Journal of Statistical Planning and Inference 148 (2014) 111–122

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Predictive measures of the conflict between frequentist and Bayesian estimators

Pierpaolo Brutti, Fulvio De Santis, Stefania Gubbiotti *

Dipartimento di Scienze Statistiche, Sapienza Università di Roma, Italy

ARTICLE INFO

Article history:
Received 21 December 2012
Received in revised form
17 December 2013
Accepted 18 December 2013
Available online 27 December 2013

Keywords:
Bayesian estimation
Clinical trials
Comparison of estimators
Exponential families
Frequentist estimation
Predictive analysis
Sample size

ABSTRACT

In the presence of prior information on an unknown parameter of a statistical model, Bayesian and frequentist estimates based on the same observed data do not coincide. However, in many standard parametric problems, this difference tends to decrease for growing sample size. In this paper we consider as a measure of discrepancy (D_n) the squared difference between Bayesian and frequentist point estimators of the parameter of a model. We derive the predictive distribution of D_n for finite sample sizes in the case of a one-dimensional exponential family and we study its behavior for increasing sample size. Numerical examples are illustrated for normal models.

1. Introduction

Bayesian statistics offers the theoretical framework for combining experimental and extra-experimental information on phenomena under study. As a consequence, Bayesian procedures for inference on an unknown parameter of a statistical model take into account both experimental data and information on the parameter incorporated in the so-called prior distribution.

In the presence of pre-experimental information, frequentist and Bayesian procedures, such as point or interval estimates based on the same observed sample, in general, do not coincide. However, in many standard parametric problems, the discrepancy between frequentist and Bayesian procedures is rather limited when sampling information dominates the prior distribution. Furthermore, this conflict tends to disappear as the sample size increases and, for sufficiently large sample sizes, frequentist procedures may provide good approximations of Bayesian methods.

A paradigmatic example is the estimation problem for the expected value of a normal random variable. In this case (see Section 4.1 for technical details), given n observations from independent and identically distributed (i.i.d.) normal random variables, the standard Bayesian estimate of θ is a linear combination of the sampling mean, $\overline{x}_n$, and of a prior guess on the parameter, μ_A:

$$(1-a_n)\overline{x}_n + a_n\mu_A, \quad a_n \in (0,1), \tag{1}$$

where a_n tends to zero as n diverges. Therefore, for a sufficiently large sample size, $(1-a_n)\overline{x}_n + a_n\mu_A \simeq \overline{x}_n$, i.e. the Bayesian estimate (1) is well approximated by the sample mean.

* Corresponding author at: Dipartimento di Scienze Statistiche, Sapienza Università di Roma, Piazzale Aldo Moro n.5, 00185 Roma.
E-mail address: stefania.gubbiotti@uniroma1.it (S. Gubbiotti).

0378-3758/$ - see front matter
http://dx.doi.org/10.1016/j.jspi.2013.12.009

In most of introductory books on Bayesian inference (see, for instance, Berger, 1985; Bernardo and Smith, 1994; Gelman et al., 2004; Lee, 2004; O'Hagan and Forster, 2004; Robert, 2001), the progressive reduction of conflict between Bayesian and frequentist procedures is typically showed only as a limiting result. In this paper we propose a predictive finite-sample look at the problem, in the case of point estimation. First, we introduce a measure of conflict between a Bayesian and a frequantist point estimator. Then, we study its predictive distribution. This pre-posterior comparative analysis allows one to:

(i) know in advance (before observing the data) what discrepancy is expected for a sample size, n^*, chosen according to a standard criterion (for instance, a criterion based on the power function of a test);
(ii) evaluate the expected impact of the prior distribution on the predictive discrepancy measure, i.e. quantifying the degree of informativeness of a prior distribution for given sample sizes.
(iii) evaluate the expected progressive reduction of discrepancy between two competing point estimates as n increases;
(iv) establish the minimum sample size that guarantees, with a chosen probability, that the discrepancy between estimates is below a given chosen threshold, i.e. how large n must be so that a frequentist estimate provides an approximation of the Bayesian estimate.

The topic of the paper is related to the more general problem of the relationships between frequentist and Bayesian point estimation, whose literature is essentially endless. See, for instance, Bayarri and Berger (2004), Berger (1985), Samaniego and Reneau (1994), Samaniego (2012), Bickel (2012) and references therein.

The paper is organized as follows. In Section 2.1, we introduce D_n, the measure of conflict between a frequentist and a Bayesian estimator. In Section 2.2 we specify the predictive distribution for D_n to be used. In Section 3 we derive the general expressions of D_n, its predictive cumulative distribution function (cdf) and its expected value in the case of one-parameter exponential family with conjugate priors. These results are specialized to the case of the normal model, assuming both known (Section 4.1) and unknown (Section 4.2) variance. In Section 4.3 we consider an illustrative example based on a superiority clinical trial. Finally, Section 5 sketches some possible extensions of the presented methodology and Section 6 contains a discussion.

2. Methodology

2.1. A measure of discrepancy between estimators

Let $\mathbf{X}_n=(X_1,X_2,\dots,X_n)$ be a random sample from a probability distribution $f_n(\cdot|\theta)$, where θ is an unknown real-valued parameter that belongs to the parameter space, Θ. Following the Bayesian inferential approach, we assume that θ is a random variable, with distribution denoted by π_A. For simplicity, assume that $\Theta\subseteq\mathbb{R}$ and that π_A is a density function. We will refer to π_A as to the *analysis-prior*. It models pre-experimental knowledge/uncertainty on θ based, for instance, on subjective opinion of experts or historical data. Given an observed sample $\mathbf{x}_n=(x_1,x_2,\dots,x_n)$ and the likelihood function of θ, $f_n(\mathbf{x}_n|\theta)$, the posterior distribution of the parameter is

$$\pi(\theta|\mathbf{x}_n)=\frac{f_n(\mathbf{x}_n|\theta)\pi_A(\theta)}{\int_\Theta f_n(\mathbf{x}_n|\theta)\pi_A(\theta)\,d\theta}.$$

We denote a Bayesian estimator of θ as $\hat{\xi}_B(\mathbf{X}_n)$ whereas $\hat{\xi}_F(\mathbf{X}_n)$ is a generic consistent frequentist estimator. In this paper we consider the posterior expectation of the parameter θ, $\mathrm{E}(\theta|\mathbf{X}_n)=\int_\Theta\theta\pi(\theta|\mathbf{x}_n)\,d\theta$, as $\hat{\xi}_B$ and the maximum likelihood estimator (MLE) as $\hat{\xi}_F$. Nevertheless, all the following can be extended to other Bayesian and frequentist point estimators. As a measure of discrepancy between $\hat{\xi}_B$ and $\hat{\xi}_F$ we consider the standard squared difference:

$$D_n(\mathbf{X}_n)=[\hat{\xi}_B(\mathbf{X}_n)-\hat{\xi}_F(\mathbf{X}_n)]^2.$$

Before observing the data, $\hat{\xi}_B$, $\hat{\xi}_F$ and D_n are random variables (functions of $\mathbf{X}_n$). We assume that, as n tends to infinity, D_n converges in probability to zero. Therefore, as the sample size increases, Bayesian and frequentist point estimators tend to become closer and closer. In the next section we specify the predictive distributions of D_n that we will use in the following.

2.2. Predictive analysis

Studying the predictive distribution of D_n is a typical pre-posterior problem. There are two main approaches to Bayesian pre-posterior analysis: the *conditional* and the *predictive* approaches. See, for instance, Chaloner and Verdinelli (1995), in the context of general Bayesian experimental design, and De Santis (2006), in the context of sample size determination. The *conditional* approach prescribes the use of the sampling distribution $f_n(\cdot|\theta)$, with $\theta=\mu_D$, a "guess-value" for the unknown parameter. This method takes into account a single value for θ at the pre-experimental phase of the analysis and leads to predictive evaluations that depend on this chosen value μ_D. The *predictive* approach implies the use of the prior predictive distribution

$$m_D(\mathbf{x}_n)=\int_\Theta f_n(\mathbf{x}_n|\theta)\pi_D(\theta)\,d\theta,$$