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

Fulvio De Santis - Stefania Gubbiotti

Sapienza University of Rome

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.

Keywords: Bayesian inference, Experimental design, Exponential family, Predictive analysis, Sample size determination, Statistical decision theory.

1 Introduction

For a decision problem involving the unknown parameter of a statistical model, consider two decision makers, $\mathcal{E}_e$ and $\mathcal{E}_o$, who have in common the same data and loss function but that do not share the same prior information and/or opinions on the parameter. Let π_e and π_o denote the prior distributions of the parameter chosen by $\mathcal{E}_e$ and $\mathcal{E}_o$ and let a_e and a_o be the actions that minimize the two posterior expected losses, i.e. the two optimal Bayesian actions of the decision makers. Furthermore, let us suppose that $\mathcal{E}_e$ is forced to take the action a_o, which 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 planner of the experiment is a third actor, $\mathcal{P}_d$, who has to select the sample size of the experiment on the grounds of a predictive distribution of the data based on a prior π_d. We assume that, in general, π_d is different from both π_o and π_e. The present article focuses on determination of minimal sample sizes such that a relative predictive measure of the additional loss incurred by $\mathcal{E}_e$ in taking the action a_o rather than a_e is sufficiently small.

In the following, as a specific inferential decision problem, we will examine closely the testing set-up. To motivate our decisional approach, consider a testing problem designed to show whether the effect of a drug (the location parameter θ of a model) is larger than a given threshold. Three parties take part to the testing experiment: an optimistic planner $\mathcal{P}_d$ (whose prior assigns relevant probability to the null hypothesis), an optimistic decision maker $\mathcal{E}_e$, who is moderately in favour of the null hypothesis, and a second decision maker, $\mathcal{E}_o$, who is more skeptical towards the effectiveness of the drug. Let us assume $\mathcal{E}_e$ has to use the test statistics of $\mathcal{E}_o$, who bases her/his choices on the skeptical prior π_o but that s/he evaluates this action according to the expected posterior loss based on her/his moderately enthusiastic prior π_e. We expect that, at least for relatively small sample sizes, the initial attitudes of the two decision makers matter a lot and that the discrepancy in the posterior expected losses (both evaluated by $\mathcal{E}_e$) associated to the use of the two test statistics is relevant. However, for larger sample sizes, this difference tends to reduce more and more. We are interested in determining how large the sample size must be so that the (predictive expected) loss incurred by $\mathcal{E}_e$ in using the test statistics of $\mathcal{E}_o$ - i.e. in taking the non-optimal choice a_o - is sufficiently close to its minimum.

The article provides essentially two contributions. First, it extends the result of the point estimation setting in De Santis and Gubbiotti (2017) [10] to the testing problem. Furthermore, it introduces a relative predictive measure of the additional loss of a non-optimal action, which makes it easier to fix thresholds for the sample size criteria, i.e. to say when a given sample size is appropriate or not.

Statistical decision problems under several actors have been previously considered, for instance, in Kadane and Seidenfeld (1989) [17], and Lindley and Singpurwalla (1991) [19]. Similar scenarios are considered in Etzioni and Kadane (1993) [11], Spiegelhalter and Freedman (1988) [28], Kadane (1990) [16] where, however, the roles of the planner $\mathcal{P}_d$ and of the decision maker $\mathcal{E}_e$ coincide. The distinction between $\mathcal{P}_d$, $\mathcal{E}_e$ and $\mathcal{E}_o$ - i.e. between the three priors π_d, π_e and π_o - has been considered by Brutti et al. (2014) [5] (in a context not formalized as a decision problem). More recently, De Santis and Gubbiotti (2017) [10] have considered the problem from a formal decision theoretic perspective for point estimation.

The article relies widely also on the literature on Bayesian sample size determination

(SSD). From a decision-theoretic point of view, see, for instance: Raiffa and Schlaifer (1961) [23], Berger (1985) [2], Bernardo (1997) [3], Pham-Gia (1997) [22], Lindley (1997) [18], Parmigiani and Inoue (2009) [25]. For non-decision theoretic methods (i.e. *performance-based* approaches) see, among others: Spiegelhalter and Freedman (1986) [27], Adcock (1997) [1], Joseph and Belisle (1997) [13], Joseph et al. (1997) [14], Joseph & Wolfson (1997) [15] Spiegelhalter et al (2004) [29], Weiss (1996) [33]. For the literature on the so-called *two-priors approach* in Bayesian SSD, see: Tsutukawa (1972) [30] and, more recently, O'Hagan and Stevens (2001) [21], Wang and Gelfand (2002) [31], De Santis (2006) [9], Sahu and Smith (2006) [24], M'Lan et al (2006) [20], Sambucini (2010) [26], Brutti et al (2014) [6] and Cellamare and Sambucini (2014) [8]. The topic of the article is also related to the wider area of agreement/consensus in Bayesian decision theory and to adversarial risk analysis. See, among others, Burt (1990) [7], Jackson et al. (1980) [12], Weerahandi and Zidek (1981) [32].

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 2.1 we briefly discuss how to determine the expression of the relative measure in point and set estimation. In Section 3 the methodology is developed for a generic testing problem for a real-valued parameter. Results are then specialized to one-sided testing (Section 3.1) and to one-sided testing of a normal mean (Section 4). Numerical results are provided in Section 4.1. Finally, Section 5 contains discussion and comments.

2 Methodology

Assume that $X_1, X_2, \ldots, X_n$ is a sample from $f_n(\cdot|\theta)$, where θ is an unknown parameter and Θ is the parameter space. Let a denote a generic action for a decision problem regarding θ, $\mathcal{A}$ the action space and $L(a, \theta)$ the loss of a when the true parameter value is θ. Following the Bayesian inferential approach, we assume that θ is a random variable and that two competing priors are available, π_o and π_e. Given an observed sample $\boldsymbol{x_n} =$

$(x_1, x_2, \ldots, x_n)$, let

$$\pi_j(\theta|\boldsymbol{x_n}) = \frac{f_n(\boldsymbol{x_n}|\theta)\pi_j(\theta)}{\int_\Theta f_n(\boldsymbol{x_n}|\theta)\pi_j(\theta)d\theta}$$

be the posterior distribution of θ from prior π_j, and

$$\rho_j(\boldsymbol{x_n}, a) = \mathbb{E}_{\pi_j}\big[L(a,\theta)|\boldsymbol{x_n}\big] = \int\limits_\Theta L(a,\theta)\pi_j(\theta|\boldsymbol{x_n})d\theta$$

be the posterior expected loss of an action a, for $j = o, e$. Let

$$a_j = a_j(\boldsymbol{x_n}) = \arg\min_{a\in\mathcal{A}} \rho_j(\boldsymbol{x_n}, a)$$

denote the optimal action with respect to $\pi_j(\theta|\boldsymbol{x_n})$. The performance of the action a_o (optimal under π_o) when the expected loss is evaluated with respect to $\pi_e(\theta|\boldsymbol{x_n})$ is then

$$\rho_e(\boldsymbol{x_n}, a_o) = \mathbb{E}_{\pi_e}\big[L(a_o,\theta)|\boldsymbol{x_n}\big].$$

If $\mathcal{E}_e$ uses a_o instead of the π_e-optimal action a_e, the additional expected loss is

$$A_{o,e}(\boldsymbol{x_n}) = \rho_e(\boldsymbol{x_n}, a_o) - \rho_e(\boldsymbol{x_n}, a_e)$$

and a relative measure is given by

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

Small values of $\bar{A}_{o,e}$ show that 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 random variable. We assume that, as n increases, $\bar{A}_{o,e}(\boldsymbol{X_n})$ converges (? in probability ?) to zero. We are interested in selecting the smallest sample size such that its expected value is smaller than a selected threshold γ, that is:

$$n^\star = \min\{n \in \mathbb{N} : e_n \le \gamma\}, \tag{1}$$

where

$$e_n = \mathbb{E}_{m_d}\big[\bar{A}_{o,e}(\boldsymbol{X_n})\big] \tag{2}$$

and where $\mathbb{E}_{m_d}[\cdot]$ denotes the expected value with respect to the sample data distribution, m_d. Following the predictive Bayesian approach, we consider

$$m_d(\boldsymbol{x_n}) = \int_\Theta f(\boldsymbol{x_n}|\theta)\pi_d(\theta)d\theta,$$

where π_d is the design prior. Therefore the optimal sample size $n^\star$ depends on three priors (π_d, π_e, π_o). In the most general case, π_d is different from both π_e and π_o. If π_d coincides with π_e, we retrieve the approach of Etzioni and Kadane (1993) [11]. Moreover, if π_d is a point-mass prior on a design value θ_d, then m_d is the sampling distribution $f(\cdot|\theta_d)$, yielding a *conditional* Bayes approach to SSD [6].

Explicit expressions of $\bar{A}_{o,e}(\boldsymbol{x_n})$ and e_n depend on the decision problem (point estimation, set estimation, test) and on the choice of the loss function. Even though the framework described above holds in more general contexts, in the following we assume that $\theta \in \Theta \subset \mathbb{R}$.

2.1 Point and set estimation

For point estimation, using the quadratic loss $L(a, \theta) = (\theta - a)^2$, it is easy to check that

$$A_{o,e}(\boldsymbol{x_n}) = (a_o - a_e)^2 = \left(\mathbb{E}_{\pi_e}\left[\theta|\boldsymbol{x_n}\right] - \mathbb{E}_{\pi_o}\left[\theta|\boldsymbol{x_n}\right]\right)^2$$

and

$$\bar{A}_{o,e}(\boldsymbol{x_n}) = \frac{\left(\mathbb{E}_{\pi_e}\left[\theta|\boldsymbol{x_n}\right] - \mathbb{E}_{\pi_o}\left[\theta|\boldsymbol{x_n}\right]\right)^2}{\left(\mathbb{E}_{\pi_e}\left[\theta|\boldsymbol{x_n}\right] - \mathbb{E}_{\pi_o}\left[\theta|\boldsymbol{x_n}\right]\right)^2 + \mathbb{V}_{\pi_e}[\theta|\boldsymbol{x_n}]},$$

where $\mathbb{E}_{\pi_j}\left[\theta|\boldsymbol{x_n}\right]$ and $\mathbb{V}_{\pi_j}[\theta|\boldsymbol{x_n}]$ denote the posterior expected value and variance of θ, $j = e, o$. Note that if π_o is a noninformative prior distribution, the optimal action a_o is typically the frequentist MLE and $A_{o,e}$ becomes the measure of the conflict between the MLE and the Bayesian estimate based on π_e, considered in Brutti et al (2014) [5]. Within the decision-theoretic framework the problem has been considered in [10].

Turning to set estimation, for a generic $S \subset \Theta$, a standard choice is the linear loss function

$$L(S, \theta) = b \cdot \text{mis}(S) - 1_S(\theta), \qquad b > 0.$$

Then,

$$\rho_j(\boldsymbol{x_n}, S) = b \cdot \text{mis}(S) - \mathbb{P}_{\pi_j}[S|\boldsymbol{x_n}]$$

where $\text{mis}(S)$ denotes the Lebesgue measure of the set S and $\mathbb{P}_{\pi_j}[S|\boldsymbol{x_n}]$ its posterior probability w.r.t $\pi_j(\theta|\boldsymbol{x_n})$, $j = e, o$. Therefore, for the optimal sets S_e and S_o we have

$$A_{o,e}(\boldsymbol{x_n}) = b[\text{mis}(S_o) - \text{mis}(S_e)] + (\mathbb{P}_{\pi_e}[S_e|\boldsymbol{x_n}] - \mathbb{P}_{\pi_e}[S_o|\boldsymbol{x_n}])$$

and

$$\bar{A}_{o,e}(\boldsymbol{x_n}) = \frac{b[\mathrm{mis}(S_o) - \mathrm{mis}(S_e)] + (\mathbb{P}_{\pi_e}[S_e|\boldsymbol{x_n}] - \mathbb{P}_{\pi_e}[S_o|\boldsymbol{x_n}])}{b \cdot \mathrm{mis}(S_o) - \mathbb{P}_{\pi_e}[S_o|\boldsymbol{x_n}]}.$$

Therefore, the additional loss $A_{o,e}$ is decomposed into the sum of (b times) the difference in length between S_o and S_e and of the extra posterior coverage of S_e w.r.t. S_o, evaluated with $\pi_e(\theta|\boldsymbol{x_n})$.

3 Testing

Consider $H_1 : \theta \in \Theta_1$ vs. $H_2 : \theta \in \Theta_2$, where $\Theta = \{\Theta_1, \Theta_2\}$ is a partition of Θ. Using the prior distribution π_j, $j = o, e$, the Bayes factor for testing H_2 vs. H_1 is

$$B_{21}^j(\boldsymbol{x_n}) = \frac{\omega_{21}^j(\boldsymbol{x_n})}{\omega_{21}^j},$$

where

$$\omega_{21}^j(\boldsymbol{x_n}) = \frac{\mathbb{P}_{\pi_j}(\Theta_2|\boldsymbol{x_n})}{\mathbb{P}_{\pi_j}(\Theta_1|\boldsymbol{x_n})} \qquad \text{and} \qquad \omega_{21}^j = \frac{\mathbb{P}_{\pi_j}(\Theta_2)}{\mathbb{P}_{\pi_j}(\Theta_1)}$$

are the posterior and the prior odds. Let $\mathcal{A} = \{a^{(1)}, a^{(2)}\}$ be the two terminal decisions, where $a^{(i)}$ denotes the choice of H_i, $i = 1, 2$ and

$$L(a^{(1)}, \theta) = b_2 \times 1_{\Theta_2}(\theta) \quad \text{and} \quad L(a^{(2)}, \theta) = b_1 \times 1_{\Theta_1}(\theta)$$

their loss functions ($b_i > 0,\ i = 1, 2$). The posterior expected losses of $a^{(1)}$ and $a^{(2)}$ w.r.t. $\pi_j(\theta|\boldsymbol{x_n})$ are

$$\rho_j(\boldsymbol{x_n}, a^{(1)}) = b_2\mathbb{P}_{\pi_j}[\Theta_2|\boldsymbol{x_n}], \qquad \text{and} \qquad \rho_j(\boldsymbol{x_n}, a^{(2)}) = b_1\mathbb{P}_{\pi_j}[\Theta_1|\boldsymbol{x_n}].$$

In this case it is easy to check that the optimal decision function $a_j(\boldsymbol{x_n})$ is

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

where

$$\mathcal{Z}_j^{(1)} = \{\boldsymbol{x_n} :\ \rho_j(\boldsymbol{x_n}, a^{(1)}) < \rho_j(\boldsymbol{x_n}, a^{(2)})\} = \{\boldsymbol{x_n} :\ b_2\mathbb{P}_{\pi_j}[\Theta_2|\boldsymbol{x_n}] < b_1\mathbb{P}_{\pi_j}[\Theta_1|\boldsymbol{x_n}]\}$$

and

$$\mathcal{Z}_j^{(2)} = \{\boldsymbol{x_n} :\ \rho_j(\boldsymbol{x_n}, a^{(1)}) > \rho_j(\boldsymbol{x_n}, a^{(2)})\} = \{\boldsymbol{x_n} :\ b_2\mathbb{P}_{\pi_j}[\Theta_2|\boldsymbol{x_n}] > b_1\mathbb{P}_{\pi_j}[\Theta_1|\boldsymbol{x_n}]\}.$$

The posterior expected loss of the optimal decision function $a_j(\boldsymbol{x_n})$ w.r.t. π_e is

$$\rho_e(\boldsymbol{x_n}, a_j) = \begin{cases} b_2\mathbb{P}_{\pi_e}[\Theta_2|\boldsymbol{x_n}] & \text{if} \quad \boldsymbol{x_n} \in \mathcal{Z}_j^{(1)} \\ b_1\mathbb{P}_{\pi_e}[\Theta_1|\boldsymbol{x_n}] & \text{if} \quad \boldsymbol{x_n} \in \mathcal{Z}_j^{(2)} \end{cases} \qquad j = o, e.$$

Therefore, noting that

$$\rho_e(\boldsymbol{x_n}, a_e) = \min\{b_1\mathbb{P}_{\pi_e}[\Theta_1|\boldsymbol{x_n}], b_2\mathbb{P}_{\pi_e}[\Theta_2|\boldsymbol{x_n}]\}$$

it follows that

$$\bar{A}_{o,e}(\boldsymbol{x_n}) = \begin{cases} 0 & \text{if} \quad a_o(\boldsymbol{x_n}) = a_e(\boldsymbol{x_n}) \\ \xi_e(\boldsymbol{x_n}) & \text{if} \quad a_o(\boldsymbol{x_n}) \neq a_e(\boldsymbol{x_n}) \end{cases} = \xi_e(\boldsymbol{x_n}) 1_{\mathcal{Z}_{o,e}(\boldsymbol{x_n})} \tag{3}$$

where 1_A is the indicator function of the set A,

$$\begin{aligned} \xi_e(\boldsymbol{x_n}) &= \frac{\rho_e(\boldsymbol{x_n}, a_o) - \rho_e(\boldsymbol{x_n}, a_e)}{\rho_e(\boldsymbol{x_n}, a_o)} = \\ &= 1 - \frac{\min\{b_1\mathbb{P}_{\pi_e}[\Theta_1|\boldsymbol{x_n}], b_2\mathbb{P}_{\pi_e}[\Theta_2|\boldsymbol{x_n}]\}}{\max\{b_1\mathbb{P}_{\pi_e}[\Theta_1|\boldsymbol{x_n}], b_2\mathbb{P}_{\pi_e}[\Theta_2|\boldsymbol{x_n}]\}} = \\ &= 1 - \min\left\{\frac{k^e}{B_{21}^e(\boldsymbol{x_n})}, \frac{B_{21}^e(\boldsymbol{x_n})}{k^e}\right\}, \qquad k_j = \frac{b_1}{b_2}\,\omega_{12}^j, \end{aligned} \tag{4}$$

and

$$\mathcal{Z}_{o,e} = \{\boldsymbol{x_n} \in \mathcal{X}^n : a_o(\boldsymbol{x_n}) \neq a_e(\boldsymbol{x_n})\}$$

is the set of $\boldsymbol{x_n}$ leading to conflicting terminal decisions under π_e and π_o respectively. Therefore

$$\mathcal{Z}_{o,e} = \left(\mathcal{Z}_o^{(1)} \cap \mathcal{Z}_e^{(2)}\right) \cup \left(\mathcal{Z}_o^{(2)} \cap \mathcal{Z}_e^{(1)}\right),$$

where

$$\mathcal{Z}_j^{(1)} = \left\{\boldsymbol{x_n} \in \mathcal{Z} : \omega_{21}^j(\boldsymbol{x_n}) < \frac{b_1}{b_2}\right\} = \{\boldsymbol{x_n} \in \mathcal{Z} : B_{21}^j(\boldsymbol{x_n}) < k_j\}$$

and

$$\mathcal{Z}_j^{(2)} = \left\{\boldsymbol{x_n} \in \mathcal{Z} : \omega_{21}^j(\boldsymbol{x_n}) > \frac{b_1}{b_2}\right\} = \{\boldsymbol{x_n} \in \mathcal{Z} : B_{21}^j(\boldsymbol{x_n}) > k_j\}.$$

Finally, from (2) and (3) we obtain

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

Remarks

i) From the above expression we can note that e_n is a monotone function of the Lebesgue measure of $\mathcal{Z}_{o,e}$. One could argue that, in our contexts, an alternative - and structurally simpler - sample size criterion could be based on $p_n = \mathbb{P}_{m_d}[\mathcal{Z}_{o,e}]$, the predictive probability of the set of samples yielding conflict. It is easy to check that $p_n = \mathbb{E}_{m_d}[I_{\mathcal{Z}_{o,e}}(\boldsymbol{X_n})]$, whereas $e_n = \mathbb{E}_{m_d}[\xi_e(\boldsymbol{X_n})I_{\mathcal{Z}_{o,e}}(\boldsymbol{X_n})]$. Recalling that, $\forall \boldsymbol{x_n} \in \mathcal{Z}$, $\xi_e(\boldsymbol{x_n}) \leq 1$, then e_n is always smaller than or equal to p_n. Therefore, given a threshold γ, the sample size needed to have $e_n \leq \gamma$ is always smaller than or equal to the sample size needed to make $p_n \leq \gamma$. The idea is that, in the expectation that defines e_n, the contribution of each sample $\boldsymbol{x_n}$ that would determine a conflicting decision depends on the strength of the discrepancy in evidence it gives to the two hypotheses. Conversely, in p_n, the contribution of each sample $\boldsymbol{x_n}$ such that $a_o(\boldsymbol{x_n}) \neq a_e(\boldsymbol{x_n})$ is inflexibly equal to one, regardless of the evidence it gives to the competing hypotheses.

ii) In general the explicit expression of e_n is not available, but it is straightforward to obtain its Monte Carlo approximation

$$e_n \approx \frac{1}{M}\sum_{r=1}^{M} \xi_e(\boldsymbol{x_n}^{(r)}) 1_{\mathcal{Z}_{o,e}}(\boldsymbol{x_n}^{(r)}),$$

where $\boldsymbol{x_n}^{(r)}$, $r = 1, \ldots, M$ are drawn from the predictive distribution $m_d(\cdot)$.

3.1 One-sided testing

The above results can be specialized to the one-sided testing set-up, where $H_1 : \theta \leq \theta_t$ vs. $H_2 : \theta > \theta_t$, with $\theta_t \in \mathbb{R}$. We show that $\bar{A}_{o,e}$ can be expressed in terms of the posterior c.d.f.'s of θ, $F_j(\cdot|\boldsymbol{x_n})$, and of their quantiles, $q^j_\epsilon(\boldsymbol{x_n})$. First, from Equation (4) we have

$$\xi_e(\boldsymbol{x_n}) = 1 - \min\left\{\frac{b_1}{b_2}\frac{\mathrm{F}_e(\theta_t|\boldsymbol{x_n})}{1-\mathrm{F}_e(\theta_t|\boldsymbol{x_n})}, \frac{b_2}{b_1}\frac{1-\mathrm{F}_e(\theta_t|\boldsymbol{x_n})}{\mathrm{F}_e(\theta_t|\boldsymbol{x_n})}\right\}. \tag{5}$$

To obtain $\mathcal{Z}_{o,e}$, note that

$$\mathcal{Z}_j^{(1)} = \left\{\boldsymbol{x_n} \in \mathcal{Z} : \frac{1-\mathrm{F}_j\left(\theta_t|\boldsymbol{x_n}\right)}{\mathrm{F}_j\left(\theta_t|\boldsymbol{x_n}\right)} < \frac{b_1}{b_2}\right\} = \left\{\boldsymbol{x_n} \in \mathcal{Z} : \theta_t > q^j_\epsilon(\boldsymbol{x_n})\right\} \tag{6}$$

where $\epsilon = \frac{b_2}{b_1+b_2}$ and $q_\epsilon^j(\boldsymbol{x_n})$ is the ϵ-quantile of the posterior distribution of θ. Therefore,

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

where

$$q_\epsilon^m(\boldsymbol{x_n}) = \min\left\{q_\epsilon^e(\boldsymbol{x_n}), q_\epsilon^o(\boldsymbol{x_n})\right\} \quad \text{and} \quad q_\epsilon^M(\boldsymbol{x_n}) = \min\left\{q_\epsilon^e(\boldsymbol{x_n}), q_\epsilon^o(\boldsymbol{x_n})\right\} \tag{7}$$

and

$$\mathcal{Z}_{o,e} = \left\{\boldsymbol{x_n} \in \mathcal{Z} : q_\epsilon^m(\boldsymbol{x_n}) < \theta_t < q_\epsilon^M(\boldsymbol{x_n})\right\}. \tag{8}$$

Let us now further assume that θ is a continuous r.v. and that F_j, $j = o, e$, are the c.d.f.'s of a location-scale model. Let $\mathbb{F}$ be the c.d.f. of the standardized r.v. $(\theta - \mu_j(\boldsymbol{x_n}))/\sigma_j(\boldsymbol{x_n})$, $\bar{q}_\epsilon$ its ϵ-quantile and

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

the two test statistics, where $\mu_j(\boldsymbol{x_n})$ and $\sigma_j(\boldsymbol{x_n})$ are the location and scale parameters of the posterior densities of θ. In this case $\bar{A}_{o,e}$ can be expressed in terms of $\mathbb{F}$, $\bar{q}_\epsilon$ and $W_j(\boldsymbol{x_n})$. First, from Equation (4) we have

$$\xi_e(\boldsymbol{x_n}) = 1 - \min\left\{\frac{b_1}{b_2}\frac{1 - \mathbb{F}(W_e(\boldsymbol{x_n}))}{\mathbb{F}(W_e(\boldsymbol{x_n}))}, \frac{b_2}{b_1}\frac{\mathbb{F}(W_e(\boldsymbol{x_n}))}{1 - \mathbb{F}(W_e(\boldsymbol{x_n}))}\right\}. \tag{9}$$

Then, from (6), it follows that

$$\begin{aligned}
\mathcal{Z}_j^{(1)} &= \left\{\boldsymbol{x_n} \in \mathcal{Z} : \mathbb{F}\left(\frac{\theta_t - \mu_j(\boldsymbol{x_n})}{\sigma_j(\boldsymbol{x_n})}\right) > \frac{b_2}{b_1 + b_2}\right\} \\
&= \left\{\boldsymbol{x_n} \in \mathcal{Z} : W_j(\boldsymbol{x_n}) = \frac{\mu_j(\boldsymbol{x_n}) - \theta_t}{\sigma_j(\boldsymbol{x_n})} < \bar{q}_{1-\epsilon}\right\}.
\end{aligned}$$

Therefore,

$$\mathcal{Z}_o^{(1)} \cap \mathcal{Z}_e^{(1)} = \left\{\boldsymbol{x_n} \in \mathcal{Z} : W_M(\boldsymbol{x_n}) < \bar{q}_{1-\epsilon}\right\} \quad \text{and} \quad \mathcal{Z}_o^{(2)} \cap \mathcal{Z}_e^{(2)} = \left\{\boldsymbol{x_n} \in \mathcal{Z} : W_m(\boldsymbol{x_n}) > \bar{q}_{1-\epsilon}\right\}.$$

Noting that $\mathcal{Z}_{o,e} = (\mathcal{Z}_o^{(1)} \cap \mathcal{Z}_e^{(1)})^C \cap (\mathcal{Z}_o^{(2)} \cap \mathcal{Z}_e^{(2)})^C$ we obtain

$$\mathcal{Z}_{o,e} = \left\{\boldsymbol{x_n} \in \mathcal{Z} : W_m(\boldsymbol{x_n}) < \bar{q}_{1-\epsilon} < W_M(\boldsymbol{x_n})\right\}, \tag{10}$$

where

$$W_m(\boldsymbol{x_n}) = \min\left\{W_o(\boldsymbol{x_n}), W_e(\boldsymbol{x_n})\right\} \quad \text{and} \quad W_M(\boldsymbol{x_n}) = \max\left\{W_o(\boldsymbol{x_n}), W_e(\boldsymbol{x_n})\right\}.$$

4 Results for the Normal mean

Let us now assume that $X_1, X_2, \ldots, X_n$ is a random sample, $X_i|\theta \sim \mathrm{N}(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$ and that $\pi_j(\cdot)$ are conjugate priors, i.e. $\theta|\sigma^2 \sim \mathrm{N}(\mu_j, \sigma^2/n_j)$, $j = o, e$. First, assume that σ^2 is known. Then the posterior distribution of θ is Normal with location and scale

$$\mu_j(\boldsymbol{x_n}) = \frac{n_j\mu_j + n\bar{x}_n}{n_j + n} \qquad \text{and} \qquad \sigma_j(\boldsymbol{x_n}) = \frac{\sigma}{\sqrt{n_j + n}}.$$

From Equation (9), setting $\mathbb{F}(\cdot) = \Phi(\cdot)$, we obtain

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

where $\Phi(\cdot)$ is the standard normal c.d.f.. From Equations (10) we have

$$\mathcal{Z}_{o,e} = \{\boldsymbol{x_n} \in \mathcal{Z} : W_m(\boldsymbol{x_n}) < z_{1-\epsilon} < W_M(\boldsymbol{x_n})\}$$

where $z_{1-\epsilon}$ is the $1 - \epsilon$ quantile of the standard normal. In this setting, we can express $\mathcal{Z}_{o,e}$ also in terms of a condition on the sample mean. In fact, it can be shown that

$$\mathcal{Z}_{o,e} = \{\boldsymbol{x_n} \in \mathcal{Z} : h_m < \bar{x}_n < h_M\}$$

where

$$h_j = \theta_t + \frac{n_j}{n}(\theta_t - \mu_j) - z_\epsilon\, \sigma\, \frac{\sqrt{n_j + n}}{n}, \qquad j = o, e$$

and

$$h_m = \min\{h_o, h_e\} \qquad \text{and} \qquad h_M = \max\{h_o, h_e\}.$$

Note that, as n increases, the contributions of the priors $\pi_j(\theta)$ in the corresponding posteriors tends to zero and the conflict between a_o and a_e vanishes: $h_j \to \theta_t$ and $\mathcal{Z}_{o,e}$ tends to the empty set.

Unknown variance. The extension to the unknown variance case is straightforward. Under the standard conjugacy assumptions (see for instance Bernardo and Smith, 1994 [4]),

$$\theta|\sigma^2 \sim \mathrm{N}\left(\mu_j, \frac{\sigma^2}{n_j}\right) \qquad \text{and} \qquad \sigma^2 \sim \mathrm{IGa}(\alpha_j, \beta_j),$$

the marginal posterior distribution of θ is a Student t distribution with parameters

$$\mu_j(\boldsymbol{x_n}) = \frac{n_j\mu_j + n\bar{x}_n}{n_j + n},\ \sigma_j(\boldsymbol{x_n}) = \frac{\beta_j + \frac{1}{2}nS_n^2 + \frac{1}{2}(n_j + n)^{-1}n_j n(\mu_j - \bar{x}_n)^2}{(n_j + n)(\alpha_j + \frac{n}{2})},\ \nu_j = 2\alpha_j + n.$$

In this case we cannot use the location-scale formulas and we must refer to the general one-sided testing set-up. The expressions of $\xi_e(\boldsymbol{x_n})$ and $\mathcal{Z}_{o,e}$ are obtained from Equations (5), (7) and (8) by setting F_j equal to the c.d.f. of a Student t of parameters $(\mu_j(\boldsymbol{x_n}), \sigma_j(\boldsymbol{x_n}), \nu_j)$ and q_ϵ^j equal to its $(1-\epsilon)$-quantile, $j = o, e$.

4.1 Numerical example

In this section we illustrate some numerical examples for the Normal case. 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 H_1 a prior probability as small as 0.056. In the following we show the behavior of the expected value of the relative expected additional loss as n increases, under two alternative choices of μ_e for different values of the prior sample sizes n_e and n_o.

First, let us assume that there is a certain contrast between the two priors, namely 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 Figure 1, for small values of the sample size 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.

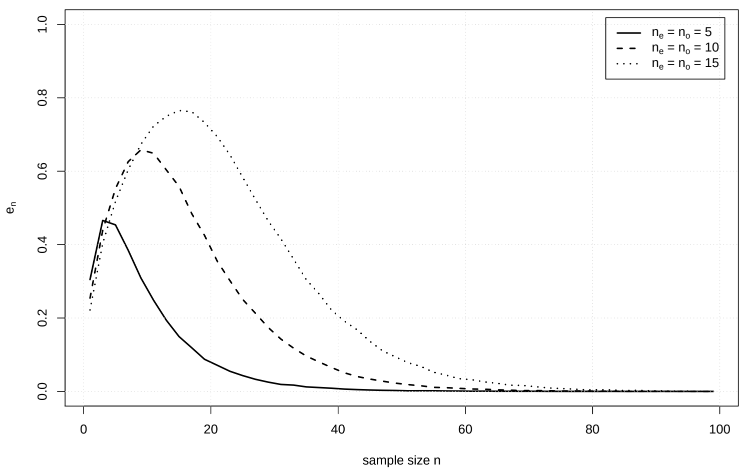

Figure 1: Expected value of the relative additional loss as a function of the sample size n, with $\mu_e = 1$ for different values of n_e and n_o, given $\theta_t = 1$, $\sigma = 1$, $\mu_d = 1.5$, $n_d = 10$, $\mu_o = 0$

In the second 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 $\mu_e = 2$). In Figure 2, 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, Table 1 displays the optimal sample sizes, obtained using criterion (1)

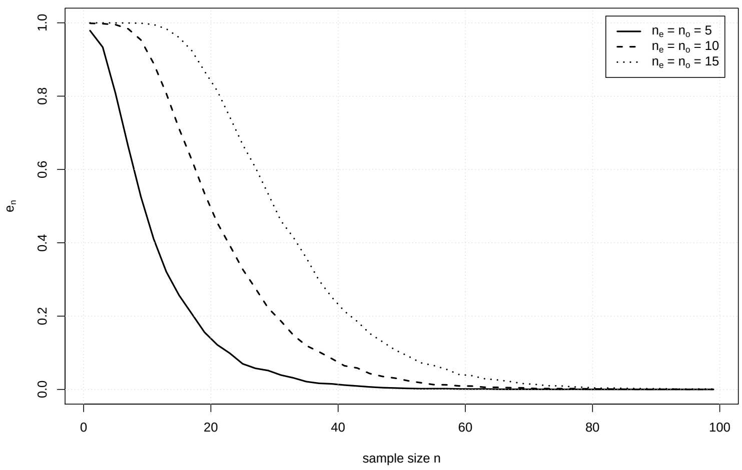

Figure 2: Expected value of the relative additional loss as a function of the sample size n, with $\mu_e = 2$ for different values of n_e and n_o, given $\theta_t = 1$, $\sigma = 1$, $\mu_d = 1.5$, $n_d = 10$, $\mu_o = 0$

for a given threshold $\gamma = 0.01$.

	$n_e = n_o = 5$	$n_e = n_o = 10$	$n_e = n_o = 15$
$\mu_e = 1$	37	56	71
$\mu_e = 2$	41	57	73

Table 1: Optimal sample sizes $n^{\star}$ based on criterion (1) with threshold $\gamma = 0.01$, for different values of μ_e, n_e and n_o, given $\theta_t = 1$, $\sigma = 1$, $\mu_d = 1.5$, $n_d = 10$, $\mu_o = 0$.

5 Conclusions

Statistical decision theory offers the ideal formal set-up to formalize sample size criteria, that can be based on the risks of the decision functions. In this paper we adopt the Bayesian perspective and we evaluate the performance of statistical decisions in terms of their posterior expected losses. The article extends and generalizes previous contributions (focused on point estimation) whose central feature is the presence of multiple actors (decision makers and planner) and the need of choosing a number of sample units such that the unavoidable additional loss one incurs by using a non-optimal action does not exceed too much the minimal loss of the best action (see [10]). Attention is here specifically devoted to the testing problem. We propose a sample size criterion built up on the predictive expected value (e_n) of the relative difference between the posterior expected losses of the chosen action and of the optimal action, respectively. A sample size such that e_n is close enough to zero guarantees that, on average, the consequence of using a non-optimal action is substantially equivalent to the consequence of taking the optimal decision. The expression of e_n for a general testing problem, given in Section 3, shows the role played both by the subset $\mathcal{Z}_{o,e}$ of the samples leading to discordant decision a_o and a_e and by the quantity $\xi_e(\boldsymbol{x_n})$, which takes into account the amount of evidence (quantified in terms of posterior probabilities or Bayes factors) a conflicting sample $\boldsymbol{x_n}$ gives to the two hypotheses. The criterion is then specialized to more specific settings: one-sided testing, testing of a location parameter, testing of a normal mean. In Section 4.1 we have considered some numerical examples in the context of testing the normal mean (known variance). Even in this basic framework, explicit expression of e_n are not available but standard Monte Carlo approximations can be easily obtained.
The article leaves open the possibility of further developments. Here are two possible areas for future work.

1. Application to non-normal models and to more challenging (not necessarily one-dimensional) testing set-ups. In these cases one should consider the possibility that, not only the expression of e_n is not available in closed form, but also the posterior probabilities (or the Bayes factors) that define $\mathcal{Z}_{o,e}$ and $\xi_e(\boldsymbol{x_n})$ may need numerical approximations. This would make the solution of the problem more intensive from

a computational point of view.

2. Instead of considering only one prior π_e, we could extend our approach by considering an entire class of prior distributions Γ. In this case, we would be interested in looking at the largest relative additional loss of a_o as π_e varies in Γ, that is $\sup_{\pi_e \in \Gamma} \bar{A}_{o,e}$. The sample size would then be chosen by replacing e_n in with $e_n^{\Gamma} = \mathbb{E}_{m_d}[\sup_{\pi_e \in \Gamma} \bar{A}_{o,e}]$ in expression (1).

References

[1] Adcock CJ. Sample Size Determination: A Review. *Journal of the Royal Statistical Society. Series D (The Statistician)* 1997; **46**(2):261-283.

[2] Berger JO. Statistical Decision Theory and Bayesian Analysis. 1985; Berlin: Springer-Verlag.

[3] Bernardo JM. Statistical inference as a decision problem: The choice of the sample size. *Statistician* 1997; **46**:151-153.

[4] Bernardo JM, Smith, AFM. Bayesian Theory. 1994; Wiley: Chichester.

[5] 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.

[6] Brutti P, De Santis F, Gubbiotti S. Bayesian frequentist sample size determination: A game of two priors. *Metron* 2014; **72**(2):133-151.

[7] Burt J. Towards Agreement: Bayesian Experimental Design by. Jameson Burt. Purdue University. *Technical Report #90-41. Department of Statistics* 1990,

[8] Cellamare M, Sambucini V. A randomized two-stage design for phase II clinical trials based on a Bayesian predictive approach. *Stat Med* 2015; **34**(6):1059–1078.

[9] De Santis F. Sample size determination for robust Bayesian analysis. *J. Am. Stat. Assoc.* 2006; **101**(473):278-291.

[10] 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.

[11] Etzioni R, Kadane JB. Optimal experimental design for anothers analysis. *J. Am. Stat. Assoc.* 1993; **88**(424):1404-1411.

[12] Jackson PH, Novick MR, DeKeyrel, DF. Adversary Preposterior Analysis for Simple Parametric Models in Bayesian Analysis in Econometrics and Statistics, ed. Arnold Zellner, Amsterdam: North-Holland. 1980; 113-132.

[13] Joseph L, Belisle P. Bayesian sample size determination for normal means and difference between normal means. *Statistician* 1997; **46**:209-226.

[14] Joseph L, du Berger R, Belisle P. Bayesian and mixed Bayesian/likelihood criteria for sample size determination. *Stat. Med.* 1997; **16**:769-781.

[15] Joseph L, Wolfson D. Interval-based versus decision theoretic criteria for the choice of sample size. *Statistician* 1997; **46**:145-149.

[16] Kadane JB. A Statistical Analysis of Adverse Impact of Employer Decisions. *Journal of the American Statistical Association* 1990; **85**(412):925-933.

[17] Kadane JB, Seidenfeld T. Randomization in a Bayesian perspective. *Journal of Statistical Planning and Inference* 1989; **25**:329-345.

[18] Lindley DV. The choice of sample size. *The Statistician* 1997; **46**:129-138.

[19] 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.

[20] M'Lan CE, Joseph L, Wolfson DB. Bayesian Sample Size Determination for Case-control Studies. *J. Am. Stat. Assoc.* 2006; **101**(474):760-772.

[21] O'Hagan A, Stevens JW. Bayesian assessment of sample size for clinical trials for cost effectiveness. *Med. Decis. Mak.* 2001; **21**:219-230.

[22] Pham-Gia T. On Bayesian analysis, Bayesian decision theory and the sample size problem. *Statistician* 1997; **46**:139-144.

[23] Raiffa H, Schlaifer R. Applied statistical decision theory. Boston: Clinton Press, Inc. 1961.

[24] Sahu SK, Smith TMF. A Bayesian method of sample size determination with practical applications. *J. Roy. Stat. Soc. A Sta.* 2006; **169**:235-253.

[25] Parmigiani G, Inoue L. Decision Theory: Principles and Approaches. Wiley Series in Probability and Statistics 2009.

[26] Sambucini V. A Bayesian predictive strategy for an adaptive two-stage design in phase II clinical trials. *Stat Med* 2010; **29**(13):1430–1442.

[27] Spiegelhalter DJ, Freedman LS. A predictive approach to selecting the size of a clinical trial, based on subjective clinical opinion. *Stat. Med.* 1986; **5**:1-13.

[28] Spiegelhalter DJ, Freedman LS. Bayesian approaches to clinical trials (with discussion). In Bayesian Statistics, 3, eds J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, 453-477. Oxford University Press, Oxford 1988.

[29] Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian approaches to clinical trials and health-care evaluation. Wiley, New York 2004.

[30] Tsutakawa RK. Design of experiment for bioassay. *J. Am. Stat. Assoc.* 1972; **67**(339):585-590.

[31] Wang F, Gelfand AE. A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. *Stati. Sci.* 2002; **17**(2):193-208.

[32] Weerahandi S, Zidek JV. Multi-bayesian statistical decision theory. *Journal of the Royal Statistical Society. Series A. General* 1981; **144**(1):85-93.

[33] Weiss RE. Bayesian Sample Size Computations for Hypothesis Testing. *Statistician* 1997; **46**:185-191.