A decision-theoretic approach to sample size determination under several priors

Fulvio De Santis - Stefania Gubbiotti

Sapienza University of Rome

Abstract

In this article we consider sample size determination (SSD) for experiments in which estimation and design are performed by multiple parties. This problem has relevant applications in contexts involving adversarial decision makers, such as control theory, marketing and drug testing. Specifically, we adopt a decision-theoretic perspective and we assume that a decision on an unknown parameter of a statistical model involves two actors, $\mathcal{E}_e$ and $\mathcal{E}_o$, who share the same data and loss function but not the same prior beliefs on the parameter. We also suppose that $\mathcal{E}_e$ has to use $\mathcal{E}_o$'s optimal action and we finally assume that the experiment is planned by a third party, $\mathcal{P}_d$. In this framework we aim at determining an appropriate sample size so that the posterior expected loss incurred by $\mathcal{E}_e$ in taking the optimal action of $\mathcal{E}_o$ is sufficiently small. We develop general results for the one-parameter exponential family under quadratic loss and analyze the interactive impact of the prior beliefs of the three different parties on the resulting sample sizes. Relationships with other SSD criteria are explored.

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

1 Introduction

Let us consider a decision problem on an unknown parameter of a statistical model and assume that $\mathcal{E}_e$ and $\mathcal{E}_o$ are two decision makers, who share the same data and loss function but not the same prior information/opinions on the parameter. Assume that π_e and π_o are the prior distributions elicited by $\mathcal{E}_e$ and $\mathcal{E}_o$ and let a_e and a_o be their optimal decisions, i.e. the actions that minimize the two posterior expected losses. Suppose also that, for some reasons, the decision maker $\mathcal{E}_e$ has to use the action a_o. $\mathcal{E}_e$ measures the performance of a_o using the posterior expected loss under π_e, which is 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 performs preposterior sample size calculations using a predictive distribution of the data based on a prior π_d that is, in general, different from both π_o and π_e. The goal of the present article is the determination of an appropriate sample size so that the loss incurred by $\mathcal{E}_e$ in taking the action a_o instead of a_e is sufficiently small.

To motivate this decisional framework, consider an industrial (pharmacological) experiment whose goal is to show that the effect of a drug (the location parameter θ of a model) is larger than a given threshold. The experiment involves three parties: a very optimistic planner $\mathcal{P}_d$, a moderately optimistic data-analyst (or decision maker) $\mathcal{E}_e$ and a skeptical final user $\mathcal{E}_o$. In order to convince $\mathcal{E}_o$, with skeptical prior π_o, $\mathcal{E}_e$ has to use a_o but s/he evaluates this action with the expected posterior loss based on her/his moderately enthusistic prior π_e. The optimism of $\mathcal{P}_d$ is formalized by π_d, used for preposterior calculations. The question is: how large the sample size must be so that the predictive expected value of the posterior expected loss of a_o, evaluated by $\mathcal{E}_e$, can be sufficiently small?

A special relevant case is obtained from the above set-up by assuming that π_e expresses the personal beliefs or information of $\mathcal{E}_e$ whereas π_o is a prior that leads to a conventional frequentist inferential procedure. For instance, in point estimation, π_o may be a noninformative prior distribution and a_o the usual maximum likelihood estimate (MLE) of θ. In this case, we look for the minimal sample size that guarantees - under the design scenario modeled by $\mathcal{P}_d$ - a sufficiently small loss incurred by the Bayesian decision maker $\mathcal{E}_e$ when the MLE a_o is used in the place of the Bayes estimator a_e.

The topic sketched above has been previously analyzed by several authors. In most of the articles we are aware of, however, the planner $\mathcal{P}_d$ coincides with $\mathcal{E}_e$ and preposterior evaluation of the performance of a_o - needed for sample size calculations - is made using the predictive distribution of the data induced by the prior distribution π_e. Kadane and Saidenfeld (1989) [17], for instance, consider the example of the proponent of a new medical treatment ($\mathcal{E}_e$) who has to design a trial to convince a skeptical expert or a regulatory agency ($\mathcal{E}_o$) that the new treatment is more effective than the standard. Similarly, in the context of quality control, Lindley and Singpurwalla (1991) [19] consider a manufacturer ($\mathcal{E}_e$) trying to sell a product to a consumer ($\mathcal{E}_o$) who accepts or rejects the product on the basis of the evidence produced by the manufacturer. Similar scenarios are considered, in Etzioni and Kadane (1993) [11], Spiegelhalter and Freedman (1988) [28], Kadane (1990)

[16].

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 recently considered by Brutti et al. (2014) [5] in a related context that, however, is not formalized as a decision problem. The authors consider point estimation, assume that π_o is a noninformative prior and measures the discrepancy between the MLE a_o and the Bayes estimator a_e in terms of their squared difference. Explicit evaluation of using a_o in the place of the Bayes action a_e is not however possible without a decisional framework. This aspect is explored here in Section 3.

The present article has a double goal. On the one hand, it is as an extension of the Etzioni-Kadane's approach, the generalization consisting in: (a) introducing a design prior π_d not necessarily coincident with the prior of the two decision makers; and (b) providing general results for the one-parameter exponential family with conjugate priors. On the other hand, it provides a decision-theoretic foundation to the analysis of the conflict between alternative inferential procedures considered in Brutti et al (2014) [5]. We limit the analysis to the point estimation problem with quadratic loss, which yields closed-form results, but it can be extended to other decision problems and loss functions as well as to other models and priors.

The central topic of the article is sample size determination (SSD) from a Bayesian perspective, whose literature has substantially increased in the last decades. From a decision-theoretic point of view, the main readings for this topics are Raiffa and Schlaifer (1961) [23], Berger (1985) [2], Bernardo (1997) [3], Pham-Gia (1997) [22], Lindley (1997) [18], Parmigiani and Inoue (2009) [25]. The literature on non-decision theoretic methods (the so-called *performance-based* approaches) is quite large, including, 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].

In the present article we support the idea of using multiple priors for SSD, and we extend the so-called *two-priors approach* in Bayesian SSD. As far as we know, the earliest reference releted to the use of two distinct priors for design and inference is Tsutukawa (1972) [30]. Recently this approach has been followed, among others, by O'Hagan and Stevens (2001) [21], Wang and Gelfand (2002) [31], De Santis (2006) [10], 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. For references 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. In Section 3 the focus is restricted to point estimation, using a quadratic loss. Explicit results are given for the exponential family with conjugate priors. We also comment on the relationships with predictive analysis of the conflict between alternative estimators of Brutti et al (2014) [5] and with the approach in Etzioni and Kadane (1993) [11]. As special cases we consider the normal and the exponential models. Numerical results are presented and discussed in Sections 3.1.1 and 3.1.2. Section 4 contains a final discussion.

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 $g(\theta)$, a function of interest of the parameter, $\mathcal{A}$ the action space and $L(a, g(\theta))$ 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

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

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

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

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

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

Small values of $\rho(\boldsymbol{x_n}, a_o; \pi_e)$ - close to $\rho(\boldsymbol{x_n}, a_e; \pi_e)$ - show that the non-optimal action a_o performs well even under the prior assumptions represented by π_e. Before observing the data, $\rho(\boldsymbol{X_n}, a_o; \pi_e)$ is a random variable. We are interested in selecting the smallest sample size such that its expected value is smaller than a selected threshold γ:

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

where

$$e_n = \mathbb{E}_{m_d}\big[\rho(\boldsymbol{X_n}, a_o; \pi_e)\big] \tag{2}$$

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].

3 Results for point estimation under quadratic loss

In this section we provide explicit results for point estimation of $g(\theta)$, where $\theta \in \Theta \subset \mathbb{R}$. Under the quadratic loss $L(a, g(\theta)) = (g(\theta) - a)^2$, it is easy to check that

$$\begin{aligned}\rho(\boldsymbol{x_n}, a_o; \pi_e) &= \mathbb{E}_{\pi_e}\big[L(a_o, g(\theta))|\boldsymbol{x_n}\big] \\ &= \mathbb{E}_{\pi_e}\big[(g(\theta) \pm a_e - a_o)^2|\boldsymbol{x_n}\big] \\ &= \rho(\boldsymbol{x_n}, a_e; \pi_e) + D_{e,o}(\boldsymbol{x_n}),\end{aligned} \tag{3}$$

where

$$\rho(\boldsymbol{x_n}, a_e; \pi_e) = \mathbb{V}_{\pi_e}\big[g(\theta)|\boldsymbol{x_n}\big]$$

and

$$D_{e,o}(\boldsymbol{x_n}) = \big(a_o - a_e\big)^2 = \big(\mathbb{E}_{\pi_e}\big[g(\theta)|\boldsymbol{x_n}\big] - \mathbb{E}_{\pi_o}\big[g(\theta)|\boldsymbol{x_n}\big]\big)^2.$$

Thus, as noted in Etzioni and Kadane (1993) [11], the posterior expected loss of a_o under π_e, is equal to the minimal expected loss under π_e plus the penalizing term $D_{e,o}(\boldsymbol{x_n})$, which measures the discrepancy between a_o and a_e. Note that if π_o is a noninformative prior distribution, the optimal action a_o is typically the frequentist MLE and $D_{e,o}(\boldsymbol{x_n})$ becomes the measure of the conflict between the MLE and the Bayesian estimate based on π_e, considered in Brutti et al (2014) [5]. From (2) and (3) it follows that

$$e_n = \mathbb{E}_{m_d}\big[\rho(\boldsymbol{X_n}, a_o; \pi_e)\big] = \mathbb{E}_{m_d}\big[\rho(\boldsymbol{X_n}, a_e; \pi_e)\big] + \mathbb{E}_{m_d}\big[D_{e,o}(\boldsymbol{X_n})\big]. \tag{4}$$

In the following section we consider the one-parameter exponential family and provide an explicit general expression for $D_{e,o}(\boldsymbol{x_n})$ and its predictive expected value. Conversely, the expression of $\mathbb{V}_{\pi_e}\big(g(\theta)|\boldsymbol{x_n}\big)$ is model-specific and is given in Sections 3.1.1 and 3.1.2 for the normal and the exponential models.

3.1 Canonical Exponential Family

Let θ be the natural parameter of the canonical exponential family with probability distribution

$$f(x|\theta) = h(x) \cdot \exp\{x\theta - b(\theta)\}. \tag{5}$$

Suppose that we are interested in $g(\theta) = \mathbb{E}_f\big[X\big] = b'(\theta)$. As prior distribution π_j for θ let us consider the natural conjugate prior distribution, that is

$$\pi_j(\theta) = c(n_j, \mu_j) \cdot \exp\{n_j\mu_j\theta - n_j b(\theta)\}, \qquad j = e, d, o \tag{6}$$

where $\mu_j = \mathbb{E}_{\pi_j}\big[b'(\theta)\big]$. If we consider a random sample $\boldsymbol{x_n}$, standard calculations (see, for instance, Bernardo & Smith, 1994 [4]) yield

$$\pi_j(\theta|\boldsymbol{x_n}) \propto \exp\big\{(n_j + n)\big[\mu_j(\boldsymbol{x_n})\theta - b(\theta)\big]\big\} \tag{7}$$

where

$$\mu_j(\boldsymbol{x_n}) = \mathbb{E}_{\pi_j}\left[b'(\theta)|\boldsymbol{x_n}\right] = w_{j,n}\mu_j + (1 - w_{j,n})\bar{x}_n,$$

with $w_{j,n} = \frac{n_j}{n+n_j}$. In this setup, $a_j = \mu_j(\boldsymbol{x_n})$ and Equation (3) becomes

$$\rho(\boldsymbol{x_n}, a_o; \pi_e) = \mathbb{V}_{\pi_e}\left[b'(\theta)|\boldsymbol{x_n}\right] + D_{e,o}(\boldsymbol{x_n}), \tag{8}$$

where $\mathbb{V}_{\pi_e}[b'(\theta)|\boldsymbol{x_n}]$ has to be specified according to the specific member of the exponential family under consideration, and $D_{e,o}(\boldsymbol{x_n})$ is

$$D_{e,o}(\boldsymbol{x_n}) = \left(a_e - a_o\right)^2 = \left(\mu_e(\boldsymbol{x_n}) - \mu_o(\boldsymbol{x_n})\right)^2 = \left(A_n\bar{x}_n - B_n\right)^2, \tag{9}$$

where

$$A_n = (w_{o,n} - w_{e,n}) \qquad \text{and} \qquad B_n = (w_{o,n}\mu_o - w_{e,n}\mu_e). \tag{10}$$

Hence, from Equation (4), e_n is

$$e_n = \mathbb{E}_{m_d}\left[\mathbb{V}_{\pi_e}\left[b'(\theta)|\boldsymbol{X_n}\right]\right] + \mathbb{E}_{m_d}\left[D_{e,o}(\boldsymbol{X_n})\right], \tag{11}$$

and an explicit expression for its second term can be found as follows

$$\begin{aligned}
\mathbb{E}_{m_d}\left[D_{e,o}(\boldsymbol{X_n})\right] &= \mathbb{E}_{m_d}\left[A_n^2\bar{X}_n^2 + B_n^2 - 2\bar{X}_nA_nB_n\right] \\
&= \left(\mathbb{V}_{m_d}\left[\bar{X}_n\right] + \mu_d^2\right)A_n^2 + B_n^2 - 2\mu_dA_nB_n \\
&= A_n^2\mathbb{V}_{m_d}\left[\bar{X}_n\right] + \left[A_n\mu_d - B_n\right]^2,
\end{aligned} \tag{12}$$

where, recalling that $\mathbb{E}_f(X) = b'(\theta)$ and $\mathbb{V}_f(X) = b''(\theta)$, for the variance decomposition formula we have that

$$\mathbb{V}_{m_d}\left[\bar{X}_n\right] = \frac{1}{n}\mathbb{E}_{\pi_d}\left[b''(\theta)\right] + \mathbb{V}_{\pi_d}\left[b'(\theta)\right]. \tag{13}$$

Remarks

a) It is easy to show that (i) $\mathbb{E}_{m_d}\left[\mathbb{V}_{\pi_e}\left[b'(\theta)|\boldsymbol{X_n}\right]\right] = o(n^{-1})$ and (ii) $\mathbb{E}_{m_d}\left[D_{e,o}(\boldsymbol{X_n})\right] = o(n^{-2})$. Hence, from (11), $e_n = o(n^{-1})$.

(i) Although a general explicit expression for $\mathbb{V}_{\pi_e}[b'(\theta)|\boldsymbol{x_n}]$ is not available, it is possible to determine the rate of convergence of $\mathbb{E}_{m_d}\left[\mathbb{V}_{\pi_e}\left[b'(\theta)|\boldsymbol{X_n}\right]\right]$ to zero as n increases. For i.i.d. observations, under the usual exponential family regularity conditions, we have that, for large n,

$$\mathbb{V}_{\pi_e}\left[b'(\theta)|\boldsymbol{x_n}\right] \simeq \frac{1}{n}b''(\hat{\theta})^2 \cdot I_1^{-1}(\hat{\theta}) = \frac{1}{n} \cdot h(\boldsymbol{x_n})$$

where $\hat{\theta}$ is the MLE and $I_1(\cdot)$ is the Fisher information for a single observation. Assuming existence and finiteness of $b''(\theta)$ and of $I_1^{-1}(\cdot)$, $\forall\theta \in \Theta$, then $h(\cdot)$ is a bounded function. Hence, for the Dominated Convergence Theorem, $\mathbb{E}_{m_d}\big[h(\boldsymbol{X_n})\big] = O(1)$ and $\mathbb{E}_{m_d}\big[\mathbb{V}_{\pi_e}\left[b'(\theta)|\boldsymbol{X_n}\right]\big] = o(n^{-1})$.

(ii) For the second term, since $\mathbb{V}_{m_d}\left[\bar{X}_n\right]$, A_n and B_n are all $o(n^{-1})$, from Equation (12) if follows that $\mathbb{E}_{m_d}\big[D_{e,o}(\boldsymbol{X_n})\big] = o(n^{-2})$.

b) Existence of the marginal distribution $m_d(\cdot)$ requires that π_d is a proper prior, i.e. $n_d > 0$. Furthermore, for any given n, when $n_d \to +\infty$, m_d reduces to the sampling distribution $f(\cdot|\theta_d)$.

c) When $n_j \to 0$ (for $j = e$ or $j = o$), π_j becomes noninformative and a_j coincides with the MLE. In particular, when π_o is a noninformative prior, the measure of discrepancy of Brutti et al (2014) [5] is retrieved as a special case of $D_{e,o}(\boldsymbol{x_n})$.

d) If $\pi_e = \pi_d$, the planner $\mathcal{P}_d$ coincides with the decision maker $\mathcal{E}_e$, as in Etzioni and Kadane (1993) [11]. Note that, in this case, $n_e = n_d > 0$.

3.1.1 Normal-Normal model

Assume that $X_1, X_2, \ldots, X_n$ is a random sample from a $\mathrm{N}(\lambda, \sigma^2)$ distribution with known variance. The density function of X can be written in the form (5) by setting

$$h(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{x^2}{2\sigma^2}\right\}, \quad \theta = \frac{\lambda}{\sigma^2}, \quad b(\theta) = \frac{\lambda^2}{2\sigma^2}.$$

Hence, $b(\theta) = \frac{\theta^2\sigma^2}{2}$, $b'(\theta) = \sigma^2\theta = \lambda = \mathbb{E}_f(X)$ and $b''(\theta) = \sigma^2 = \mathbb{V}_f(X)$. The conjugate prior for $\lambda = b'(\theta)$ is obtained from (6) by taking $c(n_j, \mu_j) = \frac{\sqrt{n_j}\sigma}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}n_j\mu_j^2\right\}$ and it is straightforward to check that

$$\mu_j = \mathbb{E}_{\pi_j}\big[b'(\theta)\big] = \mathbb{E}_{\pi_j}\left[\lambda\right], \qquad \sigma_j^2 = \mathbb{V}_{\pi_j}\big[b'(\theta)\big] = \frac{\sigma^2}{n_j}.$$

From (7), it follows that

$$\mu_j(\boldsymbol{x_n}) = \mathbb{E}_{\pi_j}\left[b'(\theta)|\boldsymbol{x_n}\right] = \frac{n_j\mu_j + n\bar{x}_n}{n_j + n}, \quad \sigma_j^2(\boldsymbol{x_n}) = \mathbb{V}_{\pi_j}\left[b'(\theta)|\boldsymbol{x_n}\right] = \frac{\sigma^2}{n_j + n},$$

and, from (8), the expected loss of a_o w.r.t. π_e is

$$\rho(\boldsymbol{x_n}, a_o; \pi_e) = \frac{\sigma^2}{n_e + n} + \big[\mu_e(\boldsymbol{x_n}) - \mu_o(\boldsymbol{x_n})\big]^2.$$

Noting that $\mathbb{E}_{\pi_d}\left[b''(\theta)\right] = \sigma^2$, $\mathbb{V}_{\pi_d}\left[b'(\theta)\right] = \frac{\sigma^2}{n_d}$ and therefore that $\mathbb{V}_{m_d}\left[\bar{X}_n\right] = \sigma^2\left(\frac{1}{n}+\frac{1}{n_d}\right)$, from Equations (12) and (13) we obtain that

$$\mathbb{E}_{m_d}\left[D_{e,o}(\boldsymbol{X_n})\right] = A_n^2\sigma^2\left(\frac{1}{n}+\frac{1}{n_d}\right) + \left[A_n\mu_d - B_n\right]^2,$$

and finally that

$$e_n = \frac{\sigma^2}{n_e+n} + A_n^2\sigma^2\left(\frac{1}{n}+\frac{1}{n_d}\right) + \left[A_n\mu_d - B_n\right]^2, \tag{14}$$

where A_n and B_n are defined in (10).

Remarks

a) If we assume that $\pi_d = \pi_e$, the expression of e_n given by Etzioni and Kadane (1993) [11] is found by replacing (μ_d, n_d) with (μ_e, n_e) in (14).

b) If π_o is the noninformative prior ($n_o = 0$), then $A_n = -w_{e,n}$, $B_n = -w_{e,n}\mu_e$ and $e_n = \frac{\sigma^2}{n_e+n} + w_{e,n}^2\left[\sigma^2\left(\frac{1}{n}+\frac{1}{n_d}\right) + \left(\mu_e-\mu_d\right)^2\right]$, where the second member of the sum coincides with the predictive expected discrepancy between the MLE $\bar{X}_n$ and the Bayesian estimator $\mu_e(\boldsymbol{X_n})$ given in Brutti et al (2014) [5].

c) Extending these results to the unknown variance case is in principle straightforward by using the standard Normal-InvertedGamma model for (λ, σ^2) but we do not consider it here.

Numerical examples

Let us illustrate some numerical examples related to the Normal case results. Normal assumption for an experimental outcome X is quite common in clinical trials applications, for instance in phase II efficacy trials, where the mean λ denotes treatment effect (e.g. a continuous measure of tumour reduction in oncology studies). However, the same basic model provides an approximation that can be used, for instance, for binary data - with λ denoting a log-odds ratio - and for survival data - with λ denoting a log-hazard function, as described in details in Spiegelhalter et al (2004) [29].

Let us first recall a typical experimental situation in which our proposed methodology can be conveniently used. For example, a pharmaceutical company - that in this case plays the role of $\mathcal{E}_e$ - wants to promote an innovative treatment by convincing a skeptical regulatory agency - $\mathcal{E}_o$ - of the superiority of its product with respect to the standard therapy available on the market. Hence, the goal of the experiment is to show that the effect of the new treatment is larger than a given threshold. In the following examples we need to elicit the parameters of the priors related to the three parties involved in this experiment, bearing in mind their respective attitude towards the trial success. Basically, we assume that planner $\mathcal{P}_d$ has a very enthusiastic opinion on the parameter of interest, $\mathcal{E}_e$ is slightly more optimistic than the final user $\mathcal{E}_o$, who may adopt either a noninformative or a skeptical prior. This underlying structure of relationships between the three actors is translated into a convenient choice of prior means – in general $\mu_d \gg \mu_e > \mu_o$ – and prior sample sizes – typically $n_d \gg n_e \geq n_o$.

In Figure 1, the posterior expected loss e_n, as defined in Equation (14), is plotted with respect to the sample size n. The contributions to e_n of the expected posterior variance and of the expected discrepancy are highlighted in two different colors. Consistently with the convergence properties shown in Section 3.1 for the canonical exponential family (see Remark a), e_n decreases to 0 as the sample size n increases, for all the considered configurations of the other parameters. The comparison of the different panels from left to right allows us to evaluate the impact of the prior sample size n_e: a more concentrated optimistic prior for $\mathcal{E}_e$ yields uniformly lower values of e_n (panels (a)-(c)-(e) with respect to panels (b)-(d)-(f)). Moreover, looking at each column of the plots, for any fixed value of μ_o, larger and larger values of μ_e imply a more relevant weight of the expected discrepancy in e_n (darker area) that, however, quickly reduces as the sample size increases. Similar considerations arise from Figure 2, where we consider a noninformative prior π_o ($n_o = 0$). With respect to the previous case ($n_o = 10$), for small values of n, the noninformative prior implies a larger contribution of the expected discrepancy to e_n but, for larger sample sizes this contribution tends to disappear more rapidly than it is observed in Figure 1.

When a point mass prior is assumed for π_d (see Section 3.1 Remark b), the situation is substantially unchanged and we omit the corresponding plots.

In Figure 3 we consider a reverted point of view: for a fixed value of the sample size, $n = 25$, we let e_n vary as a function of the prior sample size n_e for several choices of the design prior mean μ_d (see different line types in each panel) and for increasing values of n_o

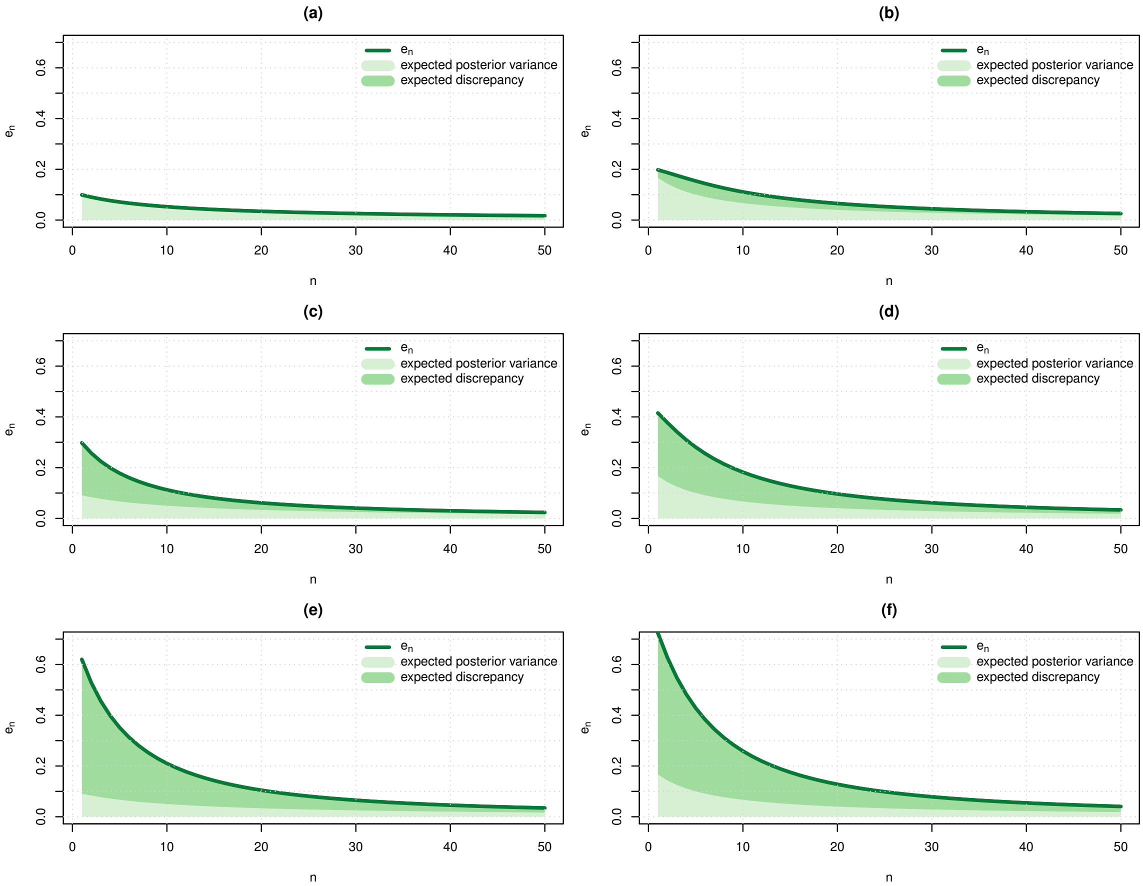

Figure 1: Behavior of e_n as a function of n, for several values of prior parameters (a) $\mu_e = 0.1, n_e = 10$; (b) $\mu_e = 0.1$, $n_e = 5$; (c) $\mu_e = 0.5, n_e = 10$; (d) $\mu_e = 0.5$, $n_e = 5$; (e) $\mu_e = 0.8, n_e = 10$; (f) $\mu_e = 0.8$, $n_e = 5$; when $\sigma^2 = 1$, $n_d = 20$, $\mu_d = 1$, $n_o = 10$, $\mu_o = 0$.

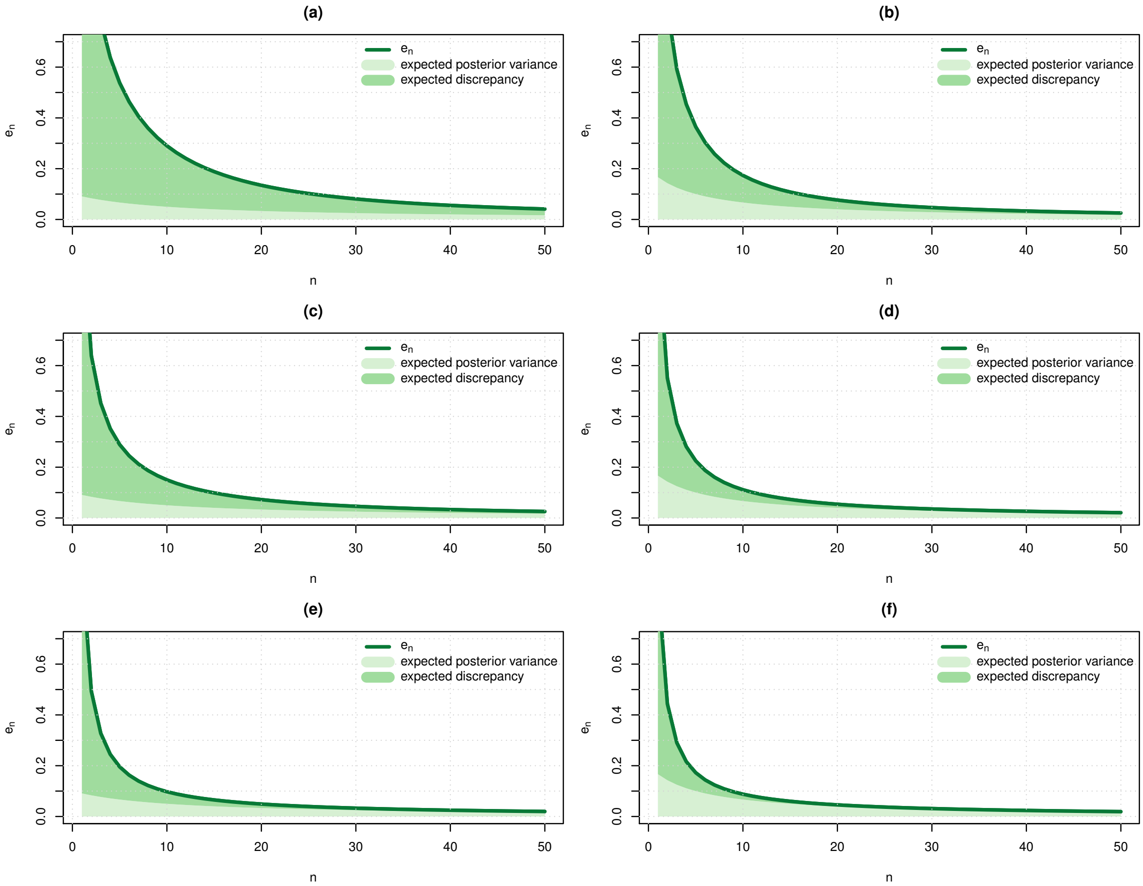

Figure 2: Behavior of e_n as a function of n, for several values of prior parameters (a) $\mu_e = 0.1, n_e = 10$; (b) $\mu_e = 0.1$, $n_e = 5$; (c) $\mu_e = 0.5, n_e = 10$; (d) $\mu_e = 0.5$, $n_e = 5$; (e) $\mu_e = 0.8, n_e = 10$; (d) $\mu_e = 0.8$, $n_e = 5$; when $\sigma^2 = 1$, $n_d = 20$, $\mu_d = 1$, $n_o = 0$, $\mu_o = 0$ (noninformative π_o).

(moving from panel (a) to panel (f)). Let us first focus on panel (a). Being π_e moderately optimistic, as expected, e_n gets larger and larger values as its prior weight n_e increases and the growth is more rapid and more pronounced for increasing values of μ_d. In the remaining panels this monotonic behaviour does not hold anymore: the effect of assuming a more and more concentrated prior π_o (moving from panel (a) to panel (f)) is that of inducing a first reduction of e_n up to a minimum value, approximately in correspondence of $n_e \approx n_o$, followed by a new increase of e_n with respect to n_e.

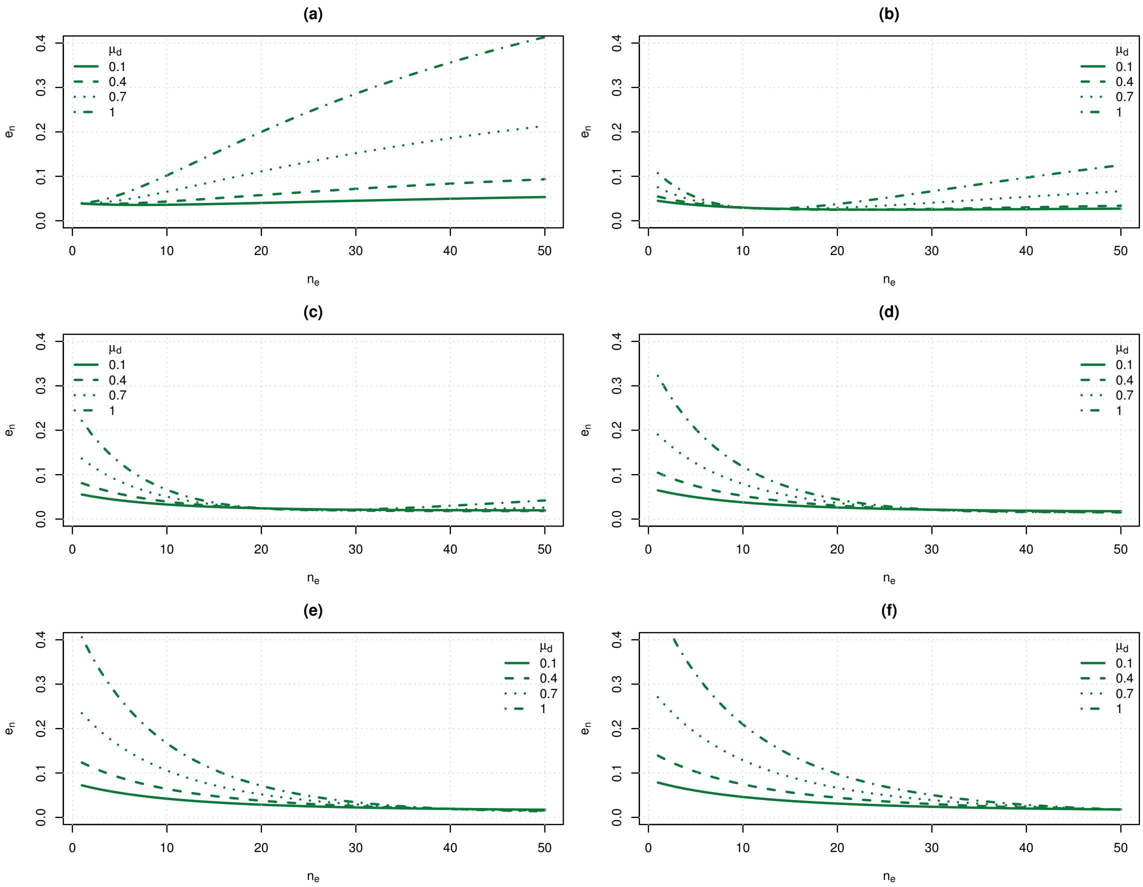

Figure 3: Behavior of e_n as a function of n_e, given a fixed sample size $n = 25$, for several values of μ_d, when $\mu_e = 0.1$, $\mu_o = 0$; $\sigma^2 = 1$, (a) $n_o = 0$; (b) $n_o = 10$; (c) $n_o = 20$; (d) $n_o = 30$; (e) $n_o = 40$; (f) $n_o = 50$.

Finally, Table 1(a) contains the values of the optimal sample sizes based on criterion (1) for several choices of the prior parameters. As a consequence of the behaviour of e_n with respect to n, if we focus on each block of the table we notice that increasingly larger

(a) $n_d = 10$

μ_d	0.1			0.5			1		
	n_e			n_e			n_e		
n_o	0	10	20	0	10	20	0	10	20
0	20	20	25	20	27	39	20	45	77
10	26	11	5	34	11	3	51	11	2
20	34	15	4	51	21	4	88	37	4
30	40	19	5	67	35	8	125	72	19
40	47	24	6	84	51	17	162	108	54
50	53	28	7	100	67	31	198	145	90

(b) $n_d = \infty$

μ_d	0.1			0.5			1		
	n_e			n_e			n_e		
n_o	0	10	20	0	10	20	0	10	20
0	20	16	16	20	23	31	20	43	72
10	23	11	5	31	11	3	49	11	2
20	25	13	4	44	20	4	84	35	4
30	28	15	5	57	30	7	118	68	18
40	29	17	5	70	42	14	153	102	50
50	31	18	6	82	54	23	188	137	85

Table 1: Optimal sample sizes based on criterion 1 for several choices of the prior parameters, given $\mu_e = 0.1$ and $\mu_o = 0$, with a threshold $\gamma = 0.05$.

values of n_e (see table rows), which means a more concentrated optimistic prior for $\mathcal{E}_e$, imply smaller and smaller optimal sample sizes. Notice that this behavior holds with the exception of the case $n_o = 0$: when π_o is noninformative, the expected discrepancy between a_o and a_e is more and more relevant as n_e increases. Conversely, looking at the table columns, when n_o increases, being $\mathcal{E}_o$ more and more skeptical about the parameter, the corresponding optimal sample sizes are uniformly larger. In other words, since the opinions of $\mathcal{E}_e$ and $\mathcal{E}_o$ progressively diverge, a higher number of experimental observations is required to bring a consensus between the two decision makers. By comparing the three blocks in each table, we also notice that larger values of μ_d yield an increase in the optimal sample size, which simply follows from expression in (14). This can be interpreted as a consequence of the impact of the extremely optimistic attitude of the planner $\mathcal{P}_d$ that makes it harder (in terms of required number of observations) to solve the conflict between the two adversaries. Similar considerations also apply to Table 1(b), where a point mass design prior ($n_d = \infty$) is considered. Notice that, as a consequence of the use of the conditional approach - which does not take into account uncertainty on the design value - the optimal sample sizes in Table 1(b) are uniformly smaller than those in Table 1(a).

	SSD criterion					
	e_n		$\mathbb{E}_{m_d}\big[\mathbb{V}_{\pi_e}[\lambda]\big]$		$\mathbb{E}_{m_d}\big[D_{e,o}(\boldsymbol{X_n})\big]$	
n_e	$n_0 = 0$	$n_0 = 30$	$n_0 = 0$	$n_0 = 30$	$n_0 = 0$	$n_0 = 30$
0	20	62	20	20	1	47
10	22	38	10	10	12	1
30	36	9	1	1	27	1

Table 2: Optimal sample sizes obtained using the three different criteria based on e_n, $\mathbb{E}_{m_d}\big[D_{e,o}(\boldsymbol{X_n})\big]$ and $\mathbb{E}_{m_d}\big[\mathbb{V}_{\pi_e}[\lambda]\big]$ respectively, for different choices of n_e and n_o, given $\mu_o = 0$, $\mu_e = 0.2$, $\mu_d = 0.5$ and $n_d = 20$ with a threshold $\gamma = 0.05$.

In Table 2 we compare the optimal sample sizes based on criterion (1) to those obtained using analogous criteria based either on the expected posterior variance $\mathbb{E}_{m_d}\big[\mathbb{V}_{\pi_e}[\lambda]\big]$ – which ignores the opinion of $\mathcal{E}_o$ – or on the expected discrepancy $\mathbb{E}_{m_d}\big[D_{e,o}(\boldsymbol{X_n})\big]$ – which corresponds to the non decisional criterion introduced in Brutti et al (2014) [5]. From the general decomposition of e_n under the quadratic loss, given in Equation (4), it easily follows that both $\mathbb{E}_{m_d}\left[D_{e,o}(\boldsymbol{X_n})\right]$ and $\mathbb{E}_{m_d}\left[\mathbb{V}_{\pi_e}[\lambda]\right]$ assume smaller values than e_n for any given n. As a consequence, the corresponding optimal sample sizes are smaller for any chosen configuration of the prior parameters. Notice that, when the expected posterior variance is used alone (see the third and the fourth columns), the optimal sample sizes are not affected by the choice of n_o because they only depend on π_e's prior parameters. Conversely, when examining the first and the fifth columns, we realize that, as the prior opinions of $\mathcal{E}_e$ and $\mathcal{E}_o$ diverge (i.e. π_o remains noninformative, while π_e is taken to be more and more concentrated on a slighlty enthusiastic mean μ_e), both criteria based on e_n and on $\mathbb{E}_{m_d}\left[D_{e,o}(\boldsymbol{X_n})\right]$ yield increasing values of optimal sample sizes.

3.1.2 Exponential-Inverted Gamma case model

Assume now that $X_1, X_2, \ldots, X_n$ is a random sample from an Exp(λ) distribution. The density function of X can be written in the form (5) by setting

$$h(x) = 1, \quad \theta = -\frac{1}{\lambda}, \quad b(\theta) = -\ln\frac{1}{\lambda}.$$

Hence, $b(\theta) = -\ln(-\theta)$, $b'(\theta) = -\frac{1}{\theta} = \lambda = \mathbb{E}_f(X)$, and $b''(\theta) = \frac{1}{\theta^2} = \lambda^2 = \mathbb{V}_f(X)$. The conjugate prior for $\lambda = b'(\theta)$, obtained from (6) by taking $c(n_j, \mu_j) = \frac{(n_j\mu_j)^{n_j+1}}{\Gamma(n_j+1)}$, is an Inverted Gamma density of parameters $(n_j+1, n_j\mu_j)$ with

$$\mu_j = \mathbb{E}_{\pi_j}\left[b'(\theta)\right] = \mathbb{E}_{\pi_j}\left[\lambda\right], \quad \sigma_j^2 = \mathbb{V}_{\pi_j}\left(b'(\theta)\right) = \mathbb{V}_{\pi_j}\left(\lambda\right) = \frac{n_j^2\mu_j^2}{n_j^2(n_j-1)} = \frac{\mu_j^2}{n_j-1}.$$

Similarly, the posterior distribution (7) is an Inverted Gamma with updated parameters $[n+n_j+1, (n+n_j)\mu_j]$ and

$$\mu_j(\boldsymbol{x_n}) = \mathbb{E}_{\pi_j}\left[b'(\theta)|\boldsymbol{x_n}\right] = \frac{n_j\mu_j + n\bar{x}_n}{n_j+n}, \quad \sigma_j^2(\boldsymbol{x_n}) = \mathbb{V}_{\pi_j}\left[b'(\theta)|\boldsymbol{x_n}\right] = \frac{(\mu_j(\boldsymbol{x_n}))^2}{n+n_j-1}.$$

From (8) the expected loss of a_o w.r.t. π_e is

$$\rho(\boldsymbol{x_n}, a_o; \pi_e) = \frac{[\mu_e(\boldsymbol{x_n})]^2}{n+n_e-1} + \left[\mu_e(\boldsymbol{x_n}) - \mu_o(\boldsymbol{x_n})\right]^2. \tag{15}$$

In order to find the predictive expectation of $\rho(\boldsymbol{X_n}, a_o; \pi_e)$ notice that, since

$$\mathbb{V}_{m_d}\left[\bar{X}_n\right] = \frac{1}{n}\mathbb{E}_{\pi_d}\left[b''(\theta)\right] + \mathbb{V}_{\pi_d}\left[b'(\theta)\right] = \mu_d^2\left[\frac{n+n_d}{n(n_d-1)}\right],$$

it follows that

$$\begin{aligned}\mathbb{E}_{m_d}\left[\mu_j(\boldsymbol{x_n})^2\right] &= \mathbb{E}_{m_d}\left[\left((1-w_{e,n})^2\bar{X}_n^2 + 2w_{e,n}(1-w_{e,n})\mu_e\bar{X}_n + w_{e,n}^2\mu_e^2\right)\right] \\ &= (1-w_{e,n})^2\left[\mathbb{V}_{m_d}\left[\bar{X}_n\right] + \mathbb{E}_{m_d}\left[\bar{X}_n\right]^2\right] + 2w_{e,n}(1-w_{e,n})\mu_e\mathbb{E}_{m_d}\left[\bar{X}_n\right] + w_{e,n}^2\mu_e^2 \\ &= (1-w_{e,n})^2\mu_d^2\left[\frac{n+n_d}{n(n_d-1)}\right] + \left[(1-w_{e,n})\mu_d + w_{e,n}\mu_e\right]^2, \qquad (16)\end{aligned}$$

and that

$$\mathbb{E}_{m_d}\left[\left(\mu_e(\boldsymbol{x_n}) - \mu_o(\boldsymbol{x_n})\right)^2\right] = A_n^2\mu_d^2\left[\frac{n+n_d}{n(n_d-1)}\right] + \left(A_n\mu_d - B_n\right)^2, \qquad (17)$$

where A_n and B_n are defined in (10). In summary, simple algebra yields

$$e_n = \mu_d^2\left[\frac{n+n_d}{n(n_d-1)}\right]\left[\frac{(1-w_{e,n})^2}{n+n_e-1} + A_n^2\right] + \frac{\left[(1-w_{e,n})\mu_d + w_{e,n}\mu_e\right]^2}{n+n_e-1} + (A_n\mu_d - B_n)^2\,.$$

Remarks

a) The case $\pi_d = \pi_e$ is obtained from the above expression of e_n by simply replacing (μ_d, n_d) with (μ_e, n_e).

b) The noninformative case for π_o is obtained by setting $\mu_o = 0$ and $n_o + 1 = 0$.

Numerical examples

In this section we provide some numerical results for the Exponential case and we discuss the impact of parameters values on the quantity e_n and, consequently, on the optimal sample sizes choice. A relevant context of application of this model is industrial and quality control, where the Exponential assumption is often made in order to model time-to-events data as, for example, completion time for a given process or lifetime of a specific device. For instance, Lindley and Singpurwalla (1991) [19] describe an experimental context in which a manufacturer ($\mathcal{E}_e$) aims at selling a product to a consumer ($\mathcal{E}_o$) who accepts or rejects the product on the basis of the evidence produced by the manifacturer. Again, in our setup the design is planned by an optimistic planner $\mathcal{P}_d$.

Under this framework, we consider the behaviour of e_n as a function of n for some interesting configurations of the parameters. Figure 4 allows us to compare the impact on e_n of different choices for the parameters of π_e. For $\mu_o = 0.1$, a larger (namely, more optimistic) prior mean μ_e has a remarkable impact on e_n, especially on the expected discrepancy term (darker area). In Figure 5 similar plots are drawn under the assumption of a noninformative π_o. Notice that, for small sample sizes, the expected discrepancy has a dramatic impact on e_n, which substantially overwhelms the expected posterior variance. However, as n increases the discrepancy tends to 0, more and more rapidly when the prior sample size n_e is smaller and when the prior mean μ_e gets closer to the design prior mean $\mu_d = 1$.

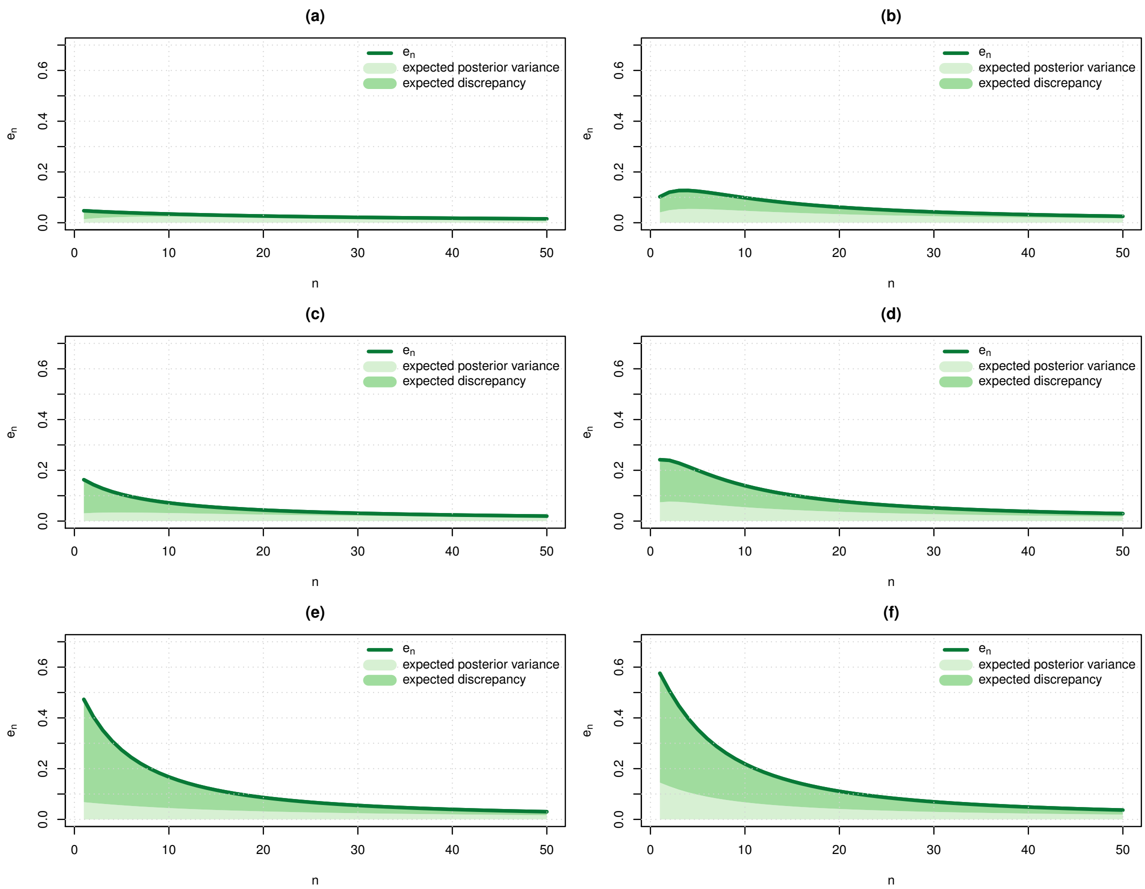

Figure 4: Behavior of e_n as a function of n, for several values of prior parameters (a) $\mu_e = 0.3, n_e = 10$; (b) $\mu_e = 0.3$, $n_e = 5$; (c) $\mu_e = 0.5, n_e = 10$; (d) $\mu_e = 0.5$, $n_e = 5$; (e) $\mu_e = 0.8, n_e = 10$; (d) $\mu_e = 0.8$, $n_e = 5$; when $n_d = 20$, $\mu_d = 1$, $n_o = 10$, $\mu_o = 0.1$.

Finally, in Table 3 the optimal sample sizes are reported for different choices of μ_d and n_d, letting the two prior sample sizes n_e and n_o vary. In each subtable, when n_e takes larger and larger values, a smaller number of units is required. The opposite behaviour is observed when π_o is noninformative, since, as discussed in the previous section, for $n_o = 0$ the expected discrepancy term is more and more relevant as n_e increases. In general, increasing values of n_o yields larger and larger sample sizes.

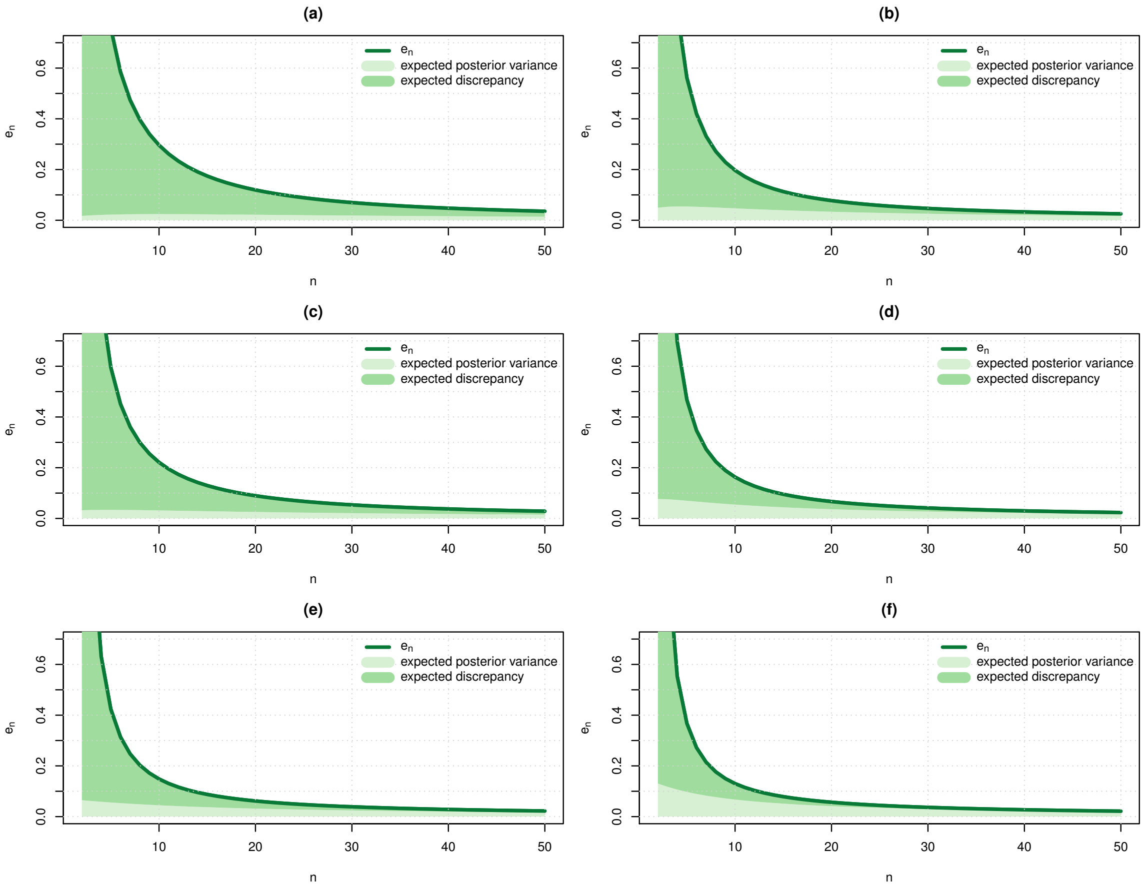

Figure 5: Behavior of e_n as a function of n, for several values of prior parameters (a) $\mu_e = 0.3, n_e = 10$; (b) $\mu_e = 0.3$, $n_e = 5$; (c) $\mu_e = 0.5, n_e = 10$; (d) $\mu_e = 0.5$, $n_e = 5$; (e) $\mu_e = 0.8, n_e = 10$; (d) $\mu_e = 0.8$, $n_e = 5$; when $n_d = 20$, $\mu_d = 1$, $n_o + 1 = 0$, $\mu_o = 0$ (noninformative π_o).

Moreover, looking at each block in the tables we notice again that, for larger values of μ_d, the optimal sample sizes increase: the values in Table 3(b) are slightly, but uniformly smaller than the values in Table 3(a), due to the choice of a point mass design prior ($n_d = \infty$) which eliminates the uncertainty in the elicitation of the prior from $\mathcal{P}_d$.

Finally, Table 4 is reported for the sake of completeness, but the comments related to Table 2 for the normal model perfectly apply to the exponential case as well.

4 Conclusions

The main goal of this article is to show some aspects of the complexity of the sample size problem. The most relevant features of our approach are the following.

1. We adopt of a complete decisional framework, in which the sample size criterion is based on the predictive behavior of the posterior expected loss of the chosen decision, rather than on the predictive features of the action itself as it is the case in the performance-based criteria. See, for instance, Brutti et. al (2014b) [6] for discussion.

2. We consider the sample size problem in the presence of multiple parties. With respect to Etzioni and Kadane (1993) [11], our main contribution here is the distinction between $\mathcal{E}_e$ and $\mathcal{P}_d$. As numerical examples show, the attitude of the planner $\mathcal{P}_d$ has a remarkable impact on the optimal dimension of the experiment. In particular, when the design prior π_d is concentrated on extremely optimistic values of the parameter, compared to the other two priors, namely $\mu_d \gg \mu_e \geq \mu_o$, then a larger number of observations is required to let $\mathcal{E}_e$ and $\mathcal{E}_o$ come to an agreement.

3. The decisional approach followed in the paper sheds new light on the analysis of the conflict between alternative procedures considered in Brutti et al. (2014)[5]: the measure of conflict between the decisions a_o and a_e, $D_{e,o}(\boldsymbol{x_n})$, is found to coincide with the difference between the corresponding posterior expected losses $\rho(\boldsymbol{x_n}, a_o; \pi_e) - \rho(\boldsymbol{x_n}, a_e; \pi_e)$, as previously noticed by Etzioni and Kadane (1993) [11]. The numerical consequences of this fact have been discussed in the examples of Section 3.1.1 and 3.1.2.

There are several potential extensions of the ideas contained in the present article. Here is a non exhaustive list of possibilities.

1. Model and prior assumptions. In the article we limit our analysis to one-parameter exponential families with conjugate priors, mainly because this choice allow us to obtain closed-form expressions that can be easily studied and commented on. Other choices are of course possible and the resulting SSD problems could be easily ad-

dressed numerically, following the simulation-based approach discussed, for instance, in [31].

2. Loss function and decision problems. We can consider loss functions and/or decision problems different from quadratic loss and point estimation. One idea is to consider the use of the logarithmic loss function for the choice of a probability distribution, as in Etzioni and Kadane (1993) [11]. Another example is the set estimation problem. In this case, if we consider the most widely used loss function for a set estimate S of θ, that is the linear loss $L(S, \theta) = \text{mis}(S) - I_S(\theta)$, where mis$(S)$ is the measure of the set S and $I_S(\cdot)$ its indicator function, then the posterior expected loss of the set C_o (optimal w.r.t. π_o) is $\rho(\boldsymbol{x_n}, C_o; \pi_e) = \text{mis}(C_o) - P_{\pi_e}(\theta \in C_o|\boldsymbol{x_n})$.

3. Robustness 1. Instead of considering only one decision maker $\mathcal{E}_e$ we could extend our approach by considering an entire class of decisioners, each with her/his own prior π_e belonging to a class of distributions Γ. In this case, we would be interested in evaluating the worst loss of a_o, when evaluated by the posterior expected losses associated to the priors in Γ, that is $\sup_{\pi_e \in \Gamma} \rho(\boldsymbol{x_n}, a_o; \pi_e)$. The sample size would then be chosen by replacing e_n in with $e_n^{\Gamma} = \mathbb{E}_{m_d}[\rho_{\Gamma}(a_o(\boldsymbol{X_n}))]$ in expression (1).

4. Robustness 2. As a final remark note also that the quantities in (3) are related to the the concept of ϵ-posterior robustness introduced in Berger (1985, p. 205) [2]. If we assume a class Γ of prior distributions for $\mathcal{E}_e$, an action a_o is defined ϵ-posterior robust with respect to Γ if, for all $\pi_e \in \Gamma$, $D_{e,o}(\boldsymbol{x_n}) = \rho(\boldsymbol{x_n}, a_o; \pi_e) - \inf_{a \in \mathcal{A}} \rho(\boldsymbol{x_n}, a; \pi_e) \leq \epsilon$. Therefore, a straightforward SSD criterion for making a_o an ϵ-posterior robust decision can be based on the predictive expected value $\mathbb{E}_{m_d}[\sup_{\pi_e \in \Gamma} D_{e,o}(\boldsymbol{x_n})]$, which is the robust version of the criterion proposed in Brutti et a. (2014) [5].

We plan to elaborate on these ideas and connections in future research.

(a) $n_d = 10$

μ_d	0.5			0.8			1		
	n_e			n_e			n_e		
$n_o + 1$	0	10	20	0	10	20	0	10	20
$0^\star$	9	9	10	18	22	30	26	35	51
10	16	10	10	33	12	7	47	15	6
20	25	20	18	56	34	20	79	45	21
30	35	29	27	80	57	38	112	78	48
40	44	38	36	103	80	60	145	111	79
50	53	48	45	127	103	82	177	143	111

(b) $n_d = \infty$

μ_d	0.5			0.8			1		
	n_e			n_e			n_e		
$n_o + 1$	0	10	20	0	10	20	0	10	20
$0^\star$	10	11	13	16	19	25	23	30	42
10	16	11	10	31	12	7	43	14	6
20	25	20	18	52	32	19	73	43	21
30	34	29	27	73	54	37	103	73	46
40	42	38	36	95	75	57	133	103	75
50	50	46	44	116	96	78	163	133	105

Table 3: Optimal sample sizes based on criterion 1 for several choices of the prior parameters, given $\mu_e = 0.5$ and $\mu_o = 0.1$ (* $\mu_o = 0$ for the noninformative prior), with a threshold $\gamma = 0.05$.

	SSD criterion					
	e_n		$\mathbb{E}_{m_d}\big[D_{e,o}(\boldsymbol{X_n})\big]$		$\mathbb{E}_{m_d}\big[\mathbb{V}_{\pi_e}[\lambda]\big]$	
n_e	(μ_0, n_0+1)					
	$(0,0)$	$(0.1,30)$	$(0,0)$	$(0.1,30)$	$(0,0)$	$(0.1,30)$
0	17	76	16	16	6	66
10	20	55	1	1	15	46
30	33	27	1	1	29	23

Table 4: Optimal sample sizes obtained using the three different criteria based on e_n, $\mathbb{E}_{m_d}\big[D_{e,o}(\boldsymbol{X_n})\big]$ and $\mathbb{E}_{m_d}\big[\mathbb{V}_{\pi_e}[\lambda]\big]$ respectively, for different choices of n_e μ_o and n_o, given $\mu_e = 0.5$, $\mu_d = 0.8$ and $n_d = 20$, with a threshold $\gamma = 0.05$.

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. 1995; 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] Chaloner K, Verdinelli I. Bayesian Experimental Design: a review. *Stat. Sci.* 1995; **10**(3):273-304.

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

[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.