A predictive measure of the additional loss of a non-optimal action under multiple priors

Una misura predittiva della perdita dovuta all'uso di un'azione non ottima in presenza di diverse a priori

Fulvio De Santis and Stefania Gubbiotti

Abstract In Bayesian decision theory, the performance of an action is measured by its posterior expected loss. In some cases it may be convenient/necessary to use a non-optimal decision instead of the optimal one. In these cases it is important to quantify the additional loss we incur and evaluate whether to use the non-optimal decision or not. In this article we study the predictive probability distribution of a relative measure of the additional loss and its use to define sample size determination criteria in one-sided testing.

Abstract L'analisi delle decisioni bayesiane prevede che la qualità di un'azione si misuri in termini della sua perdita attesa a posteriori. In alcuni casi può essere conveniente/necessario adottare una decisione non ottima al posto di quella ottima. Per valutare l'opportunità di questa scelta, è importante quantificare la perdita aggiuntiva che essa comporta. Oggetto di questo lavoro è lo studio della distribuzione predittiva di una misura relativa di tale perdita addizionale e il suo impiego per la scelta della numerosità campionaria nei problemi di test di ipotesi unilaterali.

Key words: Bayesian inference, Experimental design, One-sided testing, Predictive analysis, Sample size determination, Statistical decision theory.

1 Introduction

In a decision problem involving the unknown parameter of a statistical model, consider two decision makers who have different prior information and/or opinions on

Fulvio De Santis
Dipartimento di Scienze Statistiche, Sapienza Università di Roma,
e-mail: fulvio.desantis@uniroma1.it

Stefania Gubbiotti
Dipartimento di Scienze Statistiche, Sapienza Università di Roma,
e-mail: stefania.gubbiotti@uniroma1.it

the parameter. Let π_e and π_o denote their priors and let a_e and a_o be the actions that minimize the two posterior expected losses. Furthermore, let us suppose that the first decision makers is forced to take the action a_o, although it is not optimal from her/his point of view: under π_e, the posterior expected loss of a_o is in fact larger than the posterior expected loss of a_e. Finally, assume that the sample size of the experiment is selected by a third actor using the predictive distribution of the data based on the prior π_d, that, in general, is different from both π_o and π_e. The goal of the experiment planner is to determine the minimal sample size such that a relative predictive measure of the additional loss due to the use of a_o rather than a_e is sufficiently small.

Statistical decision problems under several actors have been previously considered, for instance, in [5], [6] and [4]. In this paper we extend to the testing problem the results of [3] (related to point-estimation) and we focus in particular on the one-sided testing set-up.

The outline of the article is as follows. In Section 2 we formalize the proposed methodology for a generic statistical decision problem: we introduce a relative measure of additional loss due to a non optimal action and the related predictive criterion for the selection of the sample size. In Section 3 the methodology is developed for a one-sided testing problem for a real-valued parameter. Results are then specialized to one-sided testing of a normal mean (Section 4) and some numerical examples are provided in Section 4.1. Finally, Section 5 contains some concluding remarks.

2 Methodology

Let $X_1, X_2, \ldots, X_n$ be a random sample from $f_n(\cdot|\theta)$, where θ is an unknown parameter, $\theta \in \Theta$. Let $a \in \mathscr{A}$ denote a generic action for a decision problem regarding θ and $L(a,\theta)$ the loss of a when the true parameter value is θ. We assume that two competing priors, π_o and π_e, are available for θ. Given an observed sample $\boldsymbol{x_n} = (x_1, x_2, \ldots, x_n)$, let $\pi_j(\theta|\boldsymbol{x_n})$ be the posterior distribution of θ from prior π_j, and

$$\rho_j(\boldsymbol{x_n}, a) = \mathbb{E}_{\pi_j}\left[L(a,\theta)|\boldsymbol{x_n}\right] = \int_{\Theta} L(a,\theta)\pi_j(\theta|\boldsymbol{x_n})d\theta$$

be the posterior expected loss of an action a, for $j = o, e$. Let a_j denote the optimal action with respect to $\pi_j(\theta|\boldsymbol{x_n})$. The performance of the action a_o when the expected loss is evaluated with respect to $\pi_e(\theta|\boldsymbol{x_n})$ is then $\rho_e(\boldsymbol{x_n}, a_o) = \mathbb{E}_{\pi_e}\left[L(a_o,\theta)|\boldsymbol{x_n}\right]$. If a_o is used instead of a_e, the *relative additional expected loss* is

$$\bar{A}_{o,e}(\boldsymbol{x_n}) = \frac{\rho_e(\boldsymbol{x_n}, a_o) - \rho_e(\boldsymbol{x_n}, a_e)}{\rho_e(\boldsymbol{x_n}, a_o)}.$$

When $\bar{A}_{o,e}$ is small the non-optimal action a_o performs well even under the prior assumptions represented by π_e. Before observing the data, $\bar{A}_{o,e}(\boldsymbol{X_n})$ is a sequence of r.v. that converges in probability to zero, as n increases. In order to define a sample size criterion we focus on $e_n = \mathbb{E}_{m_d}\left[\bar{A}_{o,e}(\boldsymbol{X_n})\right]$, where $\mathbb{E}_{m_d}[\cdot]$ denotes the expected value with respect to the sample data distribution, $m_d(\boldsymbol{x_n}) = \int_\Theta f(\boldsymbol{x_n}|\theta)\pi_d(\theta)d\theta$, where π_d is the design prior. Hence, for a desired threshold γ, $n^\star = \min\{n \in \mathbb{N} : e_n \leq \gamma\}$ is the optimal sample size that depends on three priors (π_d, π_e, π_o).

3 One-sided Testing

Consider the set up of one-sided testing, i.e. $H_1 : \theta \leq \theta_t$ vs. $H_2 : \theta > \theta_t$, with $\theta_t \in \mathbb{R}$. Let $\mathscr{A} = \{a^{(1)}, a^{(2)}\}$ be the two terminal decisions, where $a^{(i)}$ denotes the choice of H_i, $i = 1, 2$, and

$$L(a^{(1)}, \theta) = b_2 \times 1_{\{\theta:\theta>\theta_t\}}(\theta) \quad \text{and} \quad L(a^{(2)}, \theta) = b_1 \times 1_{\{\theta:\theta\leq\theta_t\}}(\theta)$$

their loss functions ($b_i > 0,\ i = 1, 2$), with $1_A(\cdot)$ the indicator function of the set A. Then the posterior expected losses of $a^{(1)}$ and $a^{(2)}$ are

$$\rho_j(\boldsymbol{x_n}, a^{(1)}) = b_2\,(1 - \mathrm{F}_j(\theta_t|\boldsymbol{x_n})) \qquad \text{and} \qquad \rho_j(\boldsymbol{x_n}, a^{(2)}) = b_1 \mathrm{F}_j(\theta_t|\boldsymbol{x_n}),$$

where $F_j(\cdot|\boldsymbol{x_n})$ is the c.d.f. associated to $\pi_j(\theta|\boldsymbol{x_n})$, $j = o, e$. In this case it is easy to check that the optimal decision function $a_j(\boldsymbol{x_n})$ is

$$a_j(\boldsymbol{x_n}) = \arg\min_{a\in\mathscr{A}} \rho_j(\boldsymbol{x_n}, a) = \begin{cases} a^{(1)} \text{ if } \boldsymbol{x_n} \in \mathscr{Z}_j^{(1)} \\ a^{(2)} \text{ if } \boldsymbol{x_n} \in \mathscr{Z}_j^{(2)} \end{cases} \qquad j = o, e.$$

where

$$\mathscr{Z}_j^{(1)} = \{\boldsymbol{x_n} :\ \rho_j(\boldsymbol{x_n}, a^{(1)}) < \rho_j(\boldsymbol{x_n}, a^{(2)})\} = \{\boldsymbol{x_n} :\ b_2\,(1 - \mathrm{F}_j(\theta_t|\boldsymbol{x_n})) < b_1 \mathrm{F}_j(\theta_t|\boldsymbol{x_n})\}$$

and $\mathscr{Z}_j^{(2)}$ is its complement. The posterior expected loss of the decision function $a_j(\boldsymbol{x_n})$ w.r.t. π_e is

$$\rho_e(\boldsymbol{x_n}, a_j) = \begin{cases} b_2\,(1 - \mathrm{F}_e(\theta_t|\boldsymbol{x_n})) & \text{if } \boldsymbol{x_n} \in \mathscr{Z}_j^{(1)} \\ b_1\,\mathrm{F}_e(\theta_t|\boldsymbol{x_n}) & \text{if } \boldsymbol{x_n} \in \mathscr{Z}_j^{(2)} \end{cases} \qquad j = o, e.$$

Therefore, noting that $\rho_e(\boldsymbol{x_n}, a_e) = \min\{b_1 \mathrm{F}_e(\theta_t|\boldsymbol{x_n}), b_2\,(1 - \mathrm{F}_e(\theta_t|\boldsymbol{x_n}))\}$, we obtain

$$\bar{A}_{o,e}(\boldsymbol{x_n}) = \xi_e(\boldsymbol{x_n}) 1_{\mathscr{Z}_{o,e}}(\boldsymbol{x_n}) \tag{1}$$

where

$$\xi_e(\boldsymbol{x_n}) = 1 - \min\left\{\frac{b_1}{b_2}\frac{\mathrm{F}_e(\theta_t|\boldsymbol{x_n})}{1 - \mathrm{F}_e(\theta_t|\boldsymbol{x_n})}, \frac{b_2}{b_1}\frac{1 - \mathrm{F}_e(\theta_t|\boldsymbol{x_n})}{\mathrm{F}_e(\theta_t|\boldsymbol{x_n})}\right\}. \tag{2}$$

and

$$\mathscr{Z}_{o,e} = \{\boldsymbol{x_n} \in \mathscr{X}^n : a_o(\boldsymbol{x_n}) \neq a_e(\boldsymbol{x_n})\} = \left(\mathscr{Z}_o^{(1)} \cap \mathscr{Z}_e^{(2)}\right) \cup \left(\mathscr{Z}_o^{(2)} \cap \mathscr{Z}_e^{(1)}\right)$$

is the set of $\boldsymbol{x_n}$ leading to conflicting terminal decisions under π_e and π_o respectively. Now, note that $\mathscr{Z}_j^{(1)}$ can be rewritten in terms of the ε-quantile of the posterior distribution of θ, $q_\varepsilon^j(\boldsymbol{x_n})$ with $\varepsilon = \frac{b_2}{b_1+b_2}$, namely

$$\mathscr{Z}_j^{(1)} = \left\{\boldsymbol{x_n} \in \mathscr{Z} : \frac{1-\mathrm{F}_j(\theta_t|\boldsymbol{x_n})}{\mathrm{F}_j(\theta_t|\boldsymbol{x_n})} < \frac{b_1}{b_2}\right\} = \left\{\boldsymbol{x_n} \in \mathscr{Z} : \theta_t > q_\varepsilon^j(\boldsymbol{x_n})\right\}. \qquad (3)$$

Therefore

$$\mathscr{Z}_o^{(1)} \cap \mathscr{Z}_e^{(1)} = \left\{\boldsymbol{x_n} \in \mathscr{Z} : q_\varepsilon^M(\boldsymbol{x_n}) < \theta_t\right\} \text{ and } \mathscr{Z}_o^{(2)} \cap \mathscr{Z}_e^{(2)} = \{\boldsymbol{x_n} \in \mathscr{Z} : q_\varepsilon^m(\boldsymbol{x_n}) > \theta_t\},$$

where $q_\varepsilon^m(\boldsymbol{x_n}) = \min\{q_\varepsilon^e(\boldsymbol{x_n}), q_\varepsilon^o(\boldsymbol{x_n})\}$ and $q_\varepsilon^M(\boldsymbol{x_n}) = \min\{q_\varepsilon^e(\boldsymbol{x_n}), q_\varepsilon^o(\boldsymbol{x_n})\}$. Hence we have

$$\mathscr{Z}_{o,e} = \left\{\boldsymbol{x_n} \in \mathscr{Z} : q_\varepsilon^m(\boldsymbol{x_n}) < \theta_t < q_\varepsilon^M(\boldsymbol{x_n})\right\}. \qquad (4)$$

Finally, from (1)

$$e_n = \int_{\mathscr{Z}} \bar{A}_{o,e}(\boldsymbol{x_n}) m_d(\boldsymbol{x_n}) d\boldsymbol{x_n} = \int_{\mathscr{Z}_{o,e}} \xi_e(\boldsymbol{x_n}) m_d(\boldsymbol{x_n}) d\boldsymbol{x_n}$$

that, in general, must be computed via Monte Carlo approximation. From the above expression we can note that e_n is a monotone function of the Lebesgue measure of $\mathscr{Z}_{o,e}$. An alternative sample size criterion could be based on the predictive probability p_n of the samples yielding conflict. Recalling that, $\forall \boldsymbol{x_n} \in \mathscr{Z}$, $\xi_e(\boldsymbol{x_n}) \leq 1$, it is easy to check that $p_n = \mathbb{P}_{m_d}[\mathscr{Z}_{o,e}] = \mathbb{E}_{m_d}[1_{\mathscr{Z}_{o,e}}(\boldsymbol{X_n})]$ is always smaller than or equal to $e_n = \mathbb{E}_{m_d}[\xi_e(\boldsymbol{X_n}) 1_{\mathscr{Z}_{o,e}}(\boldsymbol{X_n})]$. Therefore, for a given γ, e_n always yields a smaller sample size. The idea is that in e_n the contribution of each sample corresponding to a conflicting decision depends on the strength of the discrepancy in evidence it gives to the two hypotheses, whereas in p_n, it is invariably equal to one.

4 Results for the Normal mean

Let us now further assume that $X_i|\theta \sim \mathrm{N}(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$ and that $\pi_j(\cdot)$ are conjugate priors, i.e. $\theta|\sigma^2 \sim \mathrm{N}(\mu_j, \sigma^2/n_j)$, $j = o, e$. When σ^2 is assumed to be known, the posterior distribution of θ is Normal with mean $\mu_j(\boldsymbol{x_n}) = \frac{n_j\mu_j + n\bar{x}_n}{n_j+n}$ and standard deviation $\sigma_j(\boldsymbol{x_n}) = \frac{\sigma}{\sqrt{n_j+n}}$. In this case $\bar{A}_{o,e}$ can be expressed in terms of Φ, z_ε and $W_j(\boldsymbol{x_n})$, where $\Phi(\cdot)$ is the standard normal c.d.f., z_ε its ε-quantile and

$$W_j(\boldsymbol{x_n}) = \frac{\mu_j(\boldsymbol{x_n}) - \theta_t}{\sigma_j(\boldsymbol{x_n})}, \qquad j = o, e.$$

First, from Equation (2) we have

$$\xi_e(\boldsymbol{x_n}) = 1 - \min\left\{\frac{b_1}{b_2}\frac{1-\Phi(W_e(\boldsymbol{x_n}))}{\Phi(W_e(\boldsymbol{x_n}))}, \frac{b_2}{b_1}\frac{\Phi(W_e(\boldsymbol{x_n}))}{1-\Phi(W_e(\boldsymbol{x_n}))}\right\},$$

Then, from (3), it follows that

$$\mathscr{X}_j^{(1)} = \left\{\boldsymbol{x_n} \in \mathscr{X} : W_j(\boldsymbol{x_n}) = \frac{\mu_j(\boldsymbol{x_n}) - \theta_t}{\sigma_j(\boldsymbol{x_n})} < z_{1-\varepsilon}\right\}$$

and finally

$$\mathscr{X}_{o,e} = \{\boldsymbol{x_n} \in \mathscr{X} : W_m(\boldsymbol{x_n}) < z_{1-\varepsilon} < W_M(\boldsymbol{x_n})\}, \tag{5}$$

where $W_m(\boldsymbol{x_n}) = \min\{W_o(\boldsymbol{x_n}), W_e(\boldsymbol{x_n})\}$ and $W_M(\boldsymbol{x_n}) = \max\{W_o(\boldsymbol{x_n}), W_e(\boldsymbol{x_n})\}$.

4.1 Numerical example

Let us consider $\theta_t = 1$ and let the design prior be a Normal density of parameters $\mu_d = 1.5$, $n_d = 10$. Thus, π_d assigns to H_1 a prior probability as small as 0.056. Figure 1 shows the behavior of e_n as n increases, under two alternative choices of μ_e for different values of the prior sample sizes n_e and n_o. In the former case, we assume that there is a certain contrast between the two priors: π_e, centred on the threshold θ_t (e.g. $\mu_e = 1$), expresses a neutral attitude towards the two hypotheses, whereas π_o favors the null hypothesis (e.g. $\mu_o = 0$). In the left panel of Figure 1 for small values of n (due to the predominant role of the prior weights n_e and n_o) e_n increases up to a maximum value and then it definitively decreases, tending to zero more and more rapidly for smaller values of the prior sample sizes n_e and n_o. In the latter set-up, the conflict between π_e and π_o is emphasized, π_e supports the alternative hypothesis H_2 and μ_e is even larger than μ_d (i.e. $\mu_o = 0$ and $\mu_e = 2$). As shown in the right panel, e_n monotonically decreases as a function of n from 1 to 0. As before, when the two conflicting priors are more and more concentrated, the expected value of $\bar{A}_{o,e}$ is uniformly larger and, consequently, a larger number of observations is required for the conflict to be resolved.

Finally in Figure 2 we illustrate by examples the relationship that holds in general between e_n and p_n, that is $p_n < e_n$, as commented in the final remark of Section 3.

5 Future research directions

The article leaves open the possibility of further developments, such as the application to non-normal models and to more challenging (not necessarily one-dimensional) testing set-ups. Moreover instead of considering only one prior π_e, we could extend our approach by considering an entire class of priors Γ. In this case, we

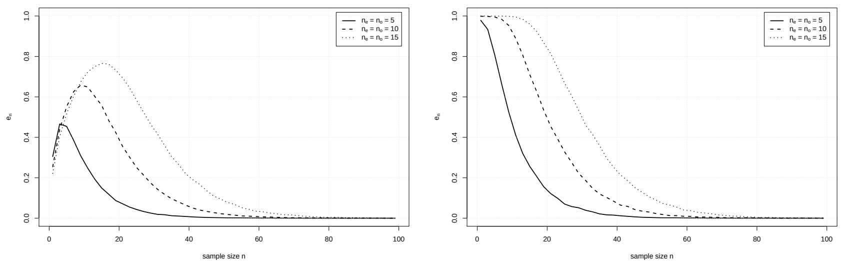

Fig. 1 e_n as a function of the sample size n, with $\mu_e = 1$ (*left panel*) and $\mu_e = 2$ (*right panel*) for different values of n_e and n_o, given $\theta_t = 1$, $\sigma = 1$, $\mu_d = 1.5$, $n_d = 10$, $\mu_o = 0$

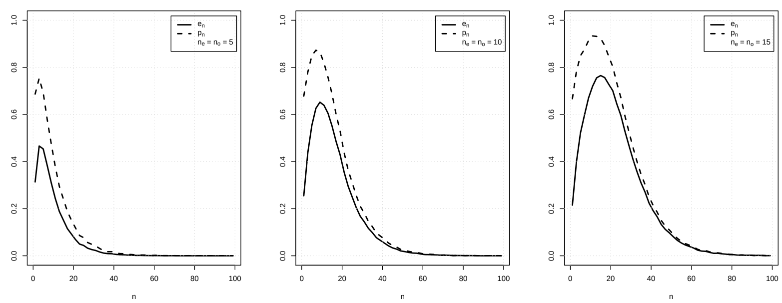

Fig. 2 Behavior of e_n and p_n for increasing values of n, with $\mu_e = 1$ (first row) and $\mu_e = 2$ (second row) for different values of n_o and n_e, given $\theta_t = 1$, $\sigma = 1$, $\mu_d = 1.5$, $n_d = 10$, $\mu_o = 0$.

would be interested in looking at the largest relative additional loss of a_o as π_e varies in Γ and the sample size is chosen by replacing e_n with $e_n^{\Gamma} = \mathbb{E}_{m_d}[\sup_{\pi_e \in \Gamma} \bar{A}_{o,e}]$.

References

1. Brutti P, De Santis F, Gubbiotti S. Predictive measures of the conflict between frequentist and Bayesian estimators. *Journal of Statistical Planning and Inference* 2014; **148**:111-122.
2. Brutti P, De Santis F, Gubbiotti S. Bayesian frequentist sample size determination: A game of two priors. *Metron* 2014; **72**(2):133-151.
3. De Santis F., Gubbiotti S. A decision-theoretic approach to sample size determination under several priors. *Applied Stochastic Models in Business and Industry* 2017; **33**(3):282-295.
4. Etzioni R, Kadane JB. Optimal experimental design for anothers analysis. *J. Am. Stat. Assoc.* 1993; **88**(424):1404-1411.
5. Kadane JB, Seidenfeld T. Randomization in a Bayesian perspective. *Journal of Statistical Planning and Inference* 1989; **25**:329-345.
6. Lindley DV, Singpurwalla N. On the Evidence Needed to Reach Agreed Action Between Adversaries, With Application to Acceptance Sampling. *Journal of the American Statistical Association* 1991; **86**(416):933-937.