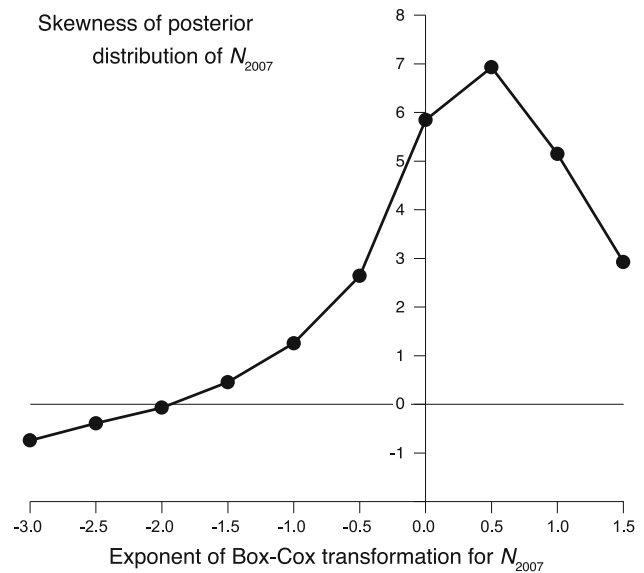

Fig. 1 Various Bayes (1763) estimates of population (N_{2007}) in thousands, obtained by modifying the exponent λ_N of the Box-Cox transformation for N_{2007}. The exponent λ_σ of the Box-Cox transformation for σ was fixed at -0.5

N_{2007}. If a researcher wants to obtain a larger estimate of population, for example, he will correctly create a large estimate of population by adopting a large λ_N such as 1.5.

## Use of Bayes (1763) estimates as approximate ML estimates

Bayes (1763) estimation permits an arbitrary creation of estimates as indicated by the above example. Such a creation will cause troubles in science. Hence, we should rather follow the recommendation given by R.A. Fisher; if we have no prior knowledge, we should calculate maximum likelihood estimates (ML estimates) instead of Bayes (1763) estimates. ML estimates of parameters are given by the parameter values that maximize the likelihood. However, we cannot easily calculate ML estimates if the model is highly complicated, while we can calculate Bayes (1763) estimates easily in most cases. Hence, we should utilize the Bayes (1763) estimates as an approximation to ML estimates. Marginal ML estimate is given by the mode of the posterior distribution if we use a uniform prior distribution. The numerical calculation of mode is relatively difficult than that of median or mean. Several methods are available for finding the mode. de Valpine (2003, 2004) used Monte Carlo kernel likelihood method. Lele et al. (2007) proposed data cloning method. Wood (2010) used quadratic regression for the sampled likelihood surface. I use another method in this paper. If we can transform the parameter space so that the posterior distribution is nearly symmetry, the mode coincides to the mean. Hence, we can use the mean as an approximate marginal ML estimate, that is, we can use the Bayes (1763) estimate as an approximate ML estimate. Furthermore, if we can use an appropriate transformation that makes the posterior distribution nearly the same symmetrical distribution such as a homoscedastic normal distribution, we can use the quantile of posterior distribution as the confidence limit. Most of the theory of ML estimation is based on the asymptotic normality. If we can make the posterior distribution nearly normal, therefore, we can use the asymptotic argument even for relatively small samples.

We should explore the procedure to find the appropriate transformation that makes the posterior distribution nearly normal. For a parameter that has the range of $(0, \infty)$, we will be able to find an appropriate transformation by changing the λ of Box-Cox transformation given by Eq. 6. We can use skewness of the posterior distribution to judge the symmetry; the distribution will be nearly symmetry if the skewness is nearly zero. We can further expect that the distribution will be nearly normal if the skewness is nearly zero. Figure 2 shows the influence of exponent λ_N on the skewness of the posterior distribution of N_{2007}. The skewness of the posterior distribution of N_{2007} became nearly zero if we use $\lambda_N = -2$. Hence,

Fig. 2 Various skewness of the posterior distribution of population (N_{2007}) obtained by modifying the exponent λ_N of the Box-Cox transformation for N_{2007}. The exponent λ_σ of the Box-Cox transformation for σ was fixed at -0.5

we can expect that the Bayes (1763) estimate of N_{2007} calculated by using $\lambda_N = -2$ is close to the marginal ML estimate. We should calculate the skewness for all parameters in the model. The influence of the exponent λ_σ on the skewness of the posterior distribution of σ is shown in Fig. 3. The skewness is nearly zero if we use $\lambda_\sigma = -0.5$. The combination of transformation, $\lambda_N = -2$ and $\lambda_\sigma = -0.5$, yields the
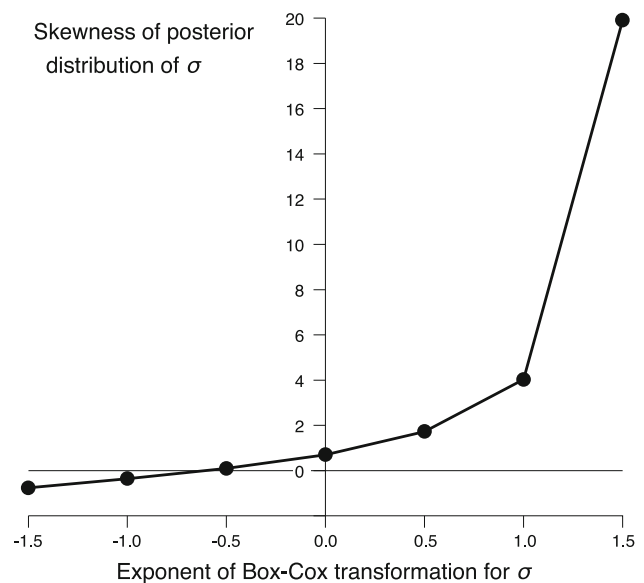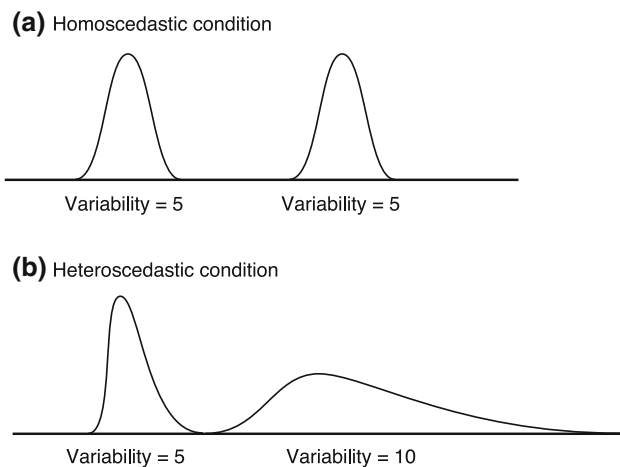

Fig. 3 Various skewness of the posterior distribution of the standard deviation of measurement error (σ) obtained by modifying the exponent λ_σ of the Box-Cox transformation for σ. The exponent λ_N of the Box-Cox transformation for N_{2007} was fixed at -2

approximate ML estimates of N_{2007} and σ in this case. Both of the mean and median coincide to the mode of distribution if the distribution is completely symmetry, but the median is less influenced by the asymmetric long tail. The median and quantiles are not influenced by the back-transformation, although the mean of back-transformed distribution is different from the back-transformed value of mean. Hence, we should use the median of back-transformed posterior distribution as the approximate ML estimate. The 2.5 and 97.5 % quantiles of the back-transformed posterior distribution can be used as the approximate lower 2.5 % confidence limit and upper 2.5 % confidence limit, respectively, in the definition of Fisher (1973, p69). In the current example, the estimates and the confidence limits are as follows: $\hat{N}_{2007} = 26.68\,(22.40 -35.37)$ thousands, $\hat{\sigma} = 0.091\,(0.052-0.218)$.

Transformation as an empirical Jeffreys prior

In the above example, I used skewness to empirically select the transformation to achieve the symmetry in the posterior distribution of parameters. This transformation is closely related to Jeffreys prior. Figure 4 schematically shows how the shape of distribution changes with the change of its mean. Skewness is generally closely related to the homoscedasticy. Skewness will be generally close to 0 if the distributions have the same variability irrespective of the mean as illustrated in Fig. 4a. If the variability of distribution increases toward right (or toward left), the distribution will be much skewed as illustrated in Fig. 4b.

(a) Homoscedastic condition

Variability = 5   Variability = 5

(b) Heteroscedastic condition

Variability = 5   Variability = 10

Fig. 4 Schematic illustration of the relation between skewness and homoscedasticity: **a** homoscedastic condition and **b** heteroscedastic condition. Under a heteroscedastic condition (*panel* **b**), the amount of dispersion changes along the axis in most cases. Hence, the amount of dispersion changes also within a distribution. Consequently, the distribution is inevitably skewed under a heteroscedastic condition as illustrated in *panel* **b**

Hence, we can find the appropriate transformation by finding a transformation that yields homoscedasticity.

Let us consider that we are estimating a single parameter μ, and that the variance of posterior distribution $L(\mu)$ is given by the function $g(\mu)$. The transformation that makes the variance constant is generally given by the following formula (e.g., Beall 1942; Bartlett 1947; Yamamura 1999).

$$f(x) = \int \frac{1}{\sqrt{g(x)}}\,dx. \tag{8}$$

On the other hand, the variance is roughly given by the inverse of Fisher information,

$$g(x) = 1 \left/ \left[-E\left(\frac{\partial^2 \log_e(L(x))}{\partial x^2} \right) \right] \right. . \tag{9}$$

Then, we obtain the appropriate transformation by substituting Eq. 9 to Eq. 8.

$$f(x) = \int \sqrt{-E\left(\frac{\partial^2 \log_e(L(x))}{\partial x^2} \right)}\,dx. \tag{10}$$

As Fisher (1922) discussed, the uniform distribution in a transformed scale corresponds to a non-uniform distribution in the original scale. Using a uniform prior distribution after performing a transformation $\int h(x)dx$ is identical to using a prior distribution $h(x)$. This relation is frequently used in generating a random number; that is, we can generate a random number that follows a distribution $h(x)$ by generating a uniform random number within (0, 1) and transforming it by the inverse function $(\int h(x)dx)^{-1}$ (see, for example, Marsaglia 1984). Therefore, the uniform distribution after the transformation by Eq. 10 is given by the following non-uniform distribution in the original scale.

$$h(x) = \sqrt{-E\left(\frac{\partial^2 \log_e(L(x))}{\partial x^2} \right)}. \tag{11}$$

This is exactly the non-informative prior of Jeffreys (1946). Thus, the uniform distribution after the transformation to yield skewness = 0 is nearly identical to Jeffreys prior in the original scale. Hence, we can call this transformation as 'empirical Jeffreys prior'. (The name was given by Dr. Mark Taper.)

Jeffreys (1946, 1961), in response to the criticism from Fisher, considered that the prior distributions should be invariant to transformation, implying that Bayesian estimates should be the same even if the parameters are transformed. Then, he proposed a prior distribution given by Eq. 11. This estimate is sometimes called 'objective Bayes'. The above argument about the empirical Jeffreys prior indicates that Jeffreys prior yields the invariant estimate simply because the estimate is essentially identical to the ML estimate which is invariant. Box and Tiao (1973) derived Eq. 11 from a different point of view. They

considered that uniform prior distributions should be used for parameters that have the following character called 'data translated': data influence only the location of the likelihood distribution of the parameter without influencing the curve of the likelihood distribution (Box and Tiao 1973, p26). Then, they derived Eq. 11. The transformation to yield a condition of 'data translated' is identical to the transformation to yield homoscedasticity, because the homoscedasticity is a necessary condition for the data translated condition. I used the homoscedasticity for different reasons; I used homoscedasticity to yield symmetrical posterior distributions.

Discussion

I discussed how we can utilize Bayes (1763) estimates as an approximation to the maximum likelihood estimates (ML estimates). We cannot directly obtain ML estimates if the model is so complicated that simulation experiments are required for calculating the predicted variables. Use of Bayes (1763) estimates as an approximation to ML estimates will be especially useful in such cases. Bayes (1763) estimates coincide to ML estimates if the amount of data is very large (Walker 1969). Hence, we can use Bayes (1763) estimates as an approximation to ML estimates, without any modification, if the amount of data is very large. However, problems arise in most cases, because the amount of data is usually not so large. I proposed the empirical Jeffreys prior that should be used in such cases; we should transform the space of parameters so that the posterior distribution becomes close to a normal distribution. We can use the skewness of posterior distribution in selecting an appropriate transformation. Then, the approximate ML estimate is given by the median of posterior distribution and the confidence limits are given by the quantiles of posterior distribution. The argument on the empirical Jeffreys prior indicates that the objective Bayesian estimates in the absence of prior knowledge are essentially identical to ML estimates, which also suggests that the Bayesian versus Fisherian debates are potentially converging.

Appropriate choice of transformation of parameters will be important also for pure ML estimation. There is a long tradition of trying to find parameter transformation in standard ML estimation (especially in nonlinear regression) so that the distribution of resulting ML estimates are nearly normal (or log-likelihood nearly paraboloid) and also nearly orthogonal, in order that (1) the convergence of estimates is improved and (2) the asymptotic confidence intervals give more accurate coverage (e.g., Ratkowsky 1983, 1990). In this paper, I considered Fisherian confidence limits instead of frequentist confidence intervals,

because we do not assume a repeated sampling in most cases, i.e., because we perform the experiment only once. Fisherian confidence limits (given by the inverse of test having a fixed probability of erroneous rejection) and frequentist confidence intervals (given by the interval having a fixed frequency of coverage) are numerically the same if the distribution is continuous, but the interpretation is very different (Neyman 1934, 1935; Fisher 1973).

Several other methods are also available for calculating confidence limits from Bayes (1763) estimation. If we use data cloning without transformation (by 50 clones) in estimating the sika deer population of Sorachi subprefecture via dclone package (Sólymos 2010), the estimates (with the lower 2.5 % and upper 2.5 % confidence limits) were as follows: $\hat{N}_{2007} = 26.67\,(22.47{-}30.88)$ thousands, $\hat{\sigma} = 0.081\,(0.031{-}0.132)$. R-code for this calculation and the comparison of $\hat{\sigma}$ are shown in ESM. The estimates were nearly the same as those obtained by using empirical Jeffreys prior. However, the estimates of confidence limits obtained by using data cloning were quite different from those obtained by using empirical Jeffreys prior. In data cloning method, as well as quadratic regression method, the estimated confidence limits become unrealistically symmetrical around ML estimates, because the calculation of confidence limits is based on the asymptotic normality in these methods. In contrast, the empirical Jeffreys prior yields asymmetrical confidence limits that correctly correspond to the inverse of test in the definition of Fisherian confidence limits (Fisher 1973, p69).

When we use empirical Jeffreys prior for a complicated model that contains many parameters, the procedure of estimation may become troublesome because we must find the appropriate transformation empirically for each of the parameters, although the computation time will be smaller than the corresponding data cloning method. If we confine the class of transformation into Box-Cox transformation, the exponent (λ in Eq. 6) that yields zero skewness may be automatically searched in each parameter by using some kind of efficient algorithm. A combination between empirical Jeffreys prior and data cloning may be also useful in estimating the parameters of complicated models; appropriate transformation will reduce the required number of clones in applying data cloning method. We cannot use Eq. 6 for binomial parameters such as the probability of occurrence. Other types of power transformation, that were indicated by Aranda-Ordaz (1981) and Guerrero and Johnson (1982), will be available in such cases.

We should reconsider about why most people including Bayes (1763) are liable to use uniform or flat distributions as prior distributions when they have no prior knowledge. Use of uniform distributions will be closely related to the Laplace definition of probability in which the true

probability is defined as the components that we cannot predict by a model that includes all knowledge we can use (Laplace 1825). In the Laplace definition of probability, we must assign the same probability to several distinct events if we have no additional knowledge. For example, let us imagine a field survey of the number of individuals; we observe a proportion p of the entire field, and record the observed number of individuals. If we have no knowledge about the spatial position of individuals, we must assume that each individual exists at each point of the field by the same possibility; that is, we must assume a uniform distribution for the spatial position of each individual. Consequently, the number of observation of each individual follows a Bernoulli distribution with the probability of occurrence p. Thus, we already used a uniform distribution in deriving the Bernoulli distribution or the binomial distribution. Bayes (1763) considered a binary situation: a certain proportion (p) of an infinite population is of a certain kind (e.g., successes) while the remaining proportion ($1-p$) is of another kind (e.g., failures). Then, he assumed a uniform distribution for the proportion (p), but the uniform distribution seems to be used duplicatively in Bayes (1763); that is, he potentially assumed a uniform distribution about which individual in the population is selected and he again assumed a uniform distribution for p. The former use of uniform distribution is correct in the Laplace definition of probability, but the latter use of uniform distribution is incorrect. We must assign equal probabilities for the components that we cannot predict by a model that includes all knowledge we can use. On the other hand, we cannot use uniform or flat distributions for the parameters that constitute our knowledge itself. We should not confuse these quite different things. If we use uniform or flat distributions for unknown parameters, such a use may have validity only as a convenient procedure to obtain ML estimates approximately.

Kolmogorov definition of probability (i.e., the mathematical definition of probability by Kolmogorov 1933) is currently adopted in most of the statistical textbooks, but Laplace definition of probability should be more reexamined when we discuss several debates on statistical problems. Yamamura (2014) proposed R_D criterion to evaluate the predictive ability of ecological models by using Laplace definition of probability. The predictive ability of models seems to be appropriately defined only if we use Laplace definition of probability. The R programs and SAS programs for calculating R_D criterion are placed in the following website with several supplementary materials: http://cse.niaes.affrc.go.jp/yamamura/RD_criterion_program.html. The fundamental debates between Bayesian and non-Bayesian scientists may be also solved to some extent if we appropriately use the philosophical definition of Laplace probability.

References

Aranda-Ordaz FJ (1981) On two families of transformations to additivity for binary response data. Biometrika 68:357–363

Bartlett MS (1947) The use of transformations. Biometrics 3:39–52

Bayes T (1763) An essay towards solving a problem in the doctrine of chances. Philos Trans R Soc 53:370–418

Beall G (1942) The transformation of data from entomological field experiments so that the analysis of variance becomes applicable. Biometrika 32:243–262

Box GEP, Tiao GC (1973) Bayesian inference in statistical analysis. Wiley, New York

Clark JS (2005) Why environmental scientists are becoming Bayesians. Ecol Lett 8:2–14

Clark JS (2007) Models for ecological data: an introduction. Princeton University Press, Princeton

de Valpine P (2003) Better inferences from population-dynamics experiments using Monte Carlo state-space likelihood methods. Ecology 84:3064–3077

de Valpine P (2004) Monte Carlo state-space likelihoods by weighted posterior kernel density estimation. J Am Stat Assoc 99:523–535

Dennis B (2004) Statistics and the scientific method in ecology (with commentary). In: Taper ML, Lele SR (eds) The nature of scientific evidence: statistical, philosophical, and empirical considerations. University of Chicago Press, Chicago, pp 327–378

Efron B (1998) R. A. Fisher in the 21st century. Stat Sci 13:95–114

Ellison AM (2004) Bayesian inference in ecology. Ecol Lett 7:509–520

Fisher RA (1922) On the mathematical foundations of theoretical statistics. Philos Trans R Soc A Math Phys Sci 222:309–368

Fisher RA (1973) Statistical methods and scientific inference, 3rd edn. Hafner Press, New York

Guerrero VM, Johnson RA (1982) Use of the Box-Cox transformation with binary response models. Biometrika 69:309–314

Hubbard R, Bayarri MJ (2003) Confusion over measures of evidence (p's) versus errors (α's) in classical statistical testing. Am Stat 57:171–178

Jeffreys H (1946) An invariant form for the prior probability in estimation problems. Proc R Soc A Math Phys Sci 186:453–461

Jeffreys H (1961) Theory of probability, 3rd edn. Oxford University Press, Oxford

Kaji K, Okada H, Yamanaka M, Matsuda H, Yabe T (2004) Irruption of a colonizing sika deer population. J Wildl Manage 68:889–899

Kolmogorov AN (1933) Foundations of the theory of probability (Translated from the 1st German edition of 1933 by N. Morrison, 1956), 2nd English edn. Chelsea Publishing Company, New York

Laplace PS (1825) A philosophical essay on probabilities (Translated from the fifth French edition of 1825 by Andrew I. Dale, 1995). Springer, New York

Lele SR, Dennis B, Lutscher F (2007) Data cloning: easy maximum likelihood estimation for complex ecological models using

Bayesian Markov chain Monte Carlo methods. Ecol Lett 10:551–563

Lele SR, Nadeem K, Schmuland B (2010) Estimability and likelihood inference for generalized linear mixed models using data cloning. J Am Stat Assoc 105:1617–1625

Marsaglia G (1984) The exact-approximation method for generating random variables in a computer. J Am Stat Assoc 79:218–221

Neyman J (1934) On the two different aspects of the representative method. J R Stat Soc 97:558–606

Neyman J (1935) On the problem of confidence intervals. Ann Math Stat 6:111–116

Ratkowsky DA (1983) Nonlinear regression modeling: a unified practical approach. Dekker, New York

Ratkowsky DA (1990) Handbook of nonlinear regression models. Dekker, New York

Salsburg D (2001) The lady tasting tea: how statistics revolutionized science in the twentieth century. Owl Books, New York

Shibamura R (2004) Statistical theory of R. A. Fisher. Kyushu University Press, Fukuoka (in Japanese)

Sólymos P (2010) dclone: data cloning in R. R J 2:29–37

Spiegelhalter DJ, Thomas A, Best N, Lunn D (2003) WinBUGS user manual, version 1.4. MRC Biostatistics Unit, Cambridge

Uno H, Kaji K, Saitoh T, Matsuda H, Hirakawa H, Yamamura K, Tamada K (2006) Evaluation of relative density indices for sika deer in eastern Hokkaido, Japan. Ecol Res 21:624–632

Walker AM (1969) On the asymptotic behaviour of posterior distributions. J R Stat Soc B 31:80–88

Wood SN (2010) Statistical inference for noisy nonlinear ecological dynamic systems. Nature 466:1102–1104

Yamamura K (1999) Transformation using $(x + 0.5)$ to stabilize the variance of populations. Res Popul Ecol 41:229–234

Yamamura K (2014) Estimation of the predictive ability of ecological models. Commun Stat Simul Comput. doi:10.1080/03610918.2014.889161

Yamamura K, Matsuda H, Yokomizo H, Kaji K, Uno H, Tamada K, Kurumada T, Saitoh T, Hirakawa H (2008) Harvest-based Bayesian estimation of sika deer populations using state-space models. Popul Ecol 50:131–144