On a predictive measure of discrepancy between classical and Bayesian estimators

Pierpaolo Brutti, Fulvio De Santis and Stefania Gubbiotti

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 it is well known that, in many standard parametric problems, their discrepancy tends to be reduced as the sample size increases. In this paper we consider a measure of discrepancy, D_n, between a frequentist and a Bayesian point estimator and we study its predictive distribution. In some specific examples we analyze the main characteristics of this predictive distribution for increasing sample sizes. We also consider the use of the predictive density of D_n for the assessment of a prior distribution informativeness. Some explicit results are given for the normal model.

Key words: Clinical trials; Conflict between estimators; Predictive approach.

1 Introduction

Bayesian methods offer the theoretical framework for combining experimental data and pre-experimental information on an unknown parameter, that is formalized by a prior probability distribution. In the presence of prior information, frequentist and Bayesian procedures, such as point or interval estimates based on the same observed sample, do not coincide. However, in many standard parametric problems, the discrepancy between frequentist and Bayesian procedures tends to disappear as the sample size increases. This is typically shown in most of introductory books on Bayesian inference (see, among others, [1]).

Let us consider the estimation problem for the expected value θ of a normal distribution. Given n observations from i.i.d. normal random variables, the standard

Pierpaolo Brutti e-mail: pierpaolo.brutti@uniroma1.it · Fulvio De Santis e-mail: fulvio.desantis@uniroma1.it · Stefania Gubbiotti e-mail: stefania.gubbiotti@uniroma1.it
Sapienza Università di Roma, Dipartimento di Scienze Statistiche

Bayesian estimate of θ is a linear combination of the sampling mean, $\bar{x}_n$, and of a prior guess on θ, μ_A, i.e. $\omega_n \bar{x}_n + (1-\omega_n)\mu_A$, where ω_n tends to one as n diverges. Therefore, for a sufficiently large sample size, the sample mean provides a good approximation of the Bayesian estimate.

In this paper, we are interested in analysing a measure of discrepancy between two competing estimators. This measure is random before observing the data. In Section 2, we introduce a specific measure of discrepancy, D_n, between a frequentist and a Bayesian estimator and in Section 2.1 we derive its explicit expression, its predictive cumulative distribution function (cdf) and its expected value for the normal model with conjugate priors. After briefly discussing the asymptotic behaviour of D_n, in Section 2.2 we focus on a fixed sample size, optimal with respect to a given criterion, and we assess the informativeness of the prior by evaluating D_n.

2 A discrepancy between estimators

Let $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ 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, $\Theta \subseteq \mathbb{R}$. Let $\mathbf{x}_n = (x_1, x_2, \ldots, x_n)$ be an observed sample, $\pi_A(\cdot)$ the prior density function of θ, $f_n(\mathbf{x}_n|\theta)$ the likelihood function and $\pi_A(\theta|\mathbf{x}_n)$ the posterior distribution. We will refer to π_A as to the *analysis-prior*. It models pre-experimental knowledge/uncertainty on θ taken into account in posterior analysis.

We denote a Bayesian estimator of θ as $\hat{\theta}_B(\mathbf{X}_n)$, whereas $\hat{\theta}_F(\mathbf{X}_n)$ is a generic classical estimator. Let $D_n(\mathbf{X}_n)$ be a measure of discrepancy between $\hat{\theta}_B$ and $\hat{\theta}_F$. Specifically, we consider the standard squared difference between estimators, i.e.

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

Before observing the data, $\hat{\theta}_B$, $\hat{\theta}_F$ and D_n are random variables (functions of $\mathbf{X}_n$). For instance, in the following we consider the posterior expectation of the parameter θ, $\mathrm{E}(\theta|\mathbf{X}_n) = \int_\Theta \theta \pi(\theta|\mathbf{x}_n) \mathrm{d}\theta$, as $\hat{\theta}_B$, and the maximum likelihood estimator (MLE) as $\hat{\theta}_F$, although the methodology could be extended to other estimators. We want to evaluate the probability of observing a small/large discrepancy between $\hat{\theta}_B$ and $\hat{\theta}_F$. As for any other pre-posterior Bayesian analysis, to this purpose two alternative distributions for the data can be used. The *conditional* approach prescribes the use of the sampling distribution $f_n(\cdot|\theta)$, with $\theta = \mu_D$, a "design value" for the unknown parameter, whereas the *predictive* approach implies the use of the predictive distribution $m_D(\mathbf{x}_n) = \int_\Theta f_n(\mathbf{x}_n|\theta)\pi_D(\theta)\mathrm{d}\theta$, where π_D (*design-prior*) is a density function that accounts for uncertainty on the design value of θ. Note that the conditional approach is a special case of the predictive one, when π_D is a point-mass prior on μ_D; for this reason in the following we adopt the most general approach. See [4] and [2] for a detailed discussion on this point.

2.1 Results for the normal model

Let $\mathbf{X}_n$ be a random sample from a $N(\theta,\sigma^2)$ distribution. The MLE of θ is $\hat{\theta}_F = \bar{x}_n$. Assume for θ a normal prior density $\pi_A(\theta) = N(\theta|\mu_A, \sigma^2/n_A)$, where n_A is given the standard interpretation of *prior sample size*. From standard results on conjugate analysis [1], the posterior distribution of θ is a normal density of mean $(n+n_A)^{-1}(n\bar{x}_n + n_A\mu_A)$ and variance $(n+n_A)^{-1}\sigma^2$. Hence $\hat{\theta}_B = \omega_n\bar{x}_n + (1-\omega_n)\mu_A$, with $\omega_n = n(n+n_A)^{-1}$.

Letting $\pi_D(\theta) = N(\theta|\mu_D, \sigma^2/n_D)$, the predictive density function of $\bar{x}_n$ is $m_D(\bar{x}_n) = N(\bar{x}_n|\mu_D, \psi_n^2)$, where $\psi_n^2 = b_n\sigma^2$ and $b_n = (n+n_D)(nn_D)^{-1}$. Given the above assumptions, letting $a_n = 1-\omega_n$ the explicit expression of D_n is as follows

$$D_n = a_n^2(\bar{X}_n - \mu_A)^2.$$

It is straightforward to check that the predictive expected value of D_n is

$$e_n = a_n^2[b_n\sigma^2 + \delta^2],$$

where $\delta = \mu_A - \mu_D$, while the cdf of D_n is

$$p_n(d) = \Phi\left[b_n^{-1/2}(\delta + a_n^{-1}d^{1/2})\,\sigma^{-1}\right] - \Phi\left[b_n^{-1/2}(\delta - a_n^{-1}d^{1/2})\,\sigma^{-1}\right],$$

where $\Phi(\cdot)$ is the standard Normal cdf.

Noting that $b_n = O(1)$ and $a_n = o(n^{-1})$, it follows that $e_n = o(n^{-2})$ and that, as n diverges, D_n converges in probability to zero as fast as n^{-2}. Both e_n and p_n depend on the prior means only through the absolute difference $|\delta|$.

2.2 Quantifying the informativeness of the prior

Let us consider the set up of an efficacy clinical trial, with positive values of θ indicating an effective treatment. Let us suppose for instance to select the minimum sample size n^* such that the frequentist conditional power reaches a desired level (see among others [3]). For a normal model, under the assumptions of the previous section the power is $\beta = \Phi\left(\frac{\theta\sqrt{n}}{\sigma} + z_{\alpha/2}\right)$, where z_α denotes the quantile of a standard normal at level α. If we set for instance a design value for θ equal to 0.5, when $\sigma = 2$ and $\alpha = 0.05$, the optimal sample size required to reach a 0.80 power is $n^* = 126$. Let us assume a design prior of parameters $\mu_D = 0.5, n_D = 20$. Based on n^*, we can compute e_{n^*} for a given analysis prior. The predictive expected discrepancy thus provides a measure of the conflict between the two alternative estimators and, at the same time, it represents the level of informativeness of the prior. In fact, the larger e_{n^*}, the stronger the impact of the prior in $\hat{\theta}_B$. In this way it is also possible to compare different choices for the analysis prior in terms of their informativeness level. In Table 1a we report the values of e_{n^*} for several choices of the analysis prior parameters. For illustrative purposes we may consider a threshold value such as $d = 0.2$ (values exceeding 0.2 are bolded in the table): when a e_{n^*} is below this threshold the corresponding analysis prior can be considered relatively non informative, otherwise its impact on the Bayes estimator is remarkable. As expected, for increasing values of n_A, the analysis prior becomes more informative and, consequently, the expected discrepancy is larger. The increments of the conflict measure appear to be smaller when the analysis prior mean coincides with θ_D. Similar considerations can be drawn by computing $p_{n^*}(d)$, but in this case a threshold on a probability scale can be set: for instance in Table 1b the values of $p_{n^*}(d)$ below a given level, say 0.5, identify the analysis prior parameters with stronger impact on the Bayesian estimator.

Table 1 e_{n^*} and p_{n^*} for several choices of the analysis prior, with $n^* = 126$.

	n_A					
μ_A	1	10	20	50	100	200
(a)						
-2	0	0.035	0.122	**0.523**	**1.269**	**2.440**
-1	0	0.013	0.047	0.200	**0.486**	**0.934**
0	0	0.003	0.009	0.039	0.094	0.181
0.5	0	0.001	0.004	0.019	0.045	0.087
(b)						
-2	1	1	0.944	**0.027**	**0.001**	**0.000**
-1	1	1	1	0.561	**0.155**	**0.055**
0	1	1	1	0.987	0.855	0.677
0.5	1	1	1	0.999	0.964	0.870

• $\mu_D = 0.5, n_D = 20$

In summary, in this work we have introduced a measure of conflict between classical and Bayesian point estimators. A similar methodology could be considered extending the idea of discrepancy both to different objects to be compared (such as interval estimators, Bayes factors, etc.) and to more complex models.

References

1. Bernardo, J.M. and Smith, A.F.M. (1994). Bayesian Theory. Wiley.
2. De Santis, F. (2006). Sample size determination for robust Bayesian analysis. Journal of the American Statistical Association, 101, n. 473, 278-291.
3. Spiegelhalter, D.J, Abrams, K.R. and Myles, J.P. (2004). Bayesian approaches to clinical trials and health-care evaluation. Wiley.
4. Wang, F., and Gelfand, A.E. (2002). A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. Statistical Science, 17, n. 2, 193-208.