Philosophy of Statistical Sciences: The Roles of Mathematics and Statistical Models in Estimation and Other Inductive Inferences

Kenichiro Shimatani*

Abstract

Statistical models have been used to test scientific hypotheses in ecological studies. The use of statistical model is critical when crucial quantities of interest are not directly measurable but statistically estimable from the acquired data. Both mathematical models and statistical models belong to mathematics, nevertheless, their roles in scientific inferences differ. A mathematical model quantitatively expresses a qualitatively expressed scientific hypothesis from which a scientist can deductively derive quantitative predictions. On the other hand, a statistical model mathematically expresses a stochastic data-generating process which allows the analyst to connect a mathematical model with data through probability distributions. Compared with mathematical models, statistical models have been insufficiently examined by philosophers, and philosophical analyses of statistical issues have focused on classical statistical tests and Bayesian inferences that are distant from up-to-date Bayesian statistics. Further, some statistical terminologies are differently used between philosophers and statisticians. Interdisciplinary studies between philosophy and statistical sciences such as ecology should consider the implications of modern statistical approaches.

Key words: Bayesian statistics, ecology, information criterion, likelihood, state-space model

1. Introduction

Statistical issues have been intensively examined in philosophical journals (e.g., Dowe et al. 2007, Henderson et al. 2010, Sprenger 2013, Rochefort-Maranda 2016, Steele and Werndl 2016); additionally, philosophical issues have been discussed in statistical papers (e.g., Gelman and Shalizi 2013, Yamamura 2016, Gelman and Hennig 2017). Philosophical issues regarding statistics are also discussed in scientific journals

* The Institute of Statistical Mathematics
E-mail: shimatani@ism.ac.jp
of a specific discipline such as ecology (e.g., Taper and Ponciano 2016, Lele 2020a, b). In this paper, the sciences in which modern statistical techniques are actively applied are called statistical science. Unfortunately, a philosopher and a statistician may associate completely different concepts with statistical terminology. For example, regarding statistics, a philosopher of science might think of a significance test, p-value, and a null hypothesis, and the term Bayes may remind them of the degree of belief in a hypothesis, prior and posterior probabilities, and discussion about subjective and objective probabilities. Conversely, for modern statistical concepts, a statistician tends to think of an information criterion, machine learning, Bayesian estimation, and the term Bayes reminds them of prior and posterior distributions for parameter estimates, Markov chain Monte Carlo (MCMC), and particle filters. In such situations, it is nearly impossible for both philosophers and statisticians to successfully communicate with each other, and for statistical scientists to be able to participate in philosophical discussions and vice versa.

For example, in the paper titled “Philosophy and the Practice of Bayesian Statistics,” citing philosophical papers, Gelman and Shalizi (2013) discussed philosophical issues in statistical science, and this paper is repeatedly cited in philosophical journals, although most of them noted this paper in footnotes (Vassend 2019, Douven 2020) or gave minor roles (e.g. Sprenger 2013, Sprenger 2014). In the paper titled “Beyond Subjective and Objective in Statistics,” although Gelman and Hennig (2017) discussed subjectivity vs. objectivity in statistical science and philosophers offered comments to the journal, Sprenger (2017) stated that this paper presented the “practitioner’s perspective,” not “philosophical perspectives by statistical scientists.” Conversely, although Gelman and Shalizi (2013) dedicated a whole section to philosophical issues, the main topics discussed are related to Thomas Kuhn and Karl Popper, and not to modern philosophical studies.

The objective of this study is to link the perspectives of the academic disciplines of philosophy and statistics using an example in the scientific discipline of population ecology.

In modern sciences, one of the primary objectives of statistics is to estimate unknown quantities of interests (Bernardo 2011), and statistical estimation is used for both applied and scientific purposes (behavioralistic and verbalistic outlook, cf. Savage 1972, Niiniluoto 2012). For applied purposes, the roles of statistical estimation are clear; decision-making should be done based on predictions derived from statistics. On the other hand, they are less clear for purely scientific purposes, and a concrete example in ecology helps us understand the basic roles of statistics. Consider the following example. If for some migratory bird species, many individuals use a large wintering site, and there are several smaller wintering sites. Observing the large wintering site, we tend to believe that this site contributes to the maintenance of the bird population (Furrer and Pasinelli 2016). However, if many immigrants
come to the large site from the smaller sites, this belief is incorrect. We should know the number of emigrants from the large wintering site and the number of immigrants to that site.

Two common ecological data collection methods for bird population dynamics are counting the number of individuals and conducting resight censuses over years after birds were captured, marked, and released (explained in Section 2.1). Although neither directly shows the numbers of emigrants and immigrants, relating the two variables with available data in a mathematical framework called a statistical model (introduced in Section 2.2) may give us their estimates. A statistical model plays a crucial role when comprehending some quantity or quantities is indispensable but these quantities are not directly measurable.

In this paper, I introduce and examine the following:

1. Concepts of a statistical model for estimation.
2. Roles and position of statistics as deduction in an inductive inference (throughout the paper, induction means inferences based on data, experiences, knowledge, and so on).
3. Differences between roles of mathematics in a mathematical model and a statistical model.

Mathematics has been used intensively from ancient times in physics. Since the late 20th century, when mathematical biology, mathematical economics, and other mathematical sciences became popular, mathematics has been commonly used in various scientific disciplines. In such disciplines, a qualitatively described or verbal hypothesis can be quantitatively written in the form of a mathematical model, which allows us to numerically test the hypothesis through experiments and measurements (discussed in Section 3). Although mathematical models and statistics are both mathematical, their roles in testing a hypothesis differ completely.

As far as relying on induction from data, scientific inferences are inevitably limited in applicability, as is indicated by David Hume. Compared with inferences that rely solely on personal experiences and knowledge, the use of systematically collected data should contribute to improving the reliability of an inductive inference. If statistics are additionally employed to analyze data, the inference is expected to be further improved. This expectation originates from the belief that statistics belong to mathematics and that mathematics (deduction) should complement inductive inferences.

The remainder of this paper is organized as follows. Section 2 introduces an ecological example to demonstrate a statistical model for estimation, presents an explanation of the concepts of a statistical model and a hierarchical Bayesian model. Section 3 summarizes the inferences using a statistical model for estimation in hypothesis testing. Sections 4 and 5 discuss the confusion between philosophical and statistical literatures regarding statistical terminology. Section 6 explores some re-
cent developments in statistics for complementing inductive inferences. The final section concludes the paper by presenting two proposals regarding the philosophy of statistical sciences.

2. Statistical model for estimation

This section explains how emigration and immigration rates, though not directly estimable, are indirectly estimable through a statistical model. Detailed explanations regarding statistical concepts can be found in Konishi and Kitagawa (2008), Link and Baker (2010), Kéris and Schaub (2012), and Antoine et al. (2017).

2.1. Estimation of survival and resight probabilities from resight data

Consider the following example. Suppose \( n \) individuals were captured, marked, and released. The next year, a resight census was conducted. If a marked individual resighted, this information directly shows the survival of that individual, whereas if the individual not resighted, it could be due to either death or survival with a failure of resight. By relying solely on the resight data, we cannot distinguish between the two cases. Incorporating a statistical model and time-series data resolves this problem as follows:

Let \( s \) and \( p \) be the (unknown) survival and resight probabilities, respectively. Let us consider that after the capture-mark-release, resight censuses were conducted twice. Consequently, there are four patterns in the resight data, as shown in Table 1, and we can derive the probability that each pattern is obtained (Table 1 and Fig. 1). If \( n_i \) individuals showed pattern \( i \) (\( i = a, b, c, d, n_a + n_b + n_c + n_d = n \)), assuming the independence of data, the probability that we obtain that data is given by the product of the probability of each of the patterns as follows:

\[
P((n_a, n_b, n_c, n_d) | (s, p)) = (spsp)^{n_a} \cdot (s(1-p)sp)^{n_b} \cdot (sp(s(1-p)+1-s)^{n_c} \cdot \{1-s+s(1-p)(s(1-p)+1-s)\}^{n_d},
\]

in which in general, when data are denoted by \( x \) and a set of unknown parameters by \( w \), the conditional probability of obtaining data \( x \) when the parameter is fixed at \( w \) is written as \( P(x|w) \). When data is given, \( x \) is fixed while parameter \( w \) is not, thus, interpreting \( P(x|w) \) as a function of \( w \), it is written as \( L(w|x) \) and called likelihood (function).

\[
L((s,p)|(n_a, n_b, n_c, n_d)) = (spsp)^{n_a} \cdot (s(1-p)sp)^{n_b} \cdot (sp(s(1-p)+1-s)^{n_c} \cdot \{1-s+s(1-p)(s(1-p)+1-s)\}^{n_d}
\]

— 8 —
Table 1  Four patterns of resight data when censuses were conducted twice after capture-mark-release, and the probability that each occurred.  and  indicate resight and not resight, respectively.  and  indicate the survival and resight probability, respectively.

| Pattern | Census | Probability | Number of individuals |
|---------|--------|-------------|-----------------------|
| a       | ○      | ○           | \(spsp\)              | \(n_a\)                |
| b       | ×      | ○           | \(s(1-p)sp\)          | \(n_b\)                |
| c       | ○      | ×           | \(sp(s(1-p) + (1-s))\)| \(n_c\)                |
| d       | ×      | ×           | \(1 - s + s(1-p)\{s(1-p) + 1 - s\}\) | \(n_d\)                |

Greater likelihood means a greater possibility that we obtain that data; thus, the parameter set that makes the likelihood greater can be interpreted as more likely. The likelihood is a measure of the support for a given set of parameters in the data. Probability and Likelihood are two sides of the same coin with the formula in (1) and (2) serving as the right hand side of both the probability and likelihood equations. The set of parameters that maximizes the likelihood is called the maximum likelihood estimate. Although the resight data directly indicate neither the survival nor resight probabilities, the application of a statistical model enables us to obtain the most likely estimates.

2.2. Statistical model

In the mathematical framework shown in Figure 1, we assumed that the survival and resight of a bird independently followed the Bernoulli distribution, and that data were randomly generated by that probability distribution; because of these assumptions, the probabilities are calculated as shown in Table 1 and Figure 1, and the likelihood is given by equation (2). In other words, Figure 1 illustrates the data-generating process in Table 1, in which probability densities play central roles in capturing and mathematically expressing stochastic uncertainty. This study calls such mathematical frameworks a statistical model, and if the primary objective of such a model is to obtain estimates of unknown quantities, it is called a statistical model for estimation.

2.3. Hierarchical model

The processes of the upper parts in each of (a) – (d) in Figure 1 (e.g. Release \(\rightarrow\) Survive, Survive \(\rightarrow\) Died) indicate an ecological process, while the processes in each lower parts (e.g. Survive \(\rightarrow\) ○, Survive \(\rightarrow\) ×) express human activities (resight census), under mathematically simplified assumptions such as the constant survival and resight probabilities regardless of individuals, years, environmental conditions, and other conditions. This mathematical framework has a hierarchical structure, both are temporal, and the data employed are time-series data. The whole hierarchical
model is called a state-space model, the upper process a state model and the lower process an observation model (Note that the terminology “hierarchical model” is used in different meanings depending on literatures, c.f. Henderson et al. 2010, Link and Baker 2010, Kéry and Schaub 2012).

The main objective of the application of the state-space model in this example is to obtain estimates of unknown parameters in the state model (the survival probability $s$); to achieve this goal, we have to additionally estimate parameters of human activities (the resight probability $p$).

### 2.4. Statistical model for estimating emigration and immigration

To construct a statistical model that enables us to estimate emigration and immigration, we need a more complex state-space model. Below is an outline of the statistical model of Weegman et al. (2016).

In the state model, we assume that the bird population consists of two subpopulations, the one comprising individuals that use the large wintering site and the other

---

**Figure 1** The illustration of a statistical model for estimating the survival and resight probabilities. The quantity labeling each arrow indicates the probability with which the transition occurs. $\bigcirc$ and $\times$ indicate resight and not resight, respectively. The thin arrows express the ecological process (survival or death) in the state model, while the vertical shorter dashed arrows show human activities in the observation model. The vertical longer dashed arrows indicate that it occurred with probability 1. These are separately illustrated for patterns (a) – (d) in Table 1. Note in pattern c and d, the same observation (not resight) can occur under of either potential states.
comprising all other individuals that use smaller sites. Each subpopulation is divided into two age classes: juveniles and adults. Thus, the bird population comprises five stages: two age classes of the two subpopulations and a stage comprising dead individuals (Figure 2). In population dynamics models, stages generally correspond to the growing stages of an individual, but here, the concept of stage is extended to a population that consists of two subpopulations.

Let $n_t^i$ be the number of individuals of stage $i$ in year $t$ ($i = 1, \ldots, 5$) and $\mathbf{n}^t$ the vertical vector whose $i$-th component is $n_t^i$. Transitions occur between the stages. For example, if an individual in stage $i$ died ($i = 1, \ldots, 4$), it moves to stage 5. If a juvenile using the large wintering site (stage 1) survived and returned to the large site, it moves to stage 2. If it emigrated to a smaller site, the transition is from stages 1 to 4. Let the transition probability that an individual in stage $i$ moves to stage $j$ be $T_{ji}$ (Figure 2). If a transition never occurs from $i$ to $j$, then $T_{ji} = 0$.

In this state model, $T_{23} + T_{24}$ and $T_{41} + T_{42}$ indicate the immigration and emigration rates per year, respectively. Thus, the hypothesis that “the large wintering site contributes to the maintenance of the population,” can be mathematically expressed.
as $T_{23} + T_{24} > T_{41} + T_{42}$.

Let $R_{ji}$ be the mean number of juveniles in stage $j$ reproduced by an adult in stage $i$ (only $R_{12}$ and $R_{34}$ are non-zero). The vector in year $t+1$, $\mathbf{n}^{t+1}$, can be written by the product of the matrix $\mathbf{P} = \mathbf{T} + \mathbf{R}$ ($\mathbf{T} = (T_{ji})$, $\mathbf{R} = (R_{ji})$) and vector $\mathbf{n}^t$ as $\mathbf{n}^{t+1} = \mathbf{P}\mathbf{n}^t$, and we use this model (matrix population model, c.f. Caswell 2000) as a state model. The number of individuals ($n_i^t$), transition probabilities ($T_{ji}$), and mean number of juveniles ($R_{ji}$) are unknown parameters in the state model.

The data is composed of four sets: (1) individual counts, separately between adults and juveniles, in the large wintering site, (2) the resight data of marked individuals in the large site, (3) detection of a dead individual with a mark (this clear evidence of mortality complements the resight data), and (4) detection of a marked individual in a smaller site (evidence of emigration). Thus, four observation models are required. The observation model for data (2) can be given by modifications of the model in Figure 1, while the other three require other mathematical equations. For mathematical details, see Weegman et al. (2016) and references therein.

Combining the state and observation models, we can derive the likelihood and probability of obtaining the four datasets. Although none of the data directly indicates the emigration and immigration rates, the count data should reflect (1) the survivors for which the survival probability is partially estimable from the resight data as explained in Section 2.1, (2) the increment by immigration, and (3) the reduction by emigration. Hence, combining the four datasets in the statistical model may enable us to obtain their estimates indirectly.

2.5. Formulation under Bayesian statistics

When a statistical model has a complex hierarchical structure and many unknown parameters, finding the maximum likelihood estimates is nearly impossible. In such cases, Bayesian statistics helps to resolve this problem.

First, a prior distribution is arbitrarily or/and subjectively given to the unknown parameters. Let $\mathbf{x}$ be data, $\mathbf{w}$ be a set of parameters, and $f(\mathbf{w})$ be the probability density function of the prior distribution. When we obtain data, by applying Bayes' theorem, the prior distribution is updated to a posterior distribution as follows:

$$f(\mathbf{w}|\mathbf{x}) = P(\mathbf{x} | \mathbf{w}) f(\mathbf{w}) / \int_D P(\mathbf{x} | \mathbf{w}) f(\mathbf{w}) d\mathbf{w}$$

(3)

where $f(\mathbf{w}|\mathbf{x})$ is the probability density function of the posterior distribution, and $D$ indicates the domain of the parameters. Replacing $P(\mathbf{x} | \mathbf{w})$ with likelihood $L(\mathbf{w} | \mathbf{x})$, we have

$$f(\mathbf{w}|\mathbf{x}) = L(\mathbf{w} | \mathbf{x}) f(\mathbf{w}) / \int_D L(\mathbf{w} | \mathbf{x}) f(\mathbf{w}) d\mathbf{w}.$$  

(4)

In this equation, $f(\mathbf{w}|\mathbf{x})$ is proportional to the product of the prior $f(\mathbf{w})$ and the
likelihood $L(w|x)$. Thus, in general, $f(w|x)$ takes high values when $L(w|x)$ is high unless $f(w)$ is extremely low, and the variance of the posterior distribution reflects uncertainty owing to the limitations of the data. Hence, we may interpret $f(w|x)$ to express how much likely is a set of parameters $w$. This study calls such mathematical frameworks a *hierarchical Bayesian model* (as noted in Section 2.3, similar terminology is used for different concepts depending on the literatures).

As $f(w|x)$ is not often explicitly written by well-known functions, we cannot know when $f(w|x)$ is high. Instead, we may produce random samples from the posterior distribution by using an algorithm such as MCMC. In general, many samples are generated around $w$ if $f(w|x)$ is high, whereas very few samples are generated near $w$ if $f(w|x)$ is very low. Thus, random samples can be interpreted as a set of likely estimates of the unknown parameters. We can also derive the mean and the standard deviation of the posterior distribution by taking the sample mean and the sample standard deviation over random samples, respectively (Monte Carlo method), which are useful for summarizing likely estimates.

The state-space model in Figure 2 does not provide a single set of values as the most likely estimates of emigration and immigration rates, but allows us to obtain their posterior distribution. If the posterior distribution of the emigration rate exceeds that of immigration, we may conclude that the large wintering site contributes to the maintenance of the population. In the case study in Weegman et al. (2016), the immigration tended to show a greater posterior distribution. Applying a similar statistical model, Fay et al. (2019) examined the effects of nest boxes on a kestrel population, in which the area with nest boxes corresponded to the large wintering site and other habitats to the smaller sites.

Finally, it is important to note that these studies were carried out for applied purposes to induce useful ecological implications for conservation biology. However, to achieve the applied purposes, a scientific hypothesis, whether emigration rate is greater than immigration, needs to be tested. Presently, for many ecological studies with applied purposes, testing a hypothesis is required to achieve this objective and the above examples are not exceptional.

### 3. Inference using statistical model for estimation

In this section, we examine inference processes when a hierarchical Bayesian model is applied.

If a hypothesis is qualitative, we first express it in the form of a mathematical model, and deductively (mathematically) derive a prediction. Subsequently, depending on whether or not the prediction is directly measurable, the inference splits into two. If the prediction is directly measurable, the process of carrying out the inference is similar to a simplified scheme of the Popperian falsification (hypothetico-deductive
testing, cf. Brittan and Bandyopadhway, 2019), as illustrated in Fig. 3b. Namely, by conducting an experiment and measurements, when we obtain an observed value, if it is not equal to the prediction, the hypothesis is falsified. On the other hand, if the prediction is not directly measurable but indirectly estimable by a statistical model, an inference is carried out by using the method demonstrated in the previous section, which is illustrated in Figure 3a. Although the two inferences contain deductions and look similar, there are essential differences.

First, as shown in Figure 3a, because a prediction is not directly measurable, we must construct a statistical model that connects the mathematical model and available data. Often, the same mathematical model that expresses the qualitative hypothesis is used in a state model, and we additionally prepare an observation model. Second, when the hierarchical model is complex, formulating the model under Bayesian statistics requires an additional artifact, a prior distribution. Third, when a posterior distribution is not expressed by well-known functions, using a deductive algorithm (this means that the computations follow a mathematically established
process), we generate random samples from the posterior distribution. Finally, comparing the prediction with the posterior distribution, we consider the hypothesis by induction (this final part is discussed again in Section 6).

If the prediction is not covered by the posterior distribution (belonging to a low-density area of random samples), the hypothesis (more precisely, its expression by the mathematical model) seems to be false; however, we should not easily falsify the hypothesis, but should examine the following possibilities.

1. The hypothesis is wrong,
2. The hypothesis is correct (more precisely, is close to the truth, captures some aspects of the truth, and so on), but because the observation model is inadequate, we got wrong estimates, which masked the suitability of the hypothesis.
3. The hypothesis is correct, but the prior distribution strongly affected the posterior distribution, resulting in incorrect estimates.
4. The hypothesis is correct, but MCMC did not work well and random samples were distant from high-density areas of the posterior distribution (mathematical propositions ensure that if infinitely many samples are produced, their distribution converges to the posterior distribution).
5. Other possibilities.

We should identify which processes in Figure 3a worked well and those that did not (cf. Duhem, 1962). Unfortunately, this is often unclear and we have to rely on induction (e.g., suggestions from previous studies), abduction (deny very unlikely possibilities), speculation, and so on. By using a statistical model for estimation, we can utilize deduction that complements an inductive inference, and at the same time, demands new inductive processes.

The roles of mathematics in a mathematical model and mathematics in a statistical model are quite different. A hypothesis tends to be first shown qualitatively, for example, “the large wintering site contributes to the maintenance of the bird population,” which is not easy to verify. If it is mathematically written as in the state model in Figure 2, we can quantitatively test the hypothesis by checking $T_{23} + T_{24} > T_{41} + T_{42}$. In contrast, mathematics in a statistical model connects the mathematical model with data using probability densities. Analyses of a mathematical model may involve a law in nature, such as periodicity, stability, synchronicity, and phase transition, while a statistical model covers human activities. Conventionally, a state-space model is called a statistical model, implying that a statistical model may contain a mathematical model in a hierarchical structure.

The contrasts between Figure 3a and Figure 3b might be invalid in modern sciences because we can rarely measure a predictive value directly. For example, if we repeated measurements and took their sample mean, we had already relied on a statistical estimation because the sample mean is nothing more than a statistical
estimator of the true mean from data. The conventional sample mean formula is commonly used because its suitability was proven in various statistical aspects, but since Stein (1956) derived another estimator of a mean, new estimators of a mean have been repeatedly proposed to fulfill other aspects.

The two inferences in Figure 3 may be unified in Figure 4, but there are large variations around the process of “Data → Statistical estimation”; some can be expressed by just one equation, some need a hierarchical model, and so on.

Finally, classical statistical tests such as the significance test follow completely different inference schemes; they begin with setting a null hypothesis and end with its rejection or not. Conventional philosophical studies about statistics strongly tend to be about this classical discipline.

4. Bayesian inference

There are two different usages of the term “Bayes.” Throughout the paper, Bayesian statistics refers to the statistical model demonstrated above, while a Bayesian inference means computations of the degree of the belief in a hypothesis.
In the Bayesian inference, we begin by setting a prior probability to quantify the degree of belief in a hypothesis. Let $H$ be a hypothesis, and $P(H)$ be its prior probability. When we obtain data $x$, by applying Bayes’ theorem, we update $P(H)$ to the posterior probability $P(H|x)$ as follows:

$$P(H|x) = \frac{P(x|H)P(H)}{P(x)}$$  \hspace{1cm} (5)

where $P(x|H)$ is the conditional probability of obtaining data $x$ under hypothesis $H$, and $P(x)$ is expressed as:

$$P(x) = \sum_h P(x|h)P(h)$$  \hspace{1cm} (6)

wherein $h$ should move all possible hypotheses that may generate data $x$. If $P(H|x)$ approaches one, the degree of belief in hypothesis $H$ is certain; thus, $H$ is likely to be true. Bayesian inference is expected to serve as an alternative to or complement inductive inferences, and has been intensively examined in philosophical studies (e.g., Savage 1972, Sprenger and Hartmann 2019, and references therein).

The Bayesian inference does not appear in the previous sections; neither a prior nor a posterior degree of belief in the hypothesis appeared. Instead, a hierarchical Bayesian model was applied in which prior distributions were given to unknown parameters and Bayes’ theorem was applied to induce equation (3) for the posterior distribution of the parameters. As is illustrated in Figure 4, we may omit “Bayes” in an inference scheme.

Bayesian inference has rarely been applied to the modern sciences. In fact, to my knowledge, nobody has ever computed the degree of belief in general relativity, the Mendelian laws, the neutral theory of molecular evolution, and so on. The main reason for this is that the Bayesian inference requires the computation of the sum regarding all possible hypotheses (equation 5), which is called the catch-all problem. It is difficult to cover all hypotheses about gravity, genetics, molecular evolution, etc. Exceptional examples are seen in introductory textbooks about the Bayesian statistics for hypothesis such as “one has cancer,” adding another hypothesis “one does not have cancer” covers all the possibilities. Unlike such examples in which the negative hypothesis is clear, in the case of general relativity, negation should cover unexplored theories.

In my knowledge, in modern statistical sciences, Bayes’ theorem is applied, in many cases, to update a probability distribution of unknown parameters, and not to update the probability of a hypothesis. The main goal is to obtain likely estimates of quantities of interest, not to quantify the degree of belief in a hypothesis. Unfortunately, “Bayes” in many of philosophical studies indicates the Bayesian inference, and few philosophers have examined up-to-date Bayesian statistics from philosophical viewpoints.
5. Different uses of the same terminology

As was demonstrated in the case of Bayes, philosophers and statisticians use the same statistical terminology differently. Another example is the objectivity and subjectivity in Bayesian arguments.

One of the biggest problems in Bayesian statistics is that a posterior distribution is not invariant for a transformation of parameters (Lele 2020a). When we set a prior distribution and run MCMC, parameters are often transformed by exponential, logarithmic, logit, and other elementary functions for mathematical convenience, and the choice of parameterization influences the posterior distribution. The parameterization is often determined subjectively (this was actually done for the above example), and this problem has been examined in statistical literatures as the riskiest subjectivity in Bayesian statistics (Lele 2020a and references therein). On the other hand, philosophers seem to have very few concerns with such a “technical” problem (Sprenger 2018).

The term “likelihood” is also differently used by statistical scientists and philosophers. In statistical studies, the likelihood is mathematically defined for a statistical model as the conditional probability that we obtain data given that model in which data are assumed to be randomly generated from a probability distribution—all this written as a function of the data. In contrast, philosophical studies often use the term “likelihood” as any conditional probability that we obtain data given a hypothesis, without explicitly expressing a data-generating process and specifying a probability distribution. For example, Weirich (2011) discussed the likelihood of Bertrand’s box paradox without specifying a probability distribution.

6. Statistical methods for complementing induction after statistical estimation

To complement the inductions after obtaining statistical estimates (the bottom process near the box of induction” in Figures 3a and 4), statistical methods were proposed in existing studies and have been applied in statistical sciences.

One is the relative evaluation of statistical models using an information criterion. The origin of this approach is the Akaike information criterion (AIC) (Akaike 1973), which uses the distance between the true data-generating process and a statistical model by the Kullback-Leibler divergence. Although the true process is unknown, because the data are assumed to be generated by the true process, ignoring a constant that does not depend on statistical models, AIC approximates the distance and is applicable to comparing statistical models. In the Bayesian context, it is known that the marginal likelihood, \( \int_D L(w|x)f(w)dw \), is proportional to the posterior degree of belief in the model under the Bayesian inference; thus, is applicable to the relative
evaluation of Bayesian statistical models. Equivalently, the ratio of marginal likelihoods is called a Bayes factor. Because the integration requires a large number of computations, approximations were derived. Examples are: the Bayesian information criterion (BIC) in Schwartz (1978), and Watanabe-BIC (WBIC) in Watanabe (2013).

These approaches switched an inference from an absolute test for one hypothesis to a relative evaluation for pairwise hypotheses. For the relative evaluation of models, the likelihood ratio is one of the oldest statistical concepts and has been intensively examined in philosophical studies (Royall 1997, Bandyopadhyay et al. 2016). The discussion has been extended to AIC (Sober 2008, Rochefort-Maranda 2016), BIC, or both (Dowe et al. 2007, Sprenger 2013, Steele and Wendl 2016). The relative evaluation between pairwise hypotheses is being developed to a new scientific inferential scheme called “evidential paradigm” (Lele 2020b and references therein).

Extending the basic ideas in AIC, Ponciano and Taper (2019) derived the equation that approximates the absolute distance between the true data-generating process and a statistical model. If Bayes’ theorem is used only to obtain likely estimates, Lele et al. (2007) developed another mathematical framework (data cloning) to avoid dependence of a prior distribution on numerical results.

Recently published papers by statistical scientists are waiting for philosophers’ investigations.

### 7. Two proposals toward interdisciplinary developments for philosophy of statistical sciences

Ecology, a discipline of modern sciences, has been often misunderstood; an ecologist often states an inductive inference based on personal natural observations as if it followed a standard scientific scheme. Since the 1980s, most ecological studies have been based on systematically collected data. On the other hand, since many ecological data have temporal and/or spatial information, and because there used to be few statistical techniques that can fully utilize long-term ecological data, ecologists often summarized spatiotemporal data into a single value and applied classical statistics. After information criteria became popular in the 1990s, ecologists were able to analyze spatiotemporal ecological data, and developments in Bayesian statistics in the 2000s promoted the use of statistical models.

A little earlier than statistics, mathematical ecology exhibited active development in the 1980s. Presently, many qualitatively described ecological hypotheses are expressible by mathematical languages and can be quantitatively tested; ecology can follow the scheme in Figure 4, and how to combine empirical and mathematical ecology has been repeatedly discussed in existing literature (e.g., Orzack and Sober 1993, Lawton 1999).

However, the complexity of statistical models prevents clear inferences (Fig. 3a),
and some ecologists have been looking for adequate scientific methodology in philosophical studies or seek opportunities to discuss this issue with philosophers (Taper and Lele 2004). Fortunately, there are many statistical keywords in common between statistical sciences and philosophy, such as Bayes and likelihood. However, many statistical scientists find philosophical studies to be difficult to understand, and the biggest reason is probably because statistical terms are used differently and have different meanings in philosophical studies.

In conclusion, I propose the following two proposals to statistical scientists who are concerned with the philosophy and philosophers of statistical science.

(1) Terminology should be unified. If a philosopher thinks of statistical scientists for potential readers, the usage of statistical terms should follow a common usage in statistics.

(2) Up-to-date philosophical studies should prioritize up-to-date statistical sciences rather than classical statistics. Up-to-date statistical scientists should not stop with studies of philosophy in the Popper era, but should communicate with modern philosophers of science. Such attempts are already ongoing (e.g., Taper and Lele 2004, Gelman and Hennig 2017 and discussion papers, Gelman et al. 2019, and papers in this special issue).

Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers 16H03050 and 20H01736. I would like thank to Dr. Matsuo, Dr. Morimoto, Dr. Ohkubo, Dr. Mark L. Taper and an anonymous referee for insightful and thoughtful comments and suggestions. I also would like to thank Editage (www.editage.com) for English language editing. The study began with an international symposium entitled with “What is a good model?” in 2016 at the Institute of Statistical Mathematics in Japan, which was supported by JSPS KAKENHI Grant Numbers JP25340110.

References

Akaike, Hirotsugu (1973). “Information Theory and an Extension of the Maximum Likelihood Principle.” In: 2nd International Symposium on Information Theory (eds. Petrov BN, Csaki F), pp.267–281. Akademiai, Kiado, Budapest, Hungary.

Antoine, Guisan, Wilfried Thuiller, Niklaus E. Zimmermann (2017). Models: With Applications in R: Habitat Suitability and Distribution. Cambridge University Press.

Bandyopadhyay, Prasanta S., Gordon Brittan Jr. and Mark L. Taper (2016). Belief, Evidence, and Uncertainty: Problems of Epistemic Inference. Springer.

Bernardo, José M. (2012). “Integrated Objective Bayesian Estimation and Hypotheses Testing.” Bayesian Statistics 9: 1–68.

Brittan Jr., Gordon and Prasanta S. Bandyopadhyay (2019). Ecology, Evidence, and Objectivity: In Search of a Bias-Free Methodology. Frontiers in Ecology and Evolution.
Caswell, Hal (2001). *Matrix Population Models: Construction, Analysis, and Interpretation.* Second edition, Sinauer Associates.

Duhem, Pierre (1962). *The Aim and Structure of Physical Theory.* ed P. Weiner. New York, NY: Atheneum.

Dowe, David L., Steve Gardner and Graham Oppy (2007). “Bayes not Bust! Why Simplicity is no Problem for Bayesians.” *British Journal for the Philosophy of Science* 58: 709–754.

Fay, Rémi, Stephanie Michler, Jacques Laasser and Michael Schaub (2019). “Integrated Population Model Reveals that Kestrels Breeding in NestBoxes Operate as a Source Population.” *Ecography* 42: 2122–2131.

Furrer Roman D. and Gilberto Pasinelli (2016). “Empirical Evidence for Source-sink Populations: A Review on Occurrence, Assessments and Implications.” *Biological Reviews* 91: 782–795.

Gelman, Andrew and Cosma Rohilla Shalizi (2013). “Philosophy and the Practice of Bayesian Statistics.” *British Journal of Mathematical and Statistical Psychology*, 66: 8–38.

Gelman, Andrew and Christian Hennig (2017). “Beyond Subjective and Objective in Statistics.” *Journal of Royal Statistical Society A*: 967–1033.

Gelman, Andrew, Brian Haig, Christian Hennig, Art Owen, Robert Cousins, Stan Young, Christian Robert, Corey Yanofsky, E. J. Wagenmakers, Ron Kenett, Daniel Lakeland (2019). “Many perspectives on Deborah Mayo’s “Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars.”” arXiv:1905.08876 [stat.OT].

Henderson, Leah, Noah D. Goodman, Joshua B. Tenenbaum, James F. Woodward (2010). “The Structure and Dynamics of Scientific Theories: A Hierarchical Bayesian Perspective”. *Philosophy of Science* 77: 172–200.

Kéris, Marc and Michael Schaub (2012). *Bayesian Population Analysis using WinBUGS.* Elsevier Inc.

Konishi, Sadanori and Genshiro Kitagawa (2008). *Information criteria and Statistical Modeling.* Springer.

Lawton, John H. (1999). “Are There General Laws in Ecology?” *OIKOS* 84: 177–192.

Lele, Subhash R., Brian Dennis and Frithjof Lutscher (2007). Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. *Ecology Letters*, 10, 551–563. doi.org/10.1111/j.1461-0248.2007.01047.x.

Lele, Subhash R. (2020a). “Consequences of Lack of Parameterization Invariance of Non-informative Bayesian Analysis for Wildlife Management: Survival of San Joaquin Kit Fox and Declines in Amphibian Populations.” *Frontiers in Ecology and Evolution*. 21, doi.org/10.3389/fevo.2019.00501

Lele, Subhash R. (2020b). “How should we quantify uncertainty in Statistical Inferences.” *Frontiers in Ecology and Evolution*: 13, doi.org/10.3389/fevo.2020.00035

Link, William A., Richard J. Barker (2010). *Bayesian Inference: with ecological applications.* Elsevier Ltd.

Niiniluoto, Ilkka (2012). “The Foundation of Statistics: Inference vs. Decision.” In: D.
Dieks et al. (eds.), Probabilities, Laws, and Structures, The Philosophy of Science in a European Perspective 3, 29–41. Springer.

Orzack, Steven Hecht and Elliot Sober (1993). “A Critical Assessment of Lewins’s the Strategy of Model Building in Population Biology (1966).” The Quarterly Review of Biology 68: 533–546.

Ponciano, José M. and Mark L. Taper (2019). Model Projections in Model Space: A Geometric Interpretation of the AIC Allows Estimating the Distance Between Truth and Approximating Models. Frontiers in Ecology and Evolution: 8. doi.org/10.3389/fevo.2019.00413

Rochefort-Maranda, Guillaume (2016). “Simplicity and Model Selection.” European Journal of Philosophy of Science 6: 261–279.

Royall, Richard (1997). Statistical Evidence: A Likelihood Paradigm. Chapman & Hall/CRC.

Savage, Leonard J. (1972). The Foundations of Statistics. Second edition. New York, Wiley.

Schwartz, Gideon (1978). “Estimating the Dimension of a Model.” Annals of Statistics 6: 461–464.

Sober, Elliot (2008). Evidence and Evolution. Cambridge University Press.

Sprenger, Jan (2013). “The Role of Bayesian philosophy within Bayesian model selection.” European Journal for Philosophy of Science 3: 101–114.

Sprenger, Jan (2018). “The Objectivity of Subjective Bayesianism”. European Journal for Philosophy of Science 8: 539–558.

Sprenger, Jan, Stephan Hartmann (2019). Bayesian Philosophy of Science. Oxford University Press.

Steele, Katie and Charlotte Werndl (2016). “The Diversity of Model Tuning Practices in Climate Science.” Philosophy of Science 83: 1133–1144.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proc. 3rd Berkeley Symp. Math. Statist. Prob., 1: 197–206, Univ. of California Press, Berkeley.

Taper, Mark L. and Subhash R. Lele (eds.) (2004). The Nature of Scientific Evidence. University of Chicago Press.

Taper, Mark L. and José M. Ponciano (2016). Evidential statistics as a statistical modern synthesis to support 21st century science. Population Ecology: 58: 9–29.

Watanabe, S. (2013). “A Widely Applicable Bayesian Information Criterion,” Journal of Machine Learning Research 14: 867–897.

Weegman, Mitch D., Stuart Bearhop, Anthony D. Fox, Geoff M. Hilton, Alyn J. Walsh, Jennifer L. McDonald, and David J. Hodgson (2016). “Integrated Population Modeling Reveals a Perceived Source to Be a Cryptic Sink,” Journal of Animal Ecology 85: 467–475.

Weirich, Paul (2011). “The Bayeian Decision-Theoretic Approach to Statistics.” In Handbook of the Philosophy of Science, Volume 7: Philosophy of Statistics. Elsevier.

Yamamura, Koji (2014). Estimation of the predictive ability of ecological models. Communications in Statistics - Simulation and Computation. DOI:10.1080/03610918.2014.889161

(Received 2021.1.13; Revised 2021.7.6; Second Revised 2021.7.29; Accepted 2021.7.31)