On the predictive performance of a non-optimal action in hypothesis testing

Fulvio De Santis[1] · Stefania Gubbiotti[1]

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 a general testing set-up.

Keywords Bayesian inference · Experimental design · Predictive analysis · Sample size determination · Statistical decision theory

1 Introduction

A decision making process typically involves more than one actor who play different roles—expert, stakeholder, analyst,...—with different responsibilities and authority. Suppose that, for a statistical decision problem on the unknown parameter θ of a model, two main actors are involved, $\mathcal{E}_e$ 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 respective posterior expected losses, (i.e. the two optimal Bayesian actions of the two actors). Let us also suppose that the relative authority of the two actors is not equivalent: $\mathcal{E}_e$ is forced to take the action a_o, which is not optimal from her/his point of view. The use of this inappropriate π_o implies an extra loss for $\mathcal{E}_e$. In fact, under π_e, the posterior expected loss of a_o is larger than the posterior expected loss of a_e.

✉ Stefania Gubbiotti
  stefania.gubbiotti@uniroma1.it

[1] Dipartimento di Scienze Statistiche, Sapienza Università di Roma, Piazzale Aldo Moro n.5, 00185 Rome, Italy

The additional loss incurred by $\mathcal{E}_e$ in taking the action a_o rather than a_e can be evaluated for any observed sample, but, before collecting the data, it is a random quantity. The present article studies the predictive distribution of this additional loss and its use for sample size determination (SSD). Preposterior analysis requires specification of a data generating mechanism. This is formalized by a probability distribution that describes a prefixed design scenario. This can be done in two alternative ways: (a) using the sampling distribution for a fixed value of the parameter θ_d; (b) assigning a prior distribution π_d to the parameter and then using the marginal distribution of the data. In general, the value of θ_d or the choice of π_d represent the goal of the experiment and they do not necessarily match with the information expressed by π_o or π_e. We refer to the actor who plans the experiment and chooses the design scenario as to $\mathcal{P}_d$. To motivate our decisional framework, we provide two examples.

Example 1 *(Efficacy clinical trial)* Consider a clinical trial whose goal is to show whether the effect of a drug (the location parameter θ of a model) developed by a pharmaceutical company, is larger than a given threshold θ_t. Two actors take part in the testing experiment: the company, represented by an optimistic decision maker $\mathcal{E}_e$, who, based on historical data, is in favour of the null hypothesis, $H_1 : \theta \geq \theta_t$; the regulatory agency, represented by the actor $\mathcal{E}_o$, who *a priori* is neutral towards the effectiveness of the drug. Therefore

- $\mathcal{E}_e$ specifies a (moderately) optimistic prior, possibly downweighting the available historical information,
- $\mathcal{E}_o$ adopts a non informative prior distribution to avoid inclusion of external information into the analysis.

When the regulatory agency $\mathcal{E}_o$ prescribes to use a_o, the company $\mathcal{E}_e$ is interested in evaluating the additional expected posterior loss based on its own prior π_e. Note that the goal of the trial implies that the data generating mechanism, i.e. the choice of θ_d and π_d made by $\mathcal{P}_d$, reflects clear efficacy of the drug, regardless of π_e and π_o. □

Example 2 *(Subjective vs default testing)* As a second example, consider the one-sided testing problem for a one-dimensional parameter θ of a model, that is: $H_1 : \theta \leq \theta_t$ vs. $H_2 : \theta > \theta_t$ and assume that prior information is represented by a subjective proper density π_e. It is well known that, under standard generalized 0–1 loss functions (see, for instance, Bernardo (1997) and also Sect. 3), optimal decisions are based on the Bayes factor B_{21}^e, i.e. the ratio between posterior and prior odds of the two hypotheses. We might be interested in evaluating the additional loss we incurr by using the Bayes factor, B_{21}^o say, computed with a noninformative prior π_o instead of the subjective Bayes factor. In this case, $\mathcal{E}_e$ and $\mathcal{E}_o$ represent the same decision maker who wants to quantify the loss (under π_e) of neglecting information from her/his subjective prior in favor of a more objective testing method and to know the number of observations needed to have high chances of making the same choice between H_1 and H_2 regardless of the prior used. Also in this case, the distribution used for pre-posterior calculations has to formalize a given scenario (for

instance: true null or true alternative) and does not have to be necessarily in accordance with π_e. $\square$

In both the situations sketched above we expect that, at least for relatively small sample sizes, the initial attitudes of the two actors 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 priors and test statistics is relevant. However, we also expect that, 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 problem of using a non optimal action a_o, which means implicitly adopting the inappropriate prior π_o, has been previously considered by DeGroot (1970), who did not discuss preposterior analysis. Statistical decision problems and predictive analysis under several actors have been dealt with, for instance, by Kadane and Seidenfeld (1989), and Lindley and Singpurwalla (1991). Similar scenarios are explored in Etzioni and Kadane (1993), Spiegelhalter and Freedman (1988) and Kadane (1990) 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. (2014a) in a context not formalized as a decision problem. More recently, De Santis and Gubbiotti (2017) have approached the problem from a formal decision theoretic perspective for point estimation.

The article provides essentially two contributions. First, it extends the result of the point estimation setting in De Santis and Gubbiotti (2017) 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 if a given sample size is appropriate or not.

The article relies widely on the literature on Bayesian sample size determination (SSD). From a decision-theoretic point of view, see, for instance: Raiffa and Schlaifer (1961), Berger (1985), Bernardo (1997), Pham-Gia (1997), Lindley (1997) and Parmigiani and Inoue (2009). For non-decision theoretic methods (i.e. *performance-based* approaches) see, among others: Spiegelhalter and Freedman (1986), Adcock (1997), Joseph and Belisle (1997), Joseph et al. (1997), Joseph and Wolfson (1997), Spiegelhalter et al. (2004) and Weiss (1997). The distinction between π_e and π_d is the so-called *two-priors approach* to Bayesian SSD, that we discuss in Sect. 2. See Tsutakawa (1972), O'Hagan and Stevens (2001), Wang and Gelfand (2002), De Santis (2006), Sahu and Smith (2006), M'Lan et al. (2006), Sambucini (2010), Brutti et al. (2014b) and Cellamare and Sambucini (2015). 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), Jackson and Novick (1980), Weerahandi and Zidek (1981).

The outline of the article is as follows. In Sect. 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 Sect. 3 the methodology is

developed for a generic testing problem for a real-valued parameter, and then it is specialized to one-sided testing (Sect. 3.1) and two-sided testing (Sect. 3.2). Results are provided for testing the normal mean (Sect. 4) and illustrated through numerical examples (Sect. 4.3). Finally, Sect. 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. For simplicity let us assume that θ is a continuous r.v. 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

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

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

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

It is straightforward to see that: (i) $A_{o,e}(\boldsymbol{x_n}) \geq 0$ for any choice of π_o and π_e; (ii) if $\pi_o = \pi_e$ then $A_{o,e}(\boldsymbol{x_n}) = 0$. Note also that, in general, $A_{o,e}(\boldsymbol{x_n}) \neq A_{e,o}(\boldsymbol{x_n})$. This reflects the asymmetry of the decision problem in which the (priviledged) actor $\mathcal{E}_e$ evaluates the additional posterior expected loss due to the (forced) use of the non optimal action a_o.

A relative measure of additional loss 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, $\bar{A}_{o,e}(X_n)$ is a random variable. We assume that, as n increases, $\bar{A}_{o,e}(X_n)$ 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:

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

where

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

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

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

where π_d is the *design prior*. Therefore the optimal sample size $n^\star$ depends on three priors (π_d, π_e, π_o). If π_d coincides with π_e, we retrieve the approach of Etzioni and Kadane (1993). However, in the most general case, π_d is different from both π_e and π_o. The design prior π_d describes a scenario in which the planner $\mathcal{P}_d$ wants to select the sample size. As mentioned in Example 1 and 2 of Sect. 1, the design scenario is not necessarily in agreement with the prior information expressed by π_e (or π_o). π_d is typically introduced to avoid local optimality of the resulting sample size as it happens in classical SSD methods. In fact, if π_d is a point-mass prior on a design value θ_d, then m_d reduces to the sampling distribution $f(\cdot|\theta_d)$, yielding optimal sample sizes that depend on θ_d (*conditional* Bayes approach). The design prior is hence used in a *what-if spirit* [see Wang and Gelfand (2002)], i.e. as an instrumental tool to incorporate uncertainty on θ_d, to avoid local optimality. See also Brutti et al. (2014b) for discussion on the use of multiple priors in SSD.

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

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, $j = o, e$, and assuming that $\mathbb{P}_{\pi_j}(\Theta_1) > 0$ and $\mathbb{P}_{\pi_j}(\Theta_2) > 0$, the Bayes factor for testing H_2 vs. H_1 is

$$B_{21}^j(x_n) = \frac{\omega_{21}^j(x_n)}{\omega_{21}^j}, \tag{3}$$

where

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

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

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

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

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

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

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

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

$$\bar{A}_{o,e}(\boldsymbol{x_n}) = \begin{cases} 0 & \text{if} & a_o(\boldsymbol{x_n}) = a_e(\boldsymbol{x_n}) \\ \xi_e(\boldsymbol{x_n}) & \text{if} & a_o(\boldsymbol{x_n}) \neq a_e(\boldsymbol{x_n}) \end{cases} = \xi_e(\boldsymbol{x_n}) 1_{\mathcal{Z}_{o,e}(\boldsymbol{x_n})} \quad (4)$$

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

$$\xi_e(\boldsymbol{x_n}) = \frac{\rho_e(\boldsymbol{x_n}, a_o) - \rho_e(\boldsymbol{x_n}, a_e)}{\rho_e(\boldsymbol{x_n}, a_o)}$$

$$= 1 - \frac{\min\{b_1 \mathbb{P}_{\pi_e}[\Theta_1|\boldsymbol{x_n}], b_2 \mathbb{P}_{\pi_e}[\Theta_2|\boldsymbol{x_n}]\}}{\max\{b_1 \mathbb{P}_{\pi_e}[\Theta_1|\boldsymbol{x_n}], b_2 \mathbb{P}_{\pi_e}[\Theta_2|\boldsymbol{x_n}]\}} \tag{5}$$

$$= 1 - \min\left\{\frac{k^e}{B_{21}^e(\boldsymbol{x_n})}, \frac{B_{21}^e(\boldsymbol{x_n})}{k^e}\right\}, \quad k_j = \frac{b_1}{b_2} \, \omega_{12}^j,$$

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

$$\mathcal{Z}_{o,e} = \left(\mathcal{Z}_o^{(1)} \cap \mathcal{Z}_e^{(2)}\right) \cup \left(\mathcal{Z}_o^{(2)} \cap \mathcal{Z}_e^{(1)}\right), \tag{6}$$

where

$$\mathcal{Z}_j^{(1)} = \left\{\boldsymbol{x_n} \in \mathcal{Z} : \omega_{21}^j(\boldsymbol{x_n}) < \frac{b_1}{b_2}\right\} = \{\boldsymbol{x_n} \in \mathcal{Z} : B_{21}^j(\boldsymbol{x_n}) < k_j\} \tag{7}$$

and

$$\mathcal{Z}_j^{(2)} = \left\{\boldsymbol{x_n} \in \mathcal{Z} : \omega_{21}^j(\boldsymbol{x_n}) > \frac{b_1}{b_2}\right\} = \{\boldsymbol{x_n} \in \mathcal{Z} : B_{21}^j(\boldsymbol{x_n}) > k_j\}. \tag{8}$$

Finally, from (2) and (4) 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, $q_\epsilon^j(\boldsymbol{x_n})$, for $j = o, e$. First, from Eq. (5) we have

$$\xi_e(\boldsymbol{x_n}) = 1 - \min \left\{ \frac{b_1}{b_2} \frac{F_e(\theta_t|\boldsymbol{x_n})}{1 - F_e(\theta_t|\boldsymbol{x_n})}, \frac{b_2}{b_1} \frac{1 - F_e(\theta_t|\boldsymbol{x_n})}{F_e(\theta_t|\boldsymbol{x_n})} \right\}. \tag{9}$$

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

$$\mathcal{Z}_j^{(1)} = \left\{ \boldsymbol{x_n} \in \mathcal{Z} : \frac{1 - F_j(\theta_t|\boldsymbol{x_n})}{F_j(\theta_t|\boldsymbol{x_n})} < \frac{b_1}{b_2} \right\} = \left\{ \boldsymbol{x_n} \in \mathcal{Z} : \theta_t > q_\epsilon^j(\boldsymbol{x_n}) \right\} \tag{10}$$

and

$$\mathcal{Z}_j^{(2)} = \left\{ \boldsymbol{x_n} \in \mathcal{Z} : \frac{1 - F_j(\theta_t|\boldsymbol{x_n})}{F_j(\theta_t|\boldsymbol{x_n})} > \frac{b_1}{b_2} \right\} = \left\{ \boldsymbol{x_n} \in \mathcal{Z} : \theta_t < q_\epsilon^j(\boldsymbol{x_n}) \right\}, \tag{11}$$

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

$$q_\epsilon^m(\boldsymbol{x_n}) = \min \left\{ q_\epsilon^e(\boldsymbol{x_n}), q_\epsilon^o(\boldsymbol{x_n}) \right\} \quad \text{and} \quad q_\epsilon^M(\boldsymbol{x_n}) = \min \left\{ q_\epsilon^e(\boldsymbol{x_n}), q_\epsilon^o(\boldsymbol{x_n}) \right\} \tag{12}$$

and, recalling Eq. (6),

$$\mathcal{Z}_{o,e} = \left\{ \boldsymbol{x_n} \in \mathcal{Z} : q_\epsilon^m(\boldsymbol{x_n}) < \theta_t < q_\epsilon^M(\boldsymbol{x_n}) \right\}. \tag{13}$$

Let $F_j(\cdot)$, $j = o, e$ be the c.d.f.'s of a location-scale model. Let $\mathbb{F}(\cdot)$ 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

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

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 $\mathbb{F}(\cdot)$, $\bar{q}_{\epsilon}$ and $W_j(x_n)$. First, from Eq. (5) we have

$$\xi_e(x_n) = 1 - \min\left\{\frac{b_1}{b_2}\frac{\mathbb{F}(-W_e(x_n))}{1 - \mathbb{F}(-W_e(x_n))}, \frac{b_2}{b_1}\frac{1 - \mathbb{F}(-W_e(x_n))}{\mathbb{F}(-W_e(x_n))}\right\}, \quad (14)$$

which, in case of symmetry, is

$$\xi_e(x_n) = 1 - \min\left\{\frac{b_1}{b_2}\frac{1 - \mathbb{F}(W_e(x_n))}{\mathbb{F}(W_e(x_n))}, \frac{b_2}{b_1}\frac{\mathbb{F}(W_e(x_n))}{1 - \mathbb{F}(W_e(x_n))}\right\}. \quad (15)$$

Then, from (10), it follows that

$$\mathcal{Z}_j^{(1)} = \left\{x_n \in \mathcal{Z} : \mathbb{F}\left(\frac{\theta_t - \mu_j(x_n)}{\sigma_j(x_n)}\right) > \frac{b_2}{b_1 + b_2}\right\}$$

$$= \left\{x_n \in \mathcal{Z} : W_j(x_n) = \frac{\mu_j(x_n) - \theta_t}{\sigma_j(x_n)} < \bar{q}_{1-\epsilon}\right\},$$

and, similarly,

$$\mathcal{Z}_j^{(2)} = \left\{x_n \in \mathcal{Z} : W_j(x_n) = \frac{\mu_j(x_n) - \theta_t}{\sigma_j(x_n)} > \bar{q}_{1-\epsilon}\right\}.$$

Therefore,

$$\mathcal{Z}_o^{(1)} \cap \mathcal{Z}_e^{(1)} = \{x_n \in \mathcal{Z} : W_M(x_n) < \bar{q}_{1-\epsilon}\}$$

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

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

$$\mathcal{Z}_{o,e} = \{x_n \in \mathcal{Z} : W_m(x_n) < \bar{q}_{1-\epsilon} < W_M(x_n)\}, \quad (16)$$

where

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

3.2 Two-sided testing

The general results of Sect. 3 are now adapted to the specific set-up of two-sided testing, i.e. $H_1 : \theta \in [\theta_L, \theta_U]$ vs. $H_2 : \theta \notin [\theta_L, \theta_U]$. In this case $\Theta_1 = [\theta_L, \theta_U]$ and $\Theta_2 = (-\infty, \theta_L) \cup (\theta_U, +\infty)$. Hence the Bayes factor is given by Eq. (3), where the prior and the posterior odds are

$$\omega_{21}^j = \frac{1 - (F_e(\theta_U) - F_e(\theta_L))}{F_e(\theta_U) - F_e(\theta_L)}, \quad (17)$$

and

$$\omega_{21}^j(\boldsymbol{x_n}) = \frac{1 - (F_e(\theta_U|\boldsymbol{x_n}) - F_e(\theta_L|\boldsymbol{x_n}))}{F_e(\theta_U|\boldsymbol{x_n}) - F_e(\theta_L|\boldsymbol{x_n})}, \tag{18}$$

respectively. Then, to obtain $\bar{A}_{o,e}$ it is convenient to use directly the general expressions of $\xi_e(\boldsymbol{x_n})$, $\mathcal{Z}_j^{(1)}$ and $\mathcal{Z}_j^{(2)}$, given by Eqs. (5), (7) and (8) respectively.

4 Results for the normal mean

Let us now assume that $X_1, X_2, \ldots, X_n$ is a random sample, $X_i|\theta \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$ and that $\pi_j(\cdot)$ are conjugate priors, i.e. $\theta|\sigma^2 \sim N(\mu_j, \sigma^2/n_j)$, $j = o, e$.

4.1 One-sided tests

Known variance. First, assume that σ^2 is known. Then the posterior distribution of θ is Normal with location and scale

$$\mu_j(\boldsymbol{x_n}) = \frac{n_j\mu_j + n\bar{x}_n}{n_j + n} \quad \text{and} \quad \sigma_j(\boldsymbol{x_n}) = \frac{\sigma}{\sqrt{n_j + n}}. \tag{19}$$

From Eq. (15), 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 Eq. (16) 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 $\bar{x}_n$. 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

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

and

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

Note that, as n increases, the contributions of the priors $\pi_j(\theta)$ in the corresponding posteriors tend 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)),

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

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

$$\left(\mu_j,\, n_j\alpha_j\beta_j^{-1},\, 2\alpha_j\right), \tag{20}$$

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

$$\begin{cases} \mu_j(\boldsymbol{x_n}) = (n_j\mu_j + n\bar{x}_n)(n_j + n)^{-1}, \\ \sigma_j(\boldsymbol{x_n}) = \left[\beta_j + \frac{1}{2}nS_n^2 + \frac{1}{2}(n_j+n)^{-1}n_jn(\mu_j - \bar{x}_n)^2\right](n_j+n)^{-1}\left(\alpha_j + \frac{n}{2}\right)^{-1}, \\ \nu_j = 2\alpha_j + n \end{cases} \tag{21}$$

where S_n^2 is the sample variance. 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 Eqs. (9), (12) and (13) by setting $F_j(\cdot)$ 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, $j = o, e$.

4.2 Two-sided tests

Under the same assumptions of Sect. 4.1, the prior and the posterior odds are given by Eqs. (17) and (18), where $F_e(\cdot)$ and $F_e(\cdot|\boldsymbol{x_n})$ are

- Normal c.d.f.'s of parameters $(\mu_j, \sigma^2/n_j)$ and $(\mu_j(\boldsymbol{x_n}), \sigma_j^2(\boldsymbol{x_n}))$ respectively, when σ^2 is assumed to be known;
- Student t c.d.f.'s of parameters given by Eqs. (20) and (21), respectively, in the unknown variance case.

4.3 Numerical examples

In this section we illustrate some numerical examples related to the Normal case for both one-sided and two-sided testing. Without loss of generality we set $b_1 = b_2 = 1$.

One-sided testing. 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. First of all in Fig. 1 we show through numerical examples the relationship that holds in general between e_n and p_n, that is $e_n \le p_n$ for each value of n (see final remark of Sect. 3).

In the following we show 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. Moreover, we consider both the known variance case, taking $\sigma^2 = 1$, and the unknown variance case, where the hyperprior parameters for σ^2 are chosen to satisfy $\mathbb{E}(\sigma^2) = 1$, i.e. $\alpha_j = \beta_j + 1$, for $j = o, e, d$. In this way we obtain results comparable to those of the known variance case and we show the mild impact of the prior assumptions on σ^2
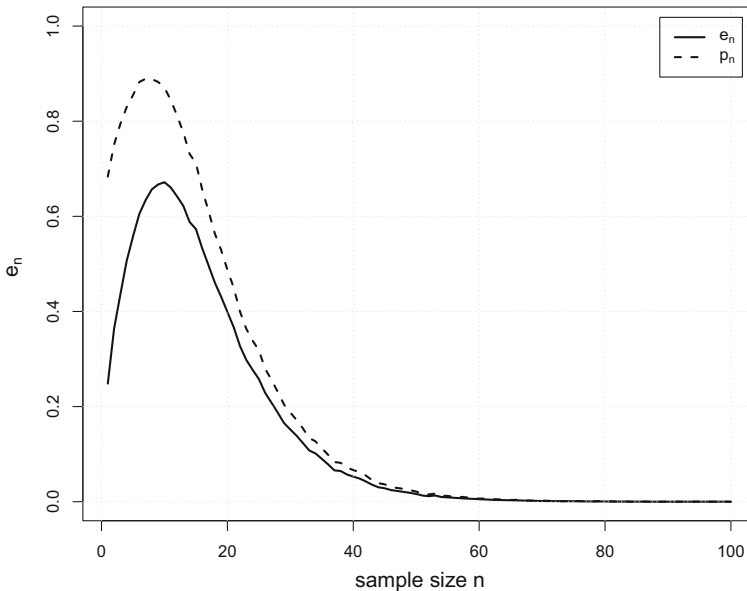
Fig. 1 Behavior of e_n and p_n for increasing values of n, given $\theta_t = 1$, $\sigma = 1$, $\mu_o = 0$, $\mu_e = 1$, $n_o = n_e = 10$, $\mu_d = 1.5$, $n_d = 10$.

on e_n and consequently on the optimal sample size $n^\star$. For a general discussion on the relationships between optimal designs for the normal model with known and unknown variance see Verdinelli (2000).

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 Fig. 2, 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. Under the unknown variance assumption (see panel (b) of Fig. 2) we notice a similar behavior of e_n with a slight inflation of each curve, uniformly with respect to n.

In the second set-up the conflict between π_e and π_o is even emphasized, π_e supports the alternative hypothesis H_2 and we take μ_e larger than μ_d (i.e. $\mu_o = 0$ and $\mu_e = 2$). In Fig. 3, 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. By comparing panel (a) and panel (b) we notice that, for each choice of $n_o = n_e$, e_n is substantially unchanged.

To sum up, Table 1 displays the optimal sample sizes, obtained using criterion (1) for a threshold γ as small as 0.01. Negligible differences are observed between the two cases, known and unknown variance.

(a) Known variance
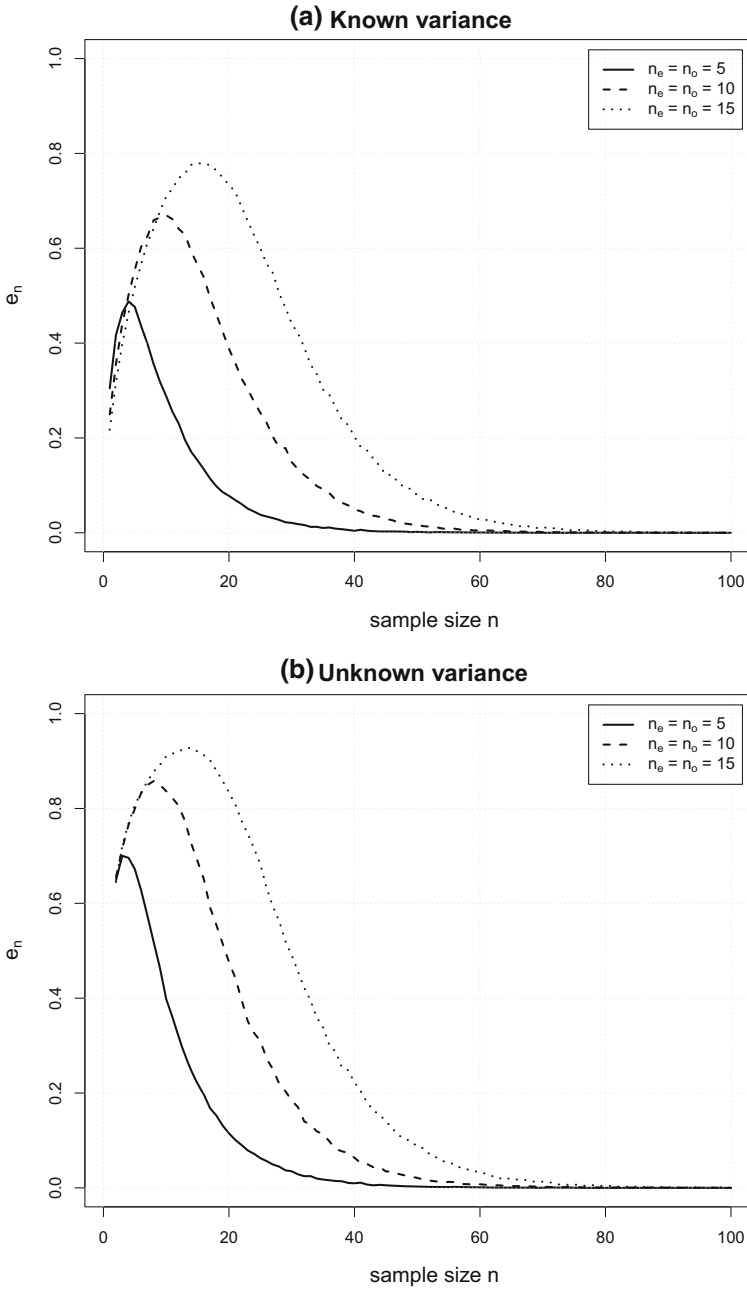
(b) Unknown variance

Fig. 2 One-sided testing: 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 $\theta_t = 1$, $\mu_d = 1.5$, $n_d = 10$, $\mu_o = 0$. **a** Known variance case: $\sigma^2 = 1$; **b** Unknown variance case: $\alpha_e = \alpha_o = 2$, $\beta_e = \beta_o = 1$ and $\alpha_d = 3$, $\beta_d = 2$

(a) Known variance
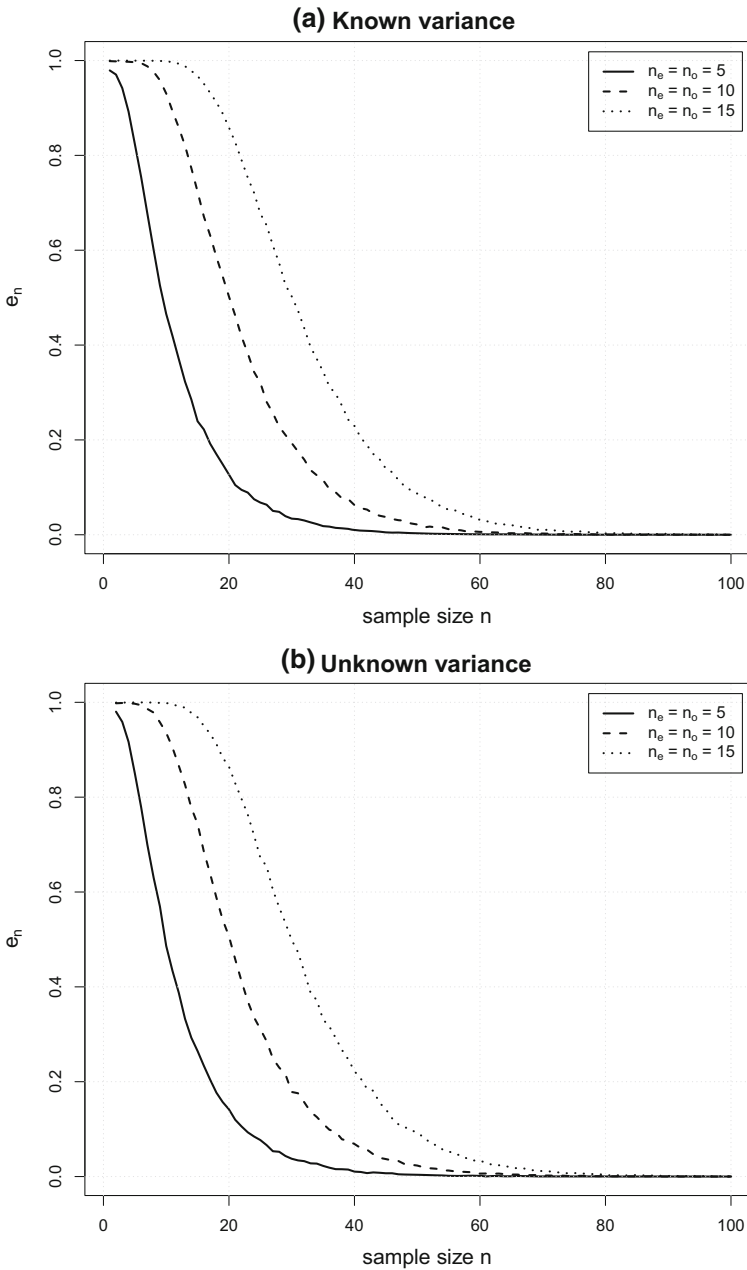
(b) Unknown variance

Fig. 3 One-sided testing: 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 $\theta_t = 1$, $\mu_d = 1.5$, $n_d = 10$, $\mu_o = 0$ **a** Known variance case: $\sigma^2 = 1$; **b** Unknown variance case: $\alpha_e = \alpha_o = 2$, $\beta_e = \beta_o = 1$ and $\alpha_d = 3$, $\beta_d = 2$

Table 1 One-sided testing: Optimal sample sizes $n^\star$ based on criterion (1) with $\gamma = 0.01$, for different values of μ_e, n_e and n_o, given $\theta_t = 1$, $\mu_d = 1.5$, $n_d = 10$, $\mu_o = 0$. (a) Known variance case: $\sigma^2 = 1$; (b) Unknown variance case: $\alpha_e = \alpha_o = 2$, $\beta_e = \beta_o = 1$ and $\alpha_d = 3$, $\beta_d = 2$

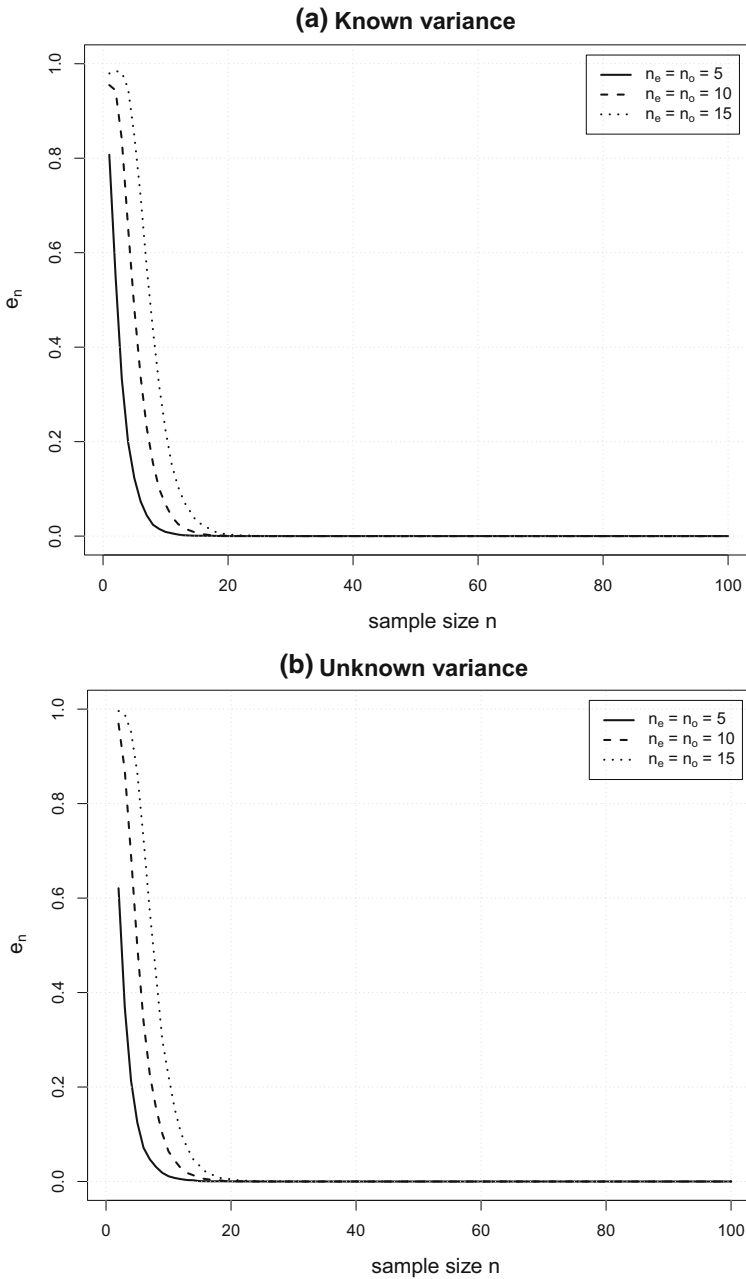| (a) Known variance			
	$n_e = n_o = 5$	$n_e = n_o = 10$	$n_e = n_o = 15$
$\mu_e = 1$	36	53	71
$\mu_e = 2$	40	56	71
(b) Unknown variance			
	$n_e = n_o = 5$	$n_e = n_o = 10$	$n_e = n_o = 15$
$\mu_e = 1$	41	56	71
$\mu_e = 2$	40	57	71

Two-sided testing Consider the interval null hypothesis $H_1 : \theta \in [\theta_L, \theta_U] = [-0.5, 0.5]$ vs $H_2 : \theta \notin [\theta_L, \theta_U]$. Let us assume that π_o is centred on the interval $[\theta_L, \theta_U]$, i.e. $\mu_o = 0$, whereas π_e and π_d both support the alternative hypothesis. More specifically we take into account two different scenarios:

(i)   in Fig. 4 $\mu_e = 1 < \mu_d = 1.5$ and the predictive distribution produces data strongly favoring large (positive) values of θ; thus, the conflict between the two competing priors rapidly reduces and e_n flattens to 0 even for small sample sizes;

(ii) conversely, in Fig. 5 $\mu_d = 1 < \mu_e = 1.5$, i.e. the prior opinion expressed by π_e is even more extreme than the design scenario depicted by π_d, and therefore the decreasing behavior of e_n appears slower than in the previous case.

For the sake of completeness, we finally report in Table 2 the optimal sample sizes obtained for a threshold $\gamma = 0.01$ in the above mentioned settings (i) and (ii). As in the one-sided testing set-up similar considerations arise from the comparison between the known variance case and the unknown variance case: again, the values of $n^\star$ are almost unaffected by the prior assumption on σ^2.

5 Conclusions

Statistical decision theory offers the ideal formal set-up to define sample size criteria, that can be based on the risks of 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 De Santis and Gubbiotti (2017)). Attention is here specifically devoted to the testing problem. We propose a sample size criterion built up on the predictive

(a) Known variance

(b) Unknown variance

Fig. 4 Two-sided testing Expected value of the relative additional loss as a function of the sample size n, with $\mu_o = 0$, $\mu_e = 1$ for different values of n_e and n_o, given $[\theta_L, \theta_U] = [-0.5, 0.5]$, $\mu_d = 1.5$, $n_d = 10$, **a** Known variance case: $\sigma = 1$; **b** Unknown variance case: $\alpha_e = \alpha_o = 2$, $\beta_e = \beta_o = 1$ and $\alpha_d = 3$, $\beta_d = 2$
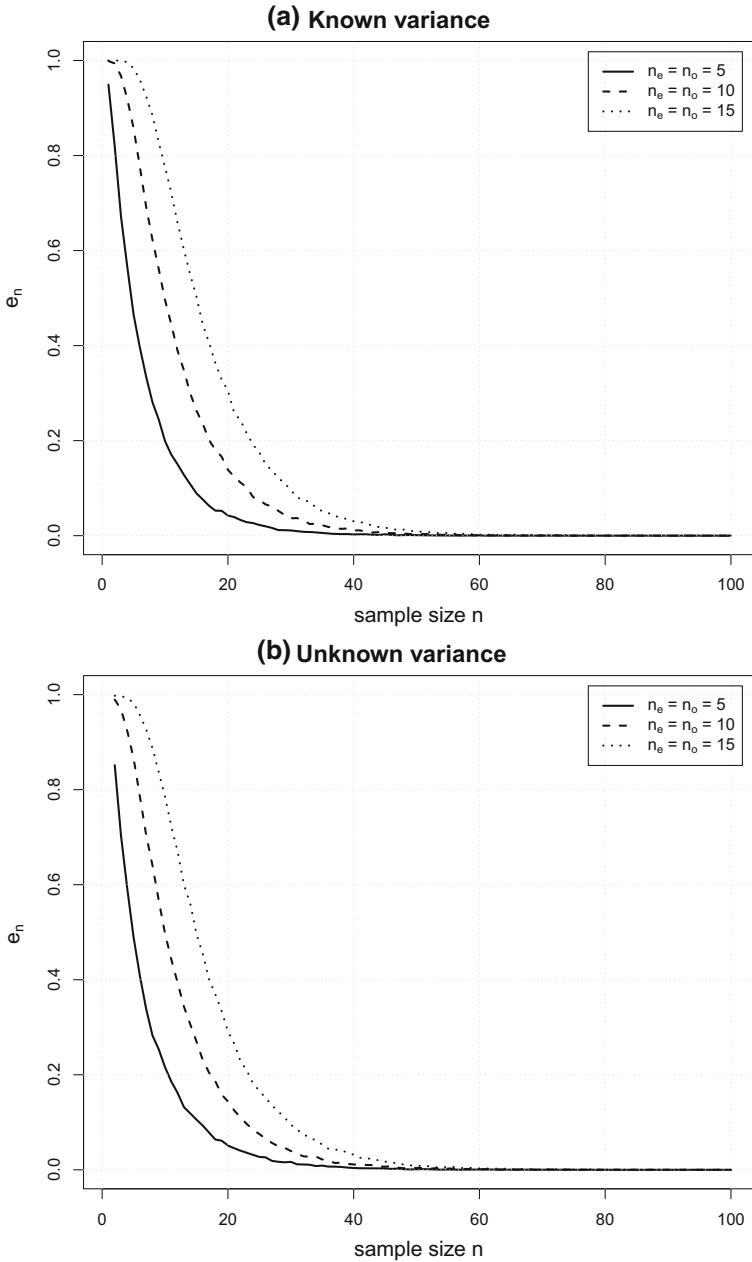
Fig. 5 Two-sided testing: Expected value of the relative additional loss as a function of the sample size n, with $\mu_o = 0$, $\mu_e = 1.5$ for different values of n_e and n_o, given $[\theta_L, \theta_U] = [-0.5, 0.5]$, $\mu_d = 1$, $n_d = 10$, **a** Known variance case: $\sigma = 1$; **b** Unknown variance case: $\alpha_e = \alpha_o = 2$, $\beta_e = \beta_o = 1$ and $\alpha_d = 3$, $\beta_d = 2$

Table 2 Two-sided testing: Optimal sample sizes $n^\star$ based on criterion (1) with $\gamma = 0.01$, for different values of μ_e and $\mu_d = 1.5$, n_e and n_o, given $[\theta_L, \theta_U] = [-0.5, 0.5]$, $\mu_o = 0$, $n_d = 10$. a Known variance case: $\sigma^2 = 1$; b Unknown variance case: $\alpha_e = \alpha_o = 2$, $\beta_e = \beta_o = 1$ and $\alpha_d = 3$, $\beta_d = 2$

(a) Known variance			
	$n_e = n_o = 5$	$n_e = n_o = 10$	$n_e = n_o = 15$
$\mu_e = 1 < \mu_d = 1.5$	9	14	17
$\mu_e = 1.5 < \mu_d = 1$	30	41	48
(b) Unknown variance			
	$n_e = n_o = 5$	$n_e = n_o = 10$	$n_e = n_o = 15$
$\mu_e = 1 < \mu_d = 1.5$	10	14	17
$\mu_e = 1.5 < \mu_d = 1$	33	42	48

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 Sect. 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 the settings of one-sided and two-sided testing. In Sect. 4.3 we have considered some numerical examples in the context of testing the normal mean (known and unknown variance). Even in this basic framework, explicit expression of e_n are not available but standard Monte Carlo approximations can be easily obtained.

Here is a non exhaustive list of some points of discussion and possible further developments.

1. *Connection with group decision making* Despite the presence of multiple parties, our framework is characterized by some specific features that make it distinct from Group decision-making methods (see French and Rios Insua (2000)). In fact, in Group decision-making many individuals make a collective choice among several alternatives and the decision is the result of a group consensus. Conversely, our approach explicitly states the asymmetry between the roles of the two actors ($\mathcal{E}_e$ and $\mathcal{E}_o$) involved in the decision problem and focuses on the evaluation of a decision made using an inappropriate prior.

2. *The use of inappropriate priors* The "consequences of using an inappropriate distribution of the parameter" is a problem that has been addressed for instance by DeGroot (1970). The Author provided a geometric representation of the difference between the risk associated to the inappropriate distribution, $\rho_e(\boldsymbol{x_n}, a_o)$, and the risk attained using the optimal decision, $\rho_e(\boldsymbol{x_n}, a_e)$, when the number of decisions is finite (non-strictly concave Bayes risk) or infinite (strictly concave Bayes risk). Our article can be regarded as a predictive analysis of this problem in the testing framework where only two decisions and

only two prior probability distributions are available. Furthermore, it would be interesting to follow the suggestion of one of the two reviewers that is to consider inappropriate priors π_o that represent extreme cases such as a least favourable prior associated with minimax decision rules or, conversely, a degenerate prior.

3. *Point-null hypothesis* In Sect. 3.2 we have considered a two-sided test with an imprecise null hypothesis. This (apparently restrictive) choice is motivated by the well known sensitivity of Bayesian testing procedures on prior distributions when a point-null hypothesis test is taken into account. However, since this kind of test can be regarded as a limiting case of the two-sided imprecise null test [see Berger and Delampady (1987), Delampady (1989) and Verdinelli (1996)], in principle, it can be addressed as a special case of the method proposed for a general testing setup in Sect. 3. We hope to elaborate on this aspect in the future.

4. *More complex models* 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.

5. *Robustness* Instead of considering only one prior π_e, we could extend our approach by considering an entire class of prior distributions Γ. See, among others, Berger (1984) and Berger (1985). 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). In future research we plan to explore more deeply the properties of e_n as a measure of dissimilarity between probability distributions. This suggestion comes from the comments of one of the two anonymous referees that we would like to thank for her/his careful and detailed review which strongly contributed to improve the final version of this paper.

References

Adcock CJ (1997) Sample size determination: a review. J R Stat Soc Ser D (The Stat) 46(2):261–283

Berger JO (1984) The robust Bayesian viewpoint (with Discussion). In: Kadane J (ed) Robustness of Bayesian analysis. North-Holland, Amsterdam

Berger JO (1985) Statistical decision theory and bayesian analysis. Springer, Berlin

Berger JO, Delampady M (1987) Testing precise hypotheses. Stat Sci 2:317–352

Bernardo JM (1997) Statistical inference as a decision problem: the choice of the sample size. Statistician 46:151–153

Bernardo JM, Smith AFM (1994) Bayesian theory. Wiley, Chichester

Brutti P, De Santis F, Gubbiotti S (2014) Predictive measures of the conflict between frequentist and Bayesian estimators. J Stat Plan Inference 148:111–122

Brutti P, De Santis F, Gubbiotti S (2014) Bayesian frequentist sample size determination: a game of two priors. Metron 72(2):133–151

Burt J (1990) Towards agreement: Bayesian experimental design by Jameson Burt, Purdue University, Technical Report #90-41, Department of Statistics

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

DeGroot MH (1970) Optimal statistical decisions. Mc-Graw Hill, New York

Delampady M (1989) Lower bounds on Bayes factors for interval null hypotheses. J Am Stat Assoc 84:120–124

De Santis F (2006) Sample size determination for robust Bayesian analysis. J Am Stat Assoc 101(473):278–291

De Santis F, Gubbiotti S (2017) A decision-theoretic approach to sample size determination under several priors. Appl Stoch Models Bus Ind 33(3):282–295

Etzioni R, Kadane JB (1993) Optimal experimental design for anothers analysis. J Am Stat Assoc 88(424):1404–1411

French S, Rios Insua D (2000) Statistical decision theory. Kendall's library of statistics. Wiley, Hoboken

Jackson PH, Novick MR, DeKeyrel DF (1980) Adversary preposterior analysis for simple parametric models in Bayesian Analysis in Econometrics and Statistics. Arnold Zellner, Amsterdam, pp 113–132

Joseph L, Belisle P (1997) Bayesian sample size determination for normal means and difference between normal means. Statistician 46:209–226

Joseph L, du Berger R, Belisle P (1997) Bayesian and mixed Bayesian/likelihood criteria for sample size determination. Stat Med 16:769–781

Joseph L, Wolfson D (1997) Interval-based versus decision theoretic criteria for the choice of sample size. Statistician 46:145–149

Kadane JB (1990) A statistical analysis of adverse impact of employer decisions. J Am Stat Assoc 85(412):925–933

Kadane JB, Seidenfeld T (1989) Randomization in a Bayesian perspective. J Stat Plan Inference 25:329–345

Lindley DV (1997) The choice of sample size. The Stat 46:129–138

Lindley DV, Singpurwalla N (1991) On the evidence needed to reach agreed action between adversaries, with application to acceptance Sampling. J Am Stat Assoc 86(416):933–937

M'Lan CE, Joseph L, Wolfson DB (2006) Bayesian sample size determination for case-control studies. J Am Stat Assoc 101(474):760–772

O'Hagan A, Stevens JW (2001) Bayesian assessment of sample size for clinical trials for cost effectiveness. Med Decis Making 21:219–230

Pham-Gia T (1997) On Bayesian analysis, Bayesian decision theory and the sample size problem. Statistician 46:139–144

Raiffa H, Schlaifer R (1961) Applied statistical decision theory. Clinton Press Inc, Boston

Sahu SK, Smith TMF (2006) A Bayesian method of sample size determination with practical applications. J R Stat Soc A Stat 169:235–253

Parmigiani G, Inoue L (2009) Decision theory: principles and approaches. Wiley series in probability and statistics. Wiley, Hoboken

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

Spiegelhalter DJ, Freedman LS (1986) A predictive approach to selecting the size of a clinical trial, based on subjective clinical opinion. Stat Med 5:1–13

Spiegelhalter DJ, Freedman LS (1988) Bayesian approaches to clinical trials (with discussion). In: Bernardo JM, DeGroot MH, Lindley DV, Smith AFM (eds) Bayesian statistics, vol 3. Oxford University Press, Oxford, pp 453–477

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

Tsutakawa RK (1972) Design of experiment for bioassay. J Am Stat Assoc 67(339):585–590

Verdinelli I (2000) A note on Bayesian design for the normal linear model with unknown error variance. Biometrika 87(1):222–227

Verdinelli I, Wasserman L (1996) Bayes Factors, nuisance parameters and imprecise tests. Bayesian Stat 5:765–771

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

Weerahandi S, Zidek JV (1981) Multi-bayesian statistical decision theory. J R Stat Soc Ser A General 144(1):85–93

Weiss RE (1997) Bayesian sample size computations for hypothesis testing. Statistician 46:185–191