Predictive Control of Posterior Robustness in a Bernoulli Model

Fulvio De Santis, Maria Clara Fasciolo and Stefania Gubbiotti [1]

Dipartimento di Scienze Statistiche, Sapienza Università di Roma

Abstract

In this article we consider the sample size determination problem in the context of robust Bayesian parameter estimation of the Bernoulli model. Following a robust approach, we consider classes of conjugate Beta prior distributions for the unknown parameter. We assume that inference is robust if posterior quantities of interest (such as point estimates and limits of credible intervals) do not change too much as the prior varies in the selected classes of priors. For the sample size problem, we consider criteria based on predictive distributions of lower bound, upper bound and range of the posterior quantity of interest. The sample size is selected so that, before observing the data, one is confident to observe a small value for the posterior range and, depending on design goals, a large (small) value of the lower (upper) bound of the quantity of interest. We also discuss relationships with and comparison to standard classical and Bayesian methods.

Keywords: Bayesian robustness; clinical trials; conjugate analysis; posterior range; sample size determination.

1 Introduction

Let $\{f(\cdot;\theta),\ \theta \in \Theta \subseteq \mathbb{R}\}$ be a standard parametric model for a random variable X. We are interested in choosing the sample size n so that inference on the unknown parameter θ is accurate. We consider a Bayesian framework that allows one to take into account pre-experimental information and uncertainty on θ by introducing a prior distribution defined over the parameter space Θ. Formally, let $\mathbf{x}_n = (x_1, x_2, ..., x_n)$ be a realization of the random sample $\mathbf{X}_n = (X_1, X_2, ..., X_n)$ and $f_n(\mathbf{x}_n;\theta)$ be the likelihood function of θ. Given a prior distribution for θ, $\pi(\cdot)$ (that we assume to be continuous), the corresponding posterior density is

$$\pi(\theta \mid \mathbf{x}_n) = \frac{f_n(\mathbf{x}_n;\theta)\,\pi(\theta)}{m(\mathbf{x}_n;\pi)},$$

where $m(\mathbf{x}_n;\pi) = \int_\Theta f_n(\mathbf{x}_n;\theta)\,\pi(\theta)\,d\theta$ is the marginal or prior predictive density of the data. We assume that, for inference on θ, we are interested on a specific functional of the posterior, $\rho(\mathbf{x}_n;\pi)$. Typical examples of posterior quantities of interest include the posterior mean, the posterior variance, the posterior probability of a set, the inferior (superior) limit of a posterior interval estimate for θ.

It is well known that critical aspects of Bayesian inference are the choice of the prior distribution and the sensitivity of posterior inference on specific prior choices. The robust Bayesian approach to statistical inference, discussed in [2] and [3], replaces a single prior distribution with a class Γ of possible priors and studies variations of the posterior quantity of interest, $\rho(\mathbf{x}_n;\pi)$, as the prior varies in Γ. More specifically, for a given sample, let

$$L_n(\mathbf{x}_n) = \inf_{\pi\in\Gamma} \rho(\mathbf{x}_n;\pi) \quad \text{and} \quad U_n(\mathbf{x}_n) = \sup_{\pi\in\Gamma} \rho(\mathbf{x}_n;\pi) \tag{1}$$

be the observed lower and upper bounds of $\rho(\mathbf{x}_n;\pi)$. In general, inference is accurate if, as the prior varies over Γ, the posterior range of $\rho(\mathbf{x}_n;\pi)$, defined as

$$R_n(\mathbf{x}_n) = U_n(\mathbf{x}_n) - L_n(\mathbf{x}_n) \tag{2}$$

[1] **Address for correspondence**: Dipartimento di Scienze Statistiche. Sapienza Università di Roma. Piazzale Aldo Moro, 5 - 00185, Roma - Italia. stefania.gubbiotti@uniroma1.it

is sufficiently small. In addition to this, depending on the inferential problem, the evidence from an experiment is conclusive only if the lower (respectively upper) bound of $\rho(\mathbf{x}_n;\pi)$ is sufficiently large (respectively small).

Two factors are the main determinants for the values of lower bound, upper bound and range of $\rho(\mathbf{x}_n;\pi)$: the strength of sample information (essentially the size of the experiment) and the width of the class of priors. At least in standard problems, in order to reduce, for instance, the posterior range, one has either to increase the sample size or to restrict the class of prior distributions, by imposing some constraints on it. Similar considerations hold for obtaining conclusive evidence in terms of sufficiently large (small) values of lower (upper) bounds. Hence, the most typical recommendations for dealing with lack of posterior robustness and/or lack of conclusive evidence are either to collect more data or to refine prior elicitation. However, both these approaches are post-experimental and may present difficulties in their implementation (see [8]).

DasGupta and Mukhopadhyay in [7] and, more recently, De Santis in [8] propose to control the risk of lack of posterior robustness by addressing the problem as a pre-experimental design issue. These authors propose methods for selecting a sample size so that the chances that data yield robust evidence are sufficiently high. In fact, in the design stage the data, the posterior distribution and any of its functionals are random objects. Hence, upper bounds, lower bounds and range of posterior quantities of interest are random as well. The idea is then to select the minimal sample size that guarantees a probabilistic control on the performance of these quantities, as π varies in the class of prior distributions. This entails computations with the predictive distribution of the data and formulation of criteria based on summaries of the predictive distributions of L_n, U_n and R_n.

Several articles are specifically dedicated to the sample size determination (SSD) problem from a Bayesian perspective, including those by [10] and [12]. A general review of Frequentist and Bayesian techniques is given in [1]. Problems of SSD for robust Bayesian analysis have been previously considered by [7], [8], [4] and [5], who formalized a methodology for SSD and developed methods for robust estimation of the normal mean with conjugate priors. In this article we follow the same approach, but we consider the Bernoulli conjugate model instead of the normal framework.

This article is structured as follows. Section 2 shows how to carry out a robust Bayesian analysis using the posterior mean. It also introduces a distinction between priors used to obtain the posterior distribution of the parameter and priors used to define the marginal distribution of the data. Section 3 formalizes robust Bayesian SSD criteria for the estimation problem sketched in Section 2. Section 4 develops the proposed methods under the assumption that θ is a Bernoulli parameter and that conjugate priors are used. Numerical results and comparisons with non-robust and non-informative Bayesian approaches are discussed in Section 5. Section 6 gives an example in the context of designing a clinical trial. Section 7 contains concluding remarks.

2 Robust estimation

Assume that we are interested in estimating θ using the posterior expectation $\mathbb{E}(\theta \mid \mathbf{x}_n;\pi)$. For a given sample $\mathbf{x}_n$, we determine the posterior range as in (2), where the posterior bounds are obtained by setting $\rho(\mathbf{x}_n;\pi) = \mathbb{E}(\theta \mid \mathbf{x}_n;\pi)$ in (1). In the following sections we introduce criteria for posterior robustness, based on the control of the probability distribution of R_n, and criteria for robust evidence, based on the control of the probability distribution of L_n and/or U_n.

The posterior range measures the variation caused by the uncertainty in the prior and the hope is that R_n is small enough that the indeterminacy in the prior is deemed to be essentially irrelevant, allowing a claim of robustness with respect to variations of π in Γ. We make no attempt here to define what is a 'small' or 'large' posterior range, that is to define when one does or does not have posterior robustness. This is a problem-specific judgement. The idea, however, is straightforward. If the value of R_n is small enough to consider the differences between the

various priors in the class essentially irrelevant, then robustness is asserted and experimental results allow conclusive inferential evidence, provided Γ is large enough to reflect the possible uncertainty in the prior. Conversely, if the posterior range is not small enough, then robustness is lacking and posterior inference on θ is not conclusive.

In general, we need to select the probability distribution to be used for pre-posterior computations and sample size choice. Frequentist methods typically consider an initial guess $\tilde{\theta}$ on the unknown parameter and use the sampling distribution $f_n(\cdot;\tilde{\theta})$ of $\mathbf{X}_n$ for pre-posterior computations. This approach does not account for uncertainty on θ and yields designs that are only locally optimal (see [6] for discussion). In the Bayesian framework, one accounts for uncertainty on the guessed value of θ by using a prior distribution π_D, that we call *design prior*, and by replacing $f_n(\cdot;\tilde{\theta})$ with the marginal distribution of the data $m(\mathbf{x}_n;\pi_D)$ for pre-posterior analysis. In Section 2.1 we illustrate the reasons why the design prior π_D does not necessarily have to be the same prior distribution used for posterior inference (note that later on this will be called *analysis prior* and denoted by π_A). Therefore we are interested in the pre-experimental problem of selecting an appropriate sample size for controlling the robustness of $\mathbb{E}(\theta \mid \mathbf{x}_n;\pi)$ with respect to Γ. The idea is to choose the minimum number of observations so that a summary of the predictive distribution of the range (either its expectation or the probability of observing a large value of R_n) is sufficiently small. This approach leads to the SSD criteria proposed in Section 3.1. However, posterior range quantifies only inferential accuracy and even a very small value of R_n does not necessarily imply posterior robustness. In fact, there are situations in which to claim that data represent conclusive evidence as the prior varies over Γ, it is necessary to have either a sufficiently large lower bound L_n or a sufficiently small upper bound U_n. A typical example is given by clinical trials where θ represents an effect parameter and the objective of the study is superiority or inferiority respectively. In these cases a pre-experimental analysis of the posterior range should be accompanied by a predictive analysis of posterior lower and upper bounds. This leads to further criteria based on summaries of the predictive distributions of L_n and U_n (see Section 3.2).

Robust Bayesian SSD methods based on the posterior mean have already been formalized in [8] where the robust inference of the normal mean with conjugate priors has been considered. The Author determined explicit formulae for SSD criteria, studied analytically their monotonicity and limiting behavior as n increases and evaluated the role of all the prior inputs (prior means, prior sample sizes, threshold values) on optimal sample sizes. The proposed methods were also used for choosing the size of a clinical trial. Here we develop an analogous methodology for the Bernoulli model with conjugate prior distributions.

2.1 Distinction between analysis and design priors

Before formally defining SSD criteria, it is important to describe a central aspect of the methodology. This approach has been proposed in previous articles, including those by [12], [8], [4], [9].

In Bayesian design problems the prior distribution is used both to obtain $\pi(\cdot;\mathbf{x}_n)$ for posterior analysis and also to induce $m(\cdot;\pi)$ for pre-posterior computations. The double role of π in the SSD-inference process leads to a distinction between these two distributions. Let π_A denote the *analysis prior* used to obtain the posterior distribution of the parameter, and let π_D be the *design prior*, used to define the marginal distribution of the data.

The idea is that, in general, π_A embodies prior opinions while π_D formalizes design expectations. More specifically, the analysis prior expresses prior knowledge on θ that we want to take into account in the posterior. The design prior describes a scenario and it serves to account for uncertainty on possible guessed values for θ in the design stage. Following [12], we may say that π_D arises in a *what if* spirit: *if* we assume that θ is most likely to be in a specific subset of Θ, then *what* sample size is appropriate for reaching conclusive inferential results, i.e., *what* are the consequences, in terms of predictive probability of a posterior quantity of interest?

The distinction between these two prior distributions offers the researcher the chance of

checking the effects of combinations of more or less optimistic analysis priors (prior inputs) with more or less optimistic design priors (design expectations). Of course, the choice of π_A and π_D may be crucial in the resulting sample size.

Furthermore, in this article we suppose that uncertainty on the analysis prior π_A is modeled through a class Γ, while we consider a robust Bayesian approach to SSD for a single fixed design prior π_D.

3 Robust Bayesian SSD criteria

In this section we formally introduce robust Bayesian SSD criteria for the estimation problem discussed in the previous section, making a distinction between criteria based on the posterior range and criteria based on the lower bound and on the upper bound.

3.1 SSD criteria based on the posterior range

When the posterior quantity of interest is the posterior expectation $\mathbb{E}(\theta \mid \mathbf{X}_n; \pi_A)$ the idea is to use summaries of the predictive distribution of

$$R_n(\mathbf{X}_n) = U_n(\mathbf{X}_n) - L_n(\mathbf{X}_n) = \sup_{\pi_A \in \Gamma} \mathbb{E}(\theta \mid \mathbf{X}_n; \pi_A) - \inf_{\pi_A \in \Gamma} \mathbb{E}(\theta \mid \mathbf{X}_n; \pi_A)$$

as measures of lack of robustness. By the analysis of their behaviour as n increases, we define SSD criteria for pre-experimental control of the posterior range. In the presence of a unique design prior π_D, values of n for controlling robustness of $\mathbb{E}(\theta \mid \mathbf{X}_n; \pi_A)$ with respect to variations of π_A in Γ are chosen by requiring that a summary of the predictive distribution of R_n is sufficiently small. Two predictive summaries and the corresponding SSD criteria are now listed.

- **Expectation Criterion**

  For a given $k > 0$ the optimal sample size is given by

  $$n_E^* = \min\{n \in \mathbb{N} : e_n < k\}$$

  where $e_n = \mathbb{E}_{\pi_D}[R_n(\mathbf{X}_n)]$ is the expected value of the posterior range with respect to the predictive distribution $m(\cdot; \pi_D)$.

- **Tail Probability Criterion**

  For given $r_0 > 0$ and $\epsilon \in (0, 1)$ the optimal sample size is defined as

  $$n_P^* = \min\{n \in \mathbb{N} : p_{n,r_0} < \epsilon\}$$

  where $p_{n,r_0} = \mathbb{P}_{\pi_D}[R_n(\mathbf{X}_n) > r_0]$ and $\mathbb{P}_{\pi_D}$ is the probability measure corresponding to the marginal distribution $m(\cdot; \pi_D)$.

As a general remark, the expectation criterion guarantees only an average control on the distribution of R_n, with no control on its variability. Then, even though the predictive expectation may be seen as a natural measure of lack of robustness, a value of n such that e_n is small does not necessarily protect one from observing large values of the posterior range. Conversely, the use of the tail probability criterion allows a more restrictive control on the distribution of R_n, but it may result highly sensitive to the threshold specification. However, since these two criteria are based on different summaries of the predictive distribution of R_n, in practice it is always possible to use them in an interactive way.

3.2 SSD criteria based on posterior lower/upper bound

We have already remarked that the posterior range only provides an indication of inferential accuracy and it may happen that firm conclusions can be drawn only when either a sufficiently large lower bound or a sufficiently small upper bound is observed. This consideration leads to SSD criteria based on pre-experimental control of posterior bounds given by (1). Formally, for a fixed design prior π_D, let us consider predictive expectations of L_n and U_n

$$e_n^L = \mathbb{E}_{\pi_D}\left[L_n\left(\mathbf{X}_n\right)\right], \qquad e_n^U = \mathbb{E}_{\pi_D}\left[U_n\left(\mathbf{X}_n\right)\right]$$

where, as before, the expected values are with respect to the marginal density $m(\cdot;\pi_D)$. For given real values r_L and r_U we also consider the predictive probabilities of observing a sufficiently large value of L_n and a sufficiently small value of U_n

$$p_{n,r_L}^L = \mathbb{P}_{\pi_D}\left[L_n\left(\mathbf{X}_n\right) > r_L\right] \qquad p_{n,r_U}^U = \mathbb{P}_{\pi_D}\left[U_n\left(\mathbf{X}_n\right) < r_U\right].$$

Depending on the design assumptions, two alternative scenarios are possible.

- **Lower bound criteria**

  In the first case we assume that clear conclusions can be drawn only if L_n is sufficiently large and that e_n^L and p_{n,r_L}^L are both increasing functions of n. Then, for $r_L > 0, k_L > 0$ and $\epsilon_L \in (0,1)$, optimal sample sizes are defined as

$$n_E^{*\,L} = \min\left\{n \in \mathbb{N} : e_n^L > k_L\right\}$$
$$n_P^{*\,L} = \min\left\{n \in \mathbb{N} : p_{n,r_L}^L > \epsilon_L\right\}$$

- **Upper bound criteria**

  In the second scenario we suppose that firm conclusions can be reached only if U_n is small enough and that e_n^U is a (bounded) decreasing function of n while p_{n,r_U}^U is an increasing function. In this case, for $r_U > 0, k_U > 0$ and $\epsilon_U \in (0,1)$, optimal sample sizes are defined as

$$n_E^{*\,U} = \min\left\{n \in \mathbb{N} : e_n^U < k_U\right\}$$
$$n_P^{*\,U} = \min\left\{n \in \mathbb{N} : p_{n,r_U}^U > \epsilon_U\right\}$$

It is worth noting that if one of the criteria for pre-experimental control of posterior bounds is used jointly with one of the criteria based on predictive summaries of R_n, the resulting sample size will be the largest among the two values obtained.

4 Robust estimation of a Bernoulli parameter

We now derive in more detail the application of SSD criteria to robust estimation of a Bernoulli parameter using conjugate beta priors. The Bernoulli model associated with two mutually exclusive outcomes plays an important role in statistics since it best describes all situations in which a trial outcome is either a "success" or a "failure", such as when tossing a coin or when modelling the positive or negative result of a treatment. For a set of independent Bernoulli trials such that the probability of response (the response rate) is θ we obtain the Binomial likelihood. Hence, the Binomial is used as a sampling distribution for empirical counts that occur as proportions like the results of voter polls, clinical trials and many other sampling procedures.

In this context, since Beta distributions form a flexible and mathematically convenient class for quantities constrained to lie between 0 and 1, they are suitable to be used as prior densities for unknown proportions. Moreover, being conjugate to the Binomial family of sampling distributions, Beta densities make the necessary computations straightforward. A further appealing property of conjugate priors is that of allowing one to begin with a certain functional form for the prior and to end up with a posterior of the same form, but with parameters updated by the sample information. This provides a straightforward way of seeing the effect of prior and sample information on posterior inference.

4.1 Conjugate analysis

Let us assume that each X_i has a Bernoulli distribution with unknown parameter θ

$$f(x_i;\theta) = Br(x_i;\theta) \qquad x_i \in \{0,1\},\ \theta \in [0,1]$$

where $Br(\cdot;\theta)$ denotes the probability mass function of a Bernoulli random variable with unknown parameter θ, with $0 < \theta < 1$. Let $Beta(\cdot;a,b)$ denote the density function of a Beta random variable with parameters a and b. Also, recall that the mean and the variance of a Beta random variable are respectively $a(a+b)^{-1}$ and $ab(a+b)^{-2}(a+b+1)^{-1}$.

As a class of analysis priors a standard choice is the class of restricted-conjugate Beta densities, Γ_B, which, for a given value μ_A in $(0,1)$, is defined as

$$\Gamma_B = \left\{ Beta(\theta;\alpha_A,\beta_A) : \mathbb{E}(\theta;\alpha_A,\beta_A) = \mu_A, \mathbb{V}(\theta;\alpha_A,\beta_A) \in \left[v_A^L, v_A^U\right] \right\},$$

where $0 < v_A^L < v_A^U < 1$ are given. By elementary algebra we have

$$\Gamma_B = \left\{ Beta(\theta;\alpha_A,\beta_A) : \alpha_A = \frac{\mu_A}{1-\mu_A}\beta_A, \beta_A \in \left[\beta_A^L, \beta_A^U\right] \right\},$$

where

$$\beta_A^L = (1-\mu_A)\left[\frac{\mu_A(1-\mu_A)}{v_A^U} - 1\right] \quad \text{and} \quad \beta_A^U = (1-\mu_A)\left[\frac{\mu_A(1-\mu_A)}{v_A^L} - 1\right]$$

with $\mu_A \in (0,1)$, $0 < v_A^L < v_A^U < \mu_A(1-\mu_A) < 1$ are given.

In the following, since Beta priors are used to formalize pre-experimental beliefs on a response probability, it may be natural for real experiments to require unimodality. For a $Beta(\cdot;a,b)$ there is a unique mode in $(a-1)(a+b-2)^{-1}$ only if $a > 1$ and $b > 1$. Imposing unimodality leads to the following restrictions on β_A^L and, consequently, on v_A^U

$$\beta_A^L > \max\left\{1, \frac{1-\mu_A}{\mu_A}\right\} = \begin{cases} \frac{1-\mu_A}{\mu_A} & \text{if} \quad \mu_A < 1/2 \\ 1 & \text{if} \quad \mu_A \geq 1/2 \end{cases}$$

$$v_A^U < \min\left\{\frac{(1-\mu_A)^2\mu_A}{2-\mu_A}, \frac{(1-\mu_A){\mu_A}^2}{1+\mu_A}\right\} = \begin{cases} \frac{(1-\mu_A){\mu_A}^2}{1+\mu_A} & \text{if} \quad \mu_A < 1/2 \\ \frac{(1-\mu_A)^2\mu_A}{2-\mu_A} & \text{if} \quad \mu_A \geq 1/2 \end{cases}$$

Finally let us suppose that we set as design prior

$$\pi_D(\theta) = Beta(\theta;\alpha_D,\beta_D)$$

where α_D and β_D are known and fixed and $\alpha_D+\beta_D$ denotes the prior sample size associated to π_D. In this case the marginal distribution of the sufficient statistic $Y_n = \sum_{i=1}^n X_i$ is a Betabinomial distribution with parameters (n,α_D,β_D) denoted by $Bb(y_n;n,\alpha_D,\beta_D)$.

Results for robust estimation of the Bernoulli parameter with conjugate priors are provided in the following subsections.

4.2 Results for a fixed conjugate design prior

From standard conjugate analysis, we know that, under the above assumptions, the posterior expectation of θ is

$$\mathbb{E}(\theta \mid \mathbf{x}_n;\pi_A) \quad = \quad \frac{\alpha_A + \sum_{i=1}^n x_i}{\alpha_A+\beta_A+n} = \frac{(\alpha_A+\beta_A)\mu_A + n\bar{x}_n}{\alpha_A+\beta_A+n}$$

where $\alpha_A+\beta_A$ has the interpretation of prior sample size associated to π_A.

Since we are concerned with the predictive analysis described in Section 3, in the following results we provide explicit expressions for R_n, e_n and p_{n,r_0}.

Result 1. Assume that X_i has probability mass function $Br(x_i;\theta)$, $i = 1, 2, ..., n$ and that $\pi_A \in \Gamma_B$. Then, the range of the posterior mean $\mathbb{E}(\theta \mid \mathbf{x}_n; \pi_A)$ is given by

$$R_n = b_n \mid Y_n - n\mu_A \mid = c_n \mid \bar{X}_n - \mu_A \mid$$

where

$$b_n = \frac{(1-\mu_A)\left(\beta_A^U - \beta_A^L\right)}{\left[\beta_A^L + n(1-\mu_A)\right]\left[\beta_A^U + n(1-\mu_A)\right]} \qquad \text{and} \qquad c_n = n\, b_n$$

Proof. The expression of R_n follows from monotonicity of the posterior mean $\mathbb{E}(\theta \mid \mathbf{x}_n; \pi_A)$ as a function of β_A.

Result 2. Under the assumptions of **Result 1**, using $\pi_D(\theta) = Beta(\theta; \alpha_D, \beta_D)$, R_n has the following properties.

- Predictive expectation:

  $$e_n = \mathbb{E}_{\pi_D}(R_n(\mathbf{X}_n)) = nb_n\left[2\zeta(\mu_A, n) - (\mu_A - \mu_D)\right]$$

  where $\zeta(\mu_A, n) = \sum_{y_n=0}^{n\mu_A}\left(\mu_A - \frac{y_n}{n}\right) p_{BBb,n,\alpha_D,\beta_D}(y_n)$ and $\mu_D = \frac{\alpha_D}{\alpha_D+\beta_D} = \mathbb{E}(\theta; \alpha_D, \beta_D)$.

- Predictive tail probability:
  given $r_0 > 0$:

  $$p_{n,r_0} = \mathbb{P}_{\pi_D}\left[R_n(\mathbf{X}_n) > r_0\right] = 1 - F_n^D\left(n\mu_A + \frac{r_0}{b_n}\right) + F_n^D\left(n\mu_A - \frac{r_0}{b_n}\right) - p_n^D\left(n\mu_A - \frac{r_0}{b_n}\right)$$

  where F_n^D and p_n^D denote the cumulative distribution function and the probability mass function of a Betabinomial random variable with parameters n, α_D and β_D.

Proof. The expressions of e_n and p_{n,r_0} follow from standard probability calculations, using the explicit form of R_n derived in **Result 1**.

The behaviour of e_n and p_{n,r_0} as n goes to infinity is described in the following corollary.

Corollary 1. Under the assumptions of **Result 1** and **Result 2**, as n goes to ∞, the following results hold

(a) The sequence $(e_n; n \in \mathcal{N})$ converges to 0 at the rate of n^{-1};

(b) The sequence of random variables $(R_n; n \in \mathcal{N})$ converges in law to 0.

Proof. Part (a) is obtained noting that as n diverges $n\left[2\zeta(\mu_A, n) - (\mu_A - \mu_D)\right] = O(n)$ and $b_n = O\left(n^{-2}\right)$.

For part (b) note that as n goes to ∞ $n\mu_A + \frac{r_0}{b_n} = O\left(n^2\right)$ and $n\mu_A - \frac{r_0}{b_n} = O\left(-n^2\right)$.

The result follows from

$$p_{n,r_0} \;=\; 1 - F_n^D\left(O\left(n^2\right)\right) + F_n^D\left(O\left(-n^2\right)\right) - p_n^D\left(O\left(-n^2\right)\right) \to 0$$

Remarks

1. It is straightforward to check that, for any finite and positive value of β_A^L and β_A^U,

$$0 < c_n\left(\beta_A^L, \beta_A^U\right) < 1$$

where we have explicitly stressed the dependence of the quantity c_n on β_A^L and β_A^U.

Now, letting $\beta_A^L \to 0$ $\beta_A^U \to +\infty$ in the definition of Γ_B we retrieve the class of Beta distributions with no restrictions on the variance, i.e.

$$\Gamma_{UB} = \{Beta\left(\theta; \alpha_A, \beta_A\right) : \mathbb{E}\left(\theta; \alpha_A, \beta_A\right) = \mu_A\}$$

where the subscript stands for *unrestricted Beta*.

In this case we have

$$\lim_{\beta_A^L \to 0\, \beta_A^U \to +\infty} c_n\left(\beta_A^L, \beta_A^U\right) = 1$$

and, consequently, we obtain $R_n \to \mid \bar{X}_n - \mu_A \mid$ and $e_n \to 2\zeta\left(\mu_A, n\right) - \left(\mu_A - \mu_D\right)$. This suggests an interpretation of c_n as a measure of the shrinkage of R_n and e_n due to the restriction of the class of priors from to Γ_B to Γ_{UB}.

2. **Corollary 1** establishes the convergence to 0 of both e_n and p_{n,r_0} as n diverges. However, these quantities are not monotonically decreasing functions of n for any arbitrary choice of prior inputs. For instance, it may be noted that e_n, as a function of the sample size, can have a maximum at a value strictly larger than 1. First of all let us notice that R_n, as a function of n has a maximum at

$$n_A^* = \frac{\sqrt{\beta_A^L \beta_A^U}}{1 - \mu_A} = \frac{\bar{\beta}_A}{1 - \mu_A} = \alpha_A\left(\bar{\beta}_A\right) + \bar{\beta}_A$$

where $\alpha_A\left(\beta_A\right) = \frac{\mu_A}{1-\mu_A}\beta_A$. This means that as long as $n < n_A^*$, R_n increases with n; as $n > n_A^*$, R_n decreases with n. This behaviour can be interpreted recalling that the posterior mean is a weighted average of μ_A and $\bar{x}_n$ with weights $\alpha_A + \beta_A$ and n. When n is small, the prior information is dominant and the weight of the prior mean μ_A in is larger than the weight of the sample mean $\bar{x}_n$. As the sample size increases the weight of $\bar{x}_n$ also increases, but as long as $n < n_A^*$ the posterior range is increasing with n. When experimental information becomes dominant, $n > n_A^*$, and $\mathbb{E}\left(\theta \mid \mathbf{x}_n; \beta_A\right)$ is closer to $\bar{x}_n$ than to μ_A, R_n starts to decrease monotonically with n.

4.3 Lower and upper bounds

We here derive results for the lower and upper bounds that parallel those obtained for the posterior range. We shall use the explicit expressions of their predictive expectations and tail probabilities in Sections 5 and 6, where numerical results are provided.

Result 3. Under the same assumptions of **Result 1** and **Result 2**, expectations and tail probabilities of L_n and U_n are as follows:

- Predictive expectations

$$e_n^L = \mathbb{E}_{\pi_D}\left(L_n(\boldsymbol{X}_n)\right) = \alpha_n^U - c_n\zeta\left(\mu_A, n\right)$$
$$e_n^U = \mathbb{E}_{\pi_D}\left(U_n(\boldsymbol{X}_n)\right) = \alpha_n^L + c_n\zeta\left(\mu_A, n\right)$$

where

$$\alpha_n^U = \frac{\beta_A^U\mu_A + n\mu_D\left(1 - \mu_A\right)}{\beta_A^U + n\left(1 - \mu_A\right)} \qquad \alpha_n^L = \frac{\beta_A^L\mu_A + n\mu_D\left(1 - \mu_A\right)}{\beta_A^L + n\left(1 - \mu_A\right)}$$

- Predictive tail probabilities: for given r_L, $r_U \in (0,1)$

$$\begin{aligned} p^L_{n,r_L} &= \mathbb{P}_{\pi_D}\left(L_n(\boldsymbol{X}_n) > r_L\right) = 1 - F^D_n\left(\min\left\{\frac{r_L\left(\beta^L_A + n\left(1-\mu_A\right)\right) - \beta^L_A\mu_A}{1-\mu_A}, n\mu_A\right\}\right) \\ &- \left[F^D_n\left(\frac{r_L\left(\beta^U_A + n\left(1-\mu_A\right)\right) - \beta^U_A\mu_A}{1-\mu_A}\right) - F^D_n\left(n\mu_A\right)\right] I_{(0,r_L)}\left(\mu_A\right) \end{aligned}$$

$$\begin{aligned} p^U_{n,r_U} &= \mathbb{P}_{\pi_D}\left(U_n(\boldsymbol{X}_n) \leq r_U\right) \\ &= F^D_n\left(\min\left\{\frac{r_U\left(\beta^U_A + n\left(1-\mu_A\right)\right) - \beta^U_A\mu_A}{1-\mu_A}, n\mu_A\right\}\right) \\ &+ \left[F^D_n\left(\frac{r_U\left(\beta^L_A + n\left(1-\mu_A\right)\right) - \beta^L_A\mu_A}{1-\mu_A}\right) - F^D_n\left(n\mu_A\right)\right] I_{0,r_U}\left(\mu_A\right) \end{aligned}$$

where $I_{(0,r_L)}(.)$ and $I_{0,r_U}(.)$ are the indicator functions of the sets $(0,r_L)$ and $(0,r_U)$.

Proof. Parts a and b follow from standard probability calculations noting that

$$L_n = \begin{cases} E\left(\theta \mid \mathbf{X}_n; \beta^L_A\right) & \text{se } Y_n \leq n\mu_A \\ E\left(\theta \mid \mathbf{X}_n; \beta^U_A\right) & \text{se } Y_n > n\mu_A \end{cases} \qquad U_n = \begin{cases} E\left(\theta \mid \mathbf{X}_n; \beta^U_A\right) & \text{se } Y_n \leq n\mu_A \\ E\left(\theta \mid \mathbf{X}_n; \beta^L_A\right) & \text{se } Y_n > n\mu_A \end{cases}.$$

Remarks

1. It is straightforward to check that, as expected, $e_n = e^U_n - e^L_n$.
2. Note that α^j_n $(j = L, U)$ may be rewritten as follows

$$\begin{aligned} \alpha^j_n &= \frac{\beta^j_A\mu_A + n\mu_D\left(1-\mu_A\right)}{\beta^j_A + n\left(1-\mu_A\right)} = \frac{\beta^j_A\mu_A + \left(1-\mu_A\right)n\mu_D}{\beta^j_A + \left(1-\mu_A\right)n} \\ &= \frac{\frac{\beta^j_A\mu_A}{1-\mu_A} + n\mu_D}{\frac{\beta^j_A}{1-\mu_A} + n} = \frac{\left(\alpha^j_A + \beta^j_A\right)\mu_A + n\mu_D}{\alpha^j_A + \beta^j_A + n} \end{aligned}$$

where

$$\alpha^j_A = \alpha^j_A(\beta^j_A) = \frac{\mu_A}{1-\mu_A}\beta^j_A, \qquad j = L, U$$

Hence, e^L_n and e^U_n are both weighted averages of μ_A and μ_D decreased and increased of the quantity $c_n\zeta\left(\mu_A, n\right)$; if $\mu_A = \mu_D$, then

$$e^L_n = \mu_D - c_n\zeta\left(\mu_D, n\right) \qquad e^U_n = \mu_D + c_n\zeta\left(\mu_A, n\right)$$

3. As n goes to infinity, both e^L_n and e^U_n converge to μ_D at the rate of n^{-1}, given that $\lim_{n\to+\infty}\alpha^j_n = \mu_D$, for $j = L, U$, and that, recalling **Corollary 1** of Section 4.2, $c_n\zeta\left(\mu_A, n\right) = O\left(n^{-1}\right)$. Of course, these limiting values have to be taken into account when fixing threshold values (here, k_L and/or k_U) so that the sample size problem ($e^L_n > k_L$ and/or $e^U_n < k_U$) is actually solvable.

5 Numerical examples

We are now interested in studying the influence of prior parameters on the optimal robust sample sizes, n_E^* and n_P^*. Specifically, we want to establish how and to what extent these quantities are influenced by the amplitude of Γ_B and by the analysis prior mean μ_A. The following numerical examples show optimal sample sizes n_E^* and n_P^* computed for several choices of $[\beta_A^L, \beta_A^U]$ and for several values of μ_A under the assumption that $\pi_D(\theta) = Beta(\theta; 2, 2)$.

5.1 Example 1: Variations of $\left[\beta_A^L, \beta_A^U\right]$

Assume, for instance, that $k = 0.1$. As expected, the wider the class Γ_B (i.e. the difference $\beta_A^U - \beta_A^L$), the larger the corresponding optimal sample size. This fact can be appreciated in Figure 1, which shows e_n as a function of n for a fixed value of $\mu_A = 0.5$ and for several choices of β_A^L and β_A^U. Table 1 exemplifies numerically the decrease in the required sample size implied by the restriction of the interval $\left[\beta_A^L, \beta_A^U\right]$. It is also interesting to observe in the graphs the property mentioned in the second remark of Section 4.2: e_n is not a monotonic function of the sample size, but it attains a maximum at a value larger than 1, then it starts decreasing definitely. Of course this behaviour has to be taken into account when applying the SSD criterion.

As shown in Figure 2 similar comments hold for the tail probabilities $p_{n,0.1}$; the corresponding optimal sample sizes are recorded in the last column of Table 1.

β_A^L	β_A^U	n_E^*	n_P^*
1	50	84	200
2	50	76	190
1	40	67	159
2	40	58	149
1	30	49	118
2	30	41	108
1	20	32	78
2	20	22	67

Table 1: Example 1. Optimal sample sizes n_E^* and $n_P^*$for several values of β_A^L and β_A^U, given $\alpha_D = \beta_D = 2, \mu_A = \mu_D = 0.5, k = 0.1, r_0 = 0.1, \epsilon = 0.2$.

5.2 Example 2: Variations of μ_A

Let us now assume that for instance $\beta_A^L = 2$ and $\beta_A^U = 30$. To illustrate the sensitivity of n_E^* and n_P^* with respect to the difference between the analysis prior mean μ_A and the design prior mean μ_D (here, for instance $\mu_D = 0.5$) we compute the required optimal sample sizes implied by the robust approach for several values of μ_A.

Let us first consider the expectation criterion. If $\mu_A = \mu_D = 0.5$, the optimal sample size is $n_E^* = 41$. When $\mu_A > \mu_D$, the values of n_E^* tend to increase as the difference $\mu_A - \mu_D$ increases. Conversely, when $\mu_A < \mu_D$, n_E^* first decreases and then starts to increase with the difference $\mu_D - \mu_A$. This fact is numerically exemplified by Table 2. The corresponding plots of e_n are shown in Figures 3 and 4. Hence, one can notice that n_E^* does not depend only on the difference $\mid \mu_A - \mu_D \mid$; moreover, for any value of $\mid \mu_A - \mu_D \mid$, optimal sample sizes determined when $\mu_A > \mu_D$ are always larger than those required if $\mu_A < \mu_D$. This behaviour is essentially due to two different factors that affect the value of the posterior range. To illustrate the first factor's effect, it is useful to rewrite the class of analysis priors Γ_B as follows

$$\Gamma_B = \left\{ Beta(\theta; \alpha_A, \beta_A) : \mathbb{E}(\theta; \alpha_A, \beta_A) = \mu_A; \alpha_A \in \left[\frac{\mu_A}{1-\mu_A}\beta_A^L, \frac{\mu_A}{1-\mu_A}\beta_A^U \right], \beta_A \in \left[\beta_A^L, \beta_A^U\right] \right\}.$$

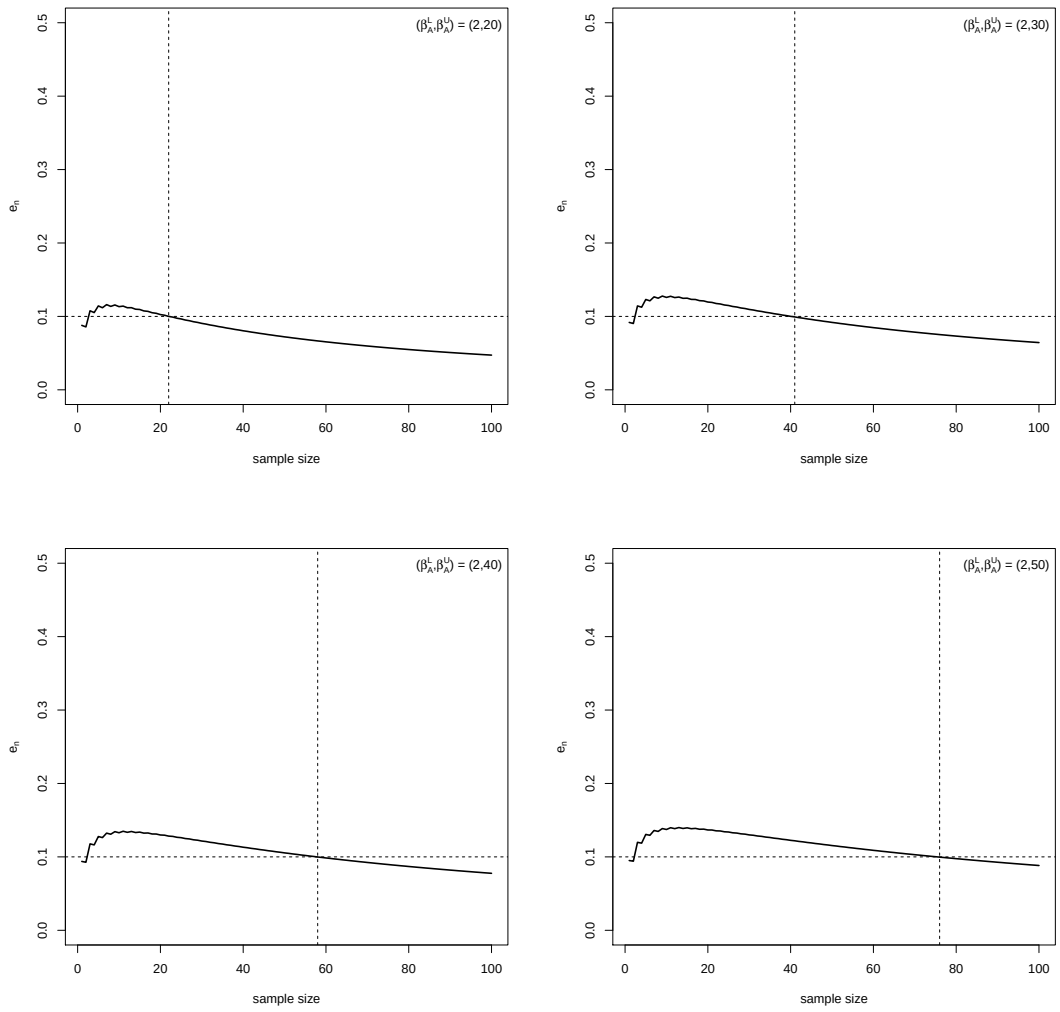

Figure 1: Example 1. Predictive expectation e_n with respect to the sample size for increasingly wide classes of Beta priors, i.e. $(\beta_A^L, \beta_A^U) = (2, 20), (2, 30), (2, 40), (2, 50)$, given $\alpha_D = \beta_D = 2, \mu_A = \mu_D = 0.5, k = 0.1$.

It can be easily checked that, for any specific choice of β_A^L and β_A^U, the length of the interval

$$\left[\frac{\mu_A}{1-\mu_A}\beta_A^L, \frac{\mu_A}{1-\mu_A}\beta_A^U\right]$$

is increasing with μ_A. This means that as μ_A decreases, the impact of α_A on the form of the analysis priors becomes smaller and smaller. Hence, for given values of β_A^L and β_A^U the restriction of the set of possible values for α_A leads to "similar" distributions in the class Γ_B, thus reducing the value of the posterior range. Conversely, the second factor, whose effect depends only on the size of the difference $\mid \mu_A - \mu_D \mid$, is immediately evident if we consider the expression of the posterior range $R_n = c_n \mid \bar{X}_n - \mu_A \mid$ where $\mathbb{E}_{\pi_D}\left(\bar{X}_n\right) = \mathbb{E}_{\pi_D}\left(\frac{Y_n}{n}\right) = \mu_D$. **Example 2** shows that when $\mu_A < \mu_D$ the impact of the first factor tends to be dominant only if μ_A is sufficiently close to μ_D ($\mu_A = 0.475$ o $\mu_A = 0.450$). The two effects are equal for $\mu_A = 0.4$: in this case we have the same sample size obtained when $\mu_A = \mu_D = 0.5$.

Similar considerations can be extended to the tail probability criterion. Figures 5 and 6 plot $p_{n,0.1}$ computed for different values of μ_A in the context of **Example 2**. Table 2 also reports the corresponding values of n_P^*.

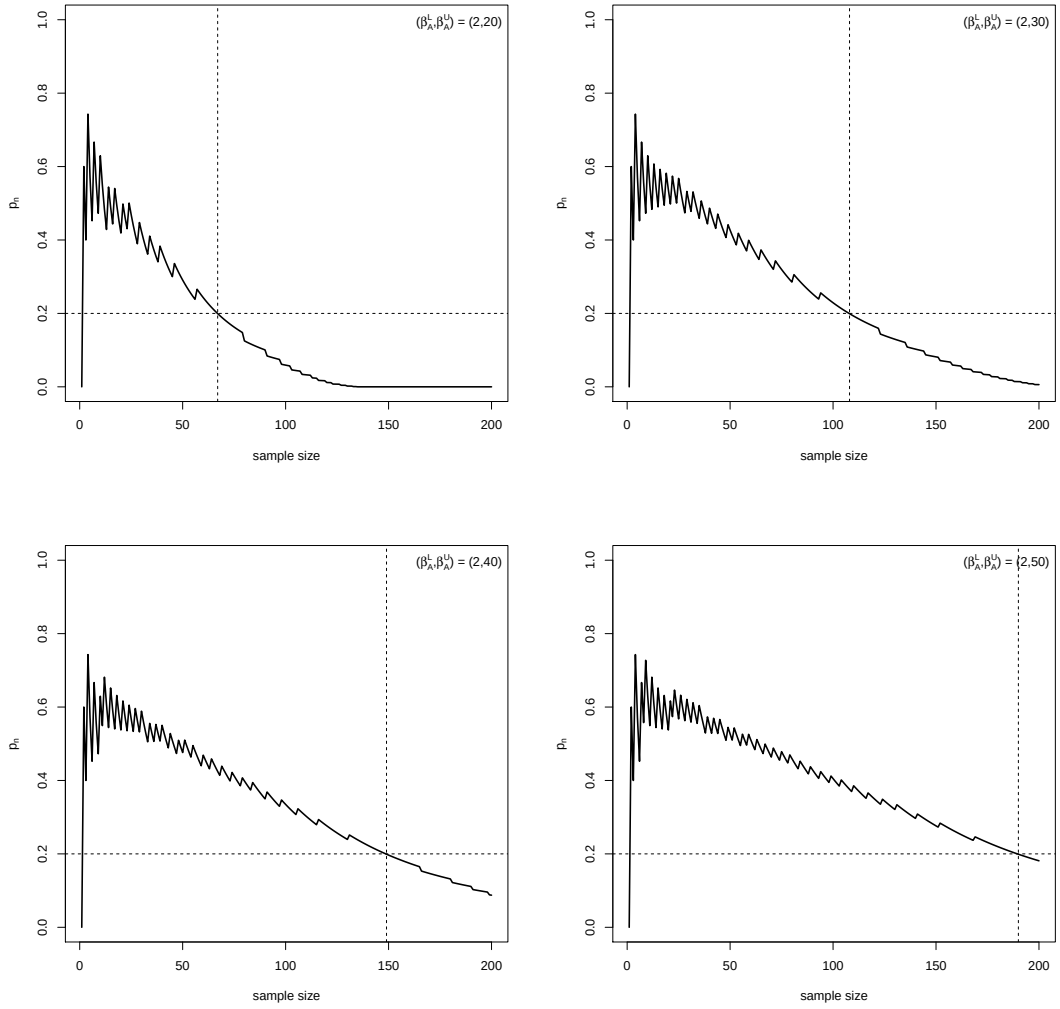

Figure 2: Example 1. Predictive probability $p_{n,0.1}$ with respect to the sample size for increasingly wide classes of Beta priors, i.e. $(\beta_A^L, \beta_A^U) = (2, 20), (2, 30), (2, 40), (2, 50)$, given $\alpha_D = \beta_D = 2, \mu_A = \mu_D = 0.5, \epsilon = 0.2$.

	μ_A	n_E^*	n_P^*
$\mu_A > \mu_D$	0.7	120	279
	0.6	60	150
	0.55	47	121
$\mu_A = \mu_D$	**0.5**	**41**	108
$\mu_A < \mu_D$	0.475	40	104
	0.45	39	101
	0.4	**41**	99
	0.3	53	119
	0.2	70	140
	0.1	89	157

Table 2: Example 1. Optimal sample sizes n_E^* and n_P^* for different values of μ_A given $(\beta_A^L, \beta_A^U) = (2, 30), \alpha_D = \beta_D = 2, \mu_D = 0.5, k = 0.1$.

5.3 Comparisons

It is interesting to compare sample sizes obtained using the robust methods of the previous section with more traditional approaches based on the use of either a conjugate beta prior or

$$\mu_A < \mu_D$$

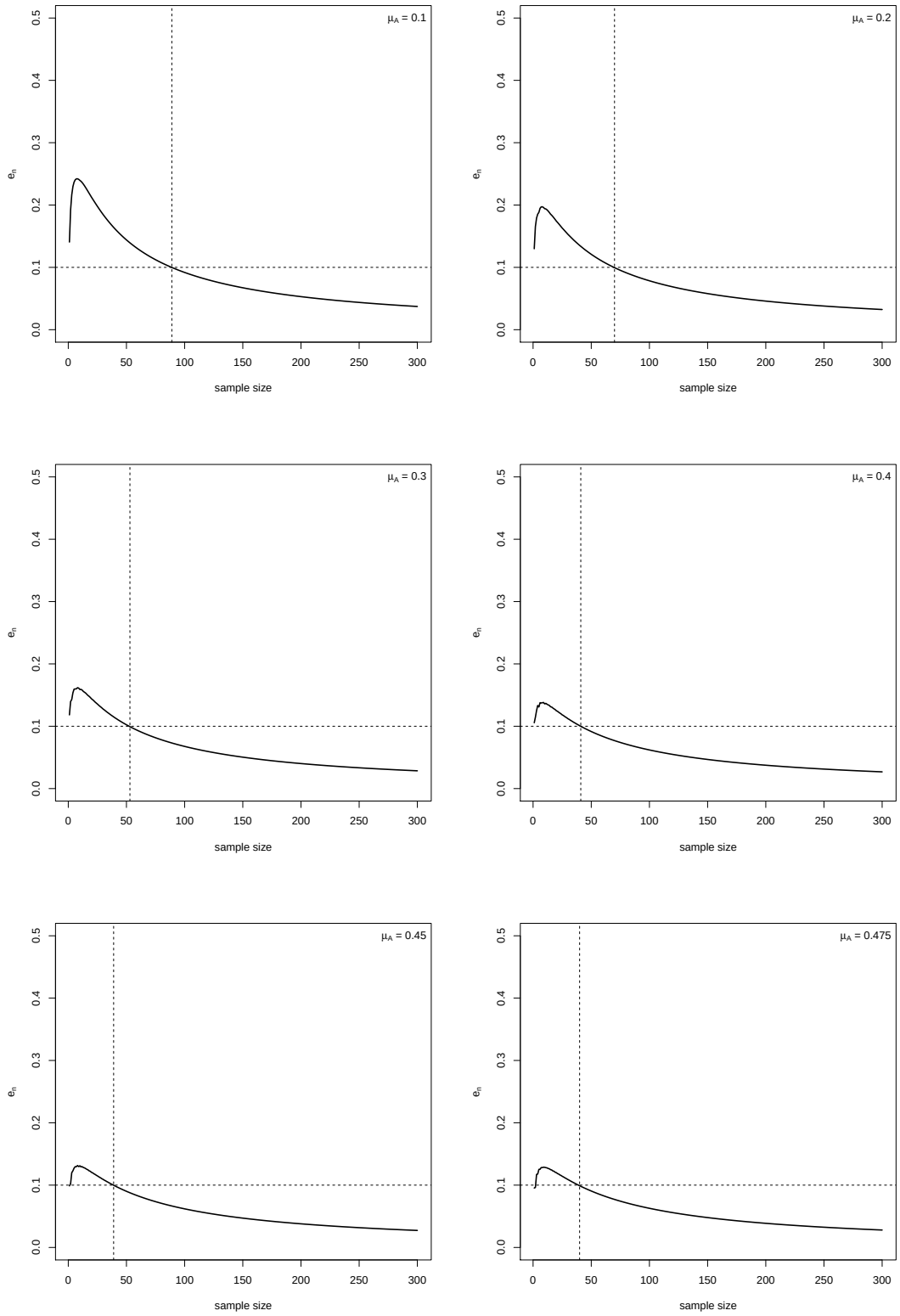

Figure 3: Example 2. Predictive expectation e_n with respect to the sample size for decreasing values of μ_A, given $(\beta_A^L, \beta_A^U) = (2, 30), \alpha_D = \beta_D = 2, \mu_D = 0.5, k = 0.1$.

$$\mu_A \geq \mu_D$$

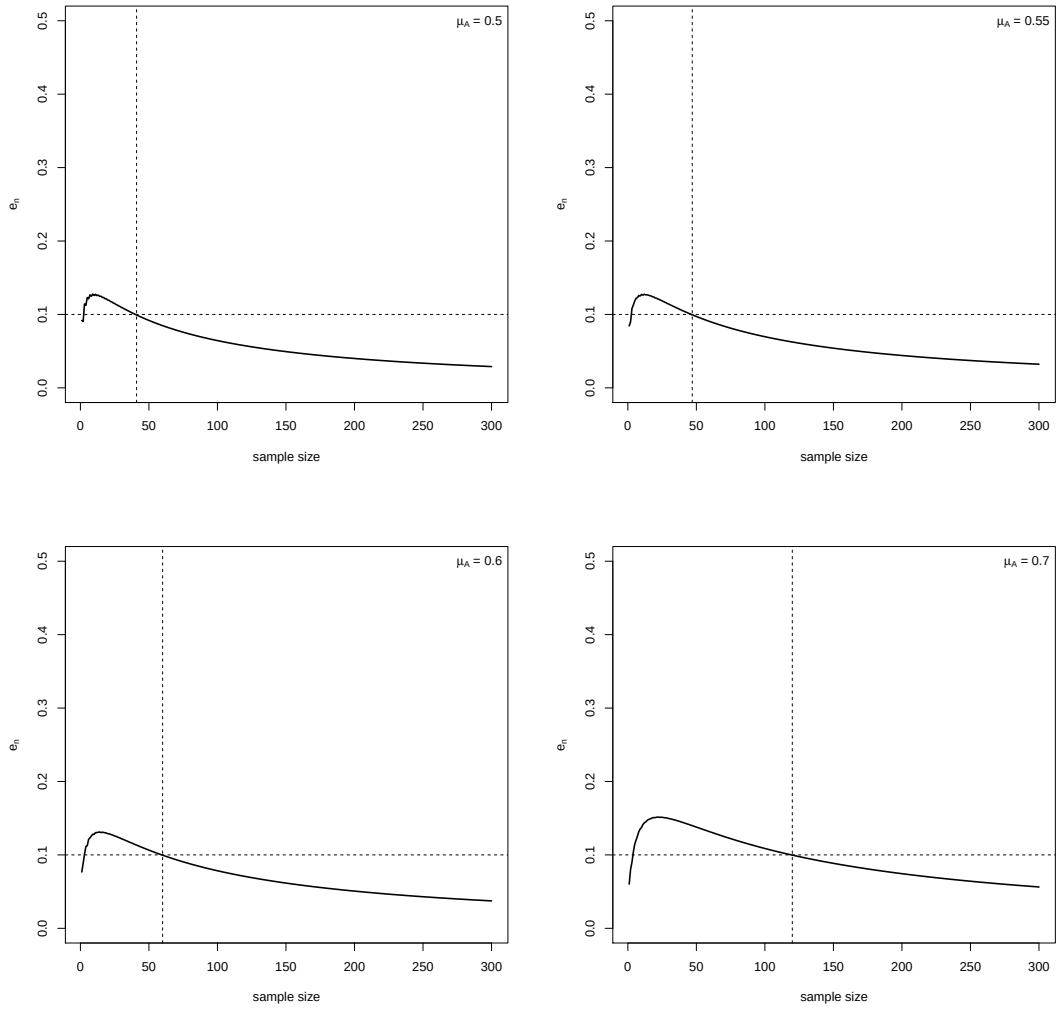

Figure 4: Example 2. Predictive expectation e_n with respect to the sample size for decreasing values of μ_A, given $(\beta_A^L, \beta_A^U) = (2, 30), \alpha_D = \beta_D = 2, \mu_D = 0.5, k = 0.1)$.

$\mu_A < \mu_D$

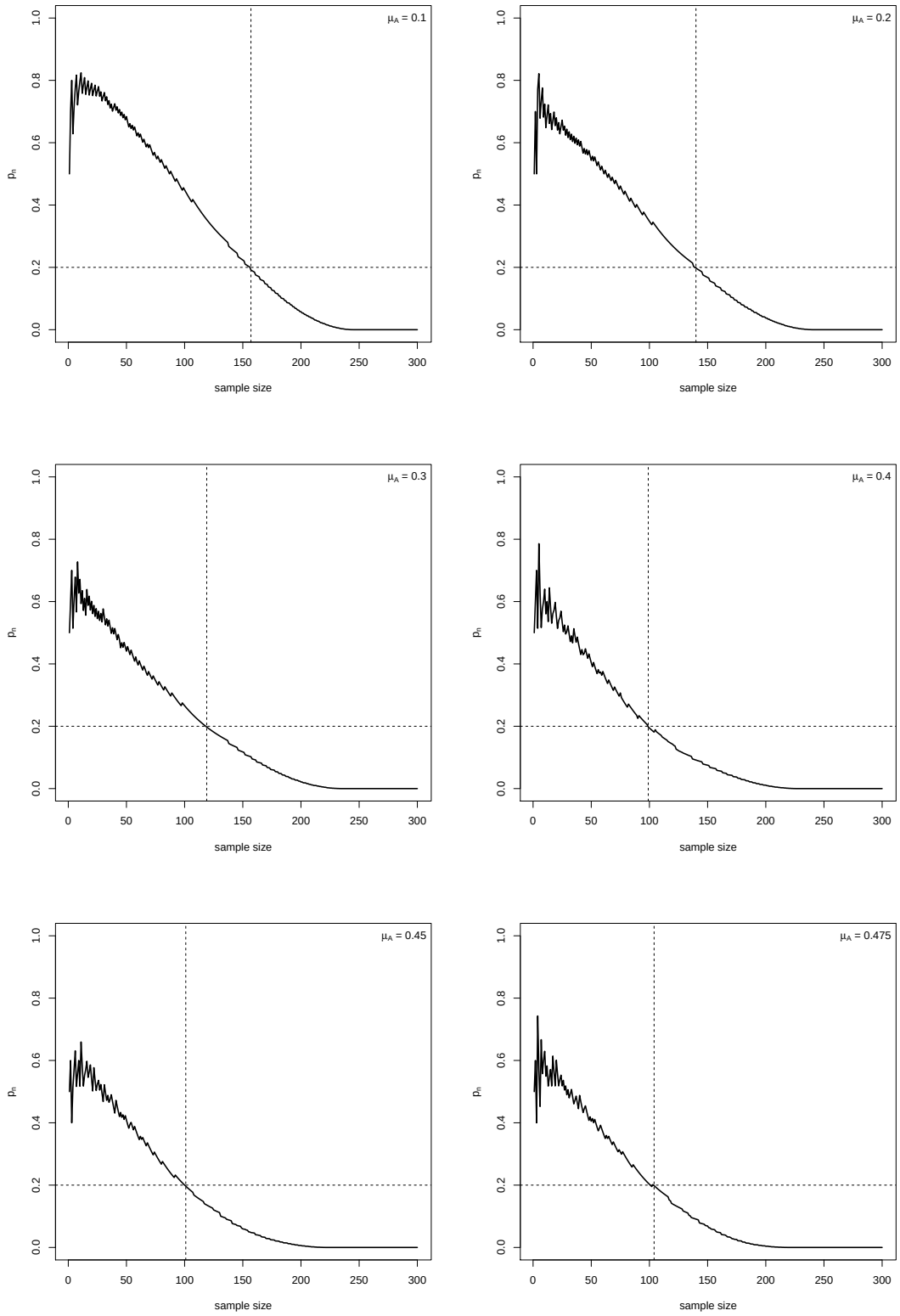

Figure 5: Example 2. Predictive probability $p_{n,0.1}$ with respect to the sample size for decreasing values of μ_A, given $(\beta_A^L, \beta_A^U) = (2, 30), \alpha_D = \beta_D = 2, \mu_D = 0.5, \epsilon = 0.2$.

$\mu_A \geq \mu_D$

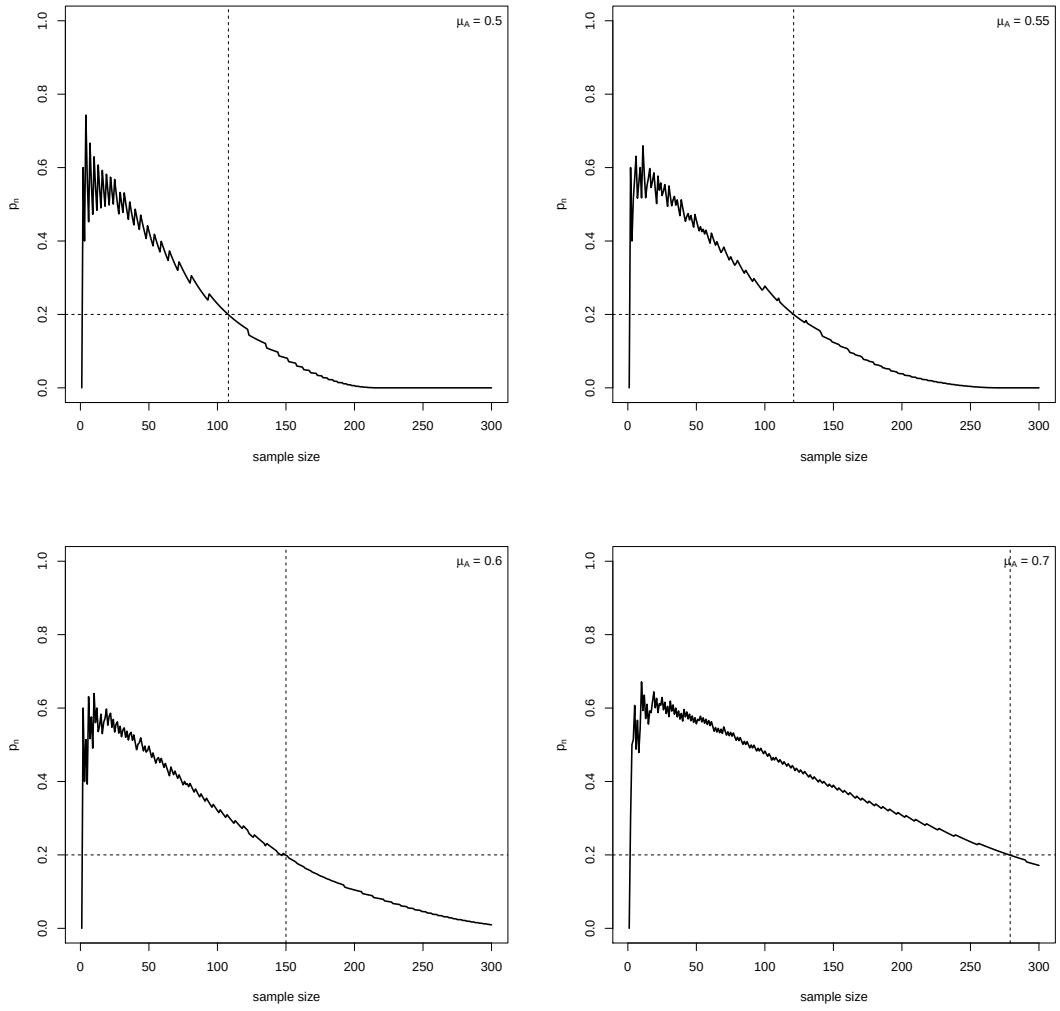

Figure 6: Example 2. Predictive probability $p_{n,0.1}$ with respect to the sample size for increasing values of μ_A, given $(\beta_A^L, \beta_A^U) = (2, 30), \alpha_D = \beta_D = 2, \mu_D = 0.5, \epsilon = 0.2$.

of a noninformative prior. In fact, even thought there are not standard versions of SSD criteria based on summaries of the posterior range, we may compare the methods based on upper and lower bounds with the corresponding non-robust approaches based on a single analysis prior and study the effect of using a class of priors rather than a single distribution.

Single prior distribution. Let us consider a specific prior π_0 for θ and let $e_n(\pi_0)$ denote the predictive expectation of the posterior mean determined using π_0. For any generic class of analysis prior distributions which includes π_0, $e_n^L \leq e_n(\pi_0) \leq e_n^U$. Hence, sample sizes needed to have $e_n^L > k_L$ (or $e_n^U < k_U$) are typically larger than those needed to have $e_n(\pi_0) > k_L$ (or $e_n(\pi_0) < k_U$) and it may be useful to quantify the increase in the required sample size implied by the robust approach. In the specific context of this section, when $\pi_0 = Beta(\theta; \alpha_0, \beta_0) \in \Gamma_B$, we obtain

$$e_n(\pi_0) = \mathbb{E}_{\pi_D}(\mathbb{E}(\theta \mid \mathbf{X}_n; \pi_0)) = \frac{\beta_0 \mu_A + (1-\mu_A) n \mu_D}{\beta_0 + n(1-\mu_A)},$$

so that

$$e_n(\pi_0) > k_L \quad \text{if and only if} \quad n > \frac{k_L - \mu_A}{\mu_D - k_L} \frac{\beta_0}{1-\mu_A} = \frac{k_L - \mu_A}{\mu_D - k_L} n_0$$

and

$$e_n(\pi_0) < k_U \quad \text{if and only if} \quad n < \frac{k_U - \mu_A}{\mu_D - k_U} \frac{\beta_0}{1-\mu_A} = \frac{k_U - \mu_A}{\mu_D - k_U} n_0.$$

Consequently, we have the following relationships between the robust and non robust optimal sample sizes

$$n_E^*(\pi_0) \leq n_E^{*\,L}(\Gamma_B) \quad \text{and} \quad n_E^*(\pi_0) \leq n_E^{*\,U}(\Gamma_B),$$

where we have stressed in the notation the dependence on the single prior or on the class of priors.

Let us first analyse the expectation criterion for the lower bound of the posterior mean assuming that $\mu_A \leq \mu_D$, $\mu_D > k_L$, which is consistent with the first scenario depicted in Section 3.2. We want to determine the smallest value of n such that $e_n^L > k_L$. Table 3 (columns a, b, c) reports optimal robust sample sizes obtained assuming $\pi_D(\theta) = Beta(\theta; 8, 2)$ (i.e., $\mu_D = 0.80$, $n_D = 10$), and $k_L = 0.75$ for several values of prior inputs. As expected, one can notice what follows:

- for any value of μ_A, the larger the difference $\beta_A^U - \beta_A^L$, the larger the corresponding value $n_E^{*\,L}$;
- for any specific choice of $[\beta_A^L, \beta_A^U]$, the larger μ_A, the smaller $n_E^{*\,L}$.

Table 3 (columns d, e, f) also shows the minimal sample sizes selected by requiring $e_n(\pi_0) > k_L = 0.75$ and allows us to quantify the number of extra observations implied by the robust approach with respect to the use of a single prior. As expected, for any value of μ_A and for any choice of $[\beta_A^L, \beta_A^U]$, $n_E^*(\pi_0) \leq n_E^{*\,L}(\Gamma_B)$. Moreover, for any specific choice of β_0, as the value of μ_A approaches k_L, $n_E^*(\pi_0)$ tends to decrease. Note that $n_E^*(\pi_0) < 0$ if $\mu_A > k_L$. On the other hand, for any value of μ_A, $n_E^*(\pi_0)$ is increasing with β_0.

Non informative prior distribution. Let us now consider the uniform density $\pi_U = Beta(\theta; 1, 1)$ that corresponds to assigning a weight $n_U = \alpha_U + \beta_U = 2$ to the analysis prior mean $\mu_A = 0.5$. It can be checked that

$$e_n(\pi_U) = \frac{1 + n\mu_D}{2+n} > k_L \quad \text{if and only if} \quad n > \frac{2k_L - 1}{\mu_D - k_L}$$

Note that the optimal sample size $n_E^*(\pi_U)$ is a decreasing function of the difference $\mu_D - k_L$ and diverges as $\mu_D - k_L$ becomes smaller and smaller.

μ_A	(a) $\beta_A^L = 5$ $\beta_A^U = 7$	(b) $\beta_A^L = 4$ $\beta_A^U = 8$	(c) $\beta_A^L = 3$ $\beta_A^U = 9$	(d) $\beta_0 = 5$	(e) $\beta_0 = 6$	(f) $\beta_0 = 7$	(g) $\alpha_U = 1$ $\beta_U = 1$
0.20	97	111	124	69	83	97	-
0.25	94	107	121	67	80	94	-
0.30	91	103	116	65	78	90	-
0.35	87	99	111	62	74	87	-
0.40	82	94	106	59	70	82	-
0.45	77	88	99	55	66	77	-
0.50	71	81	91	50	60	70	10
0.55	63	72	81	45	54	63	-
0.60	54	62	70	38	45	53	-
0.65	42	49	56	29	35	40	-
0.70	26	32	38	17	20	24	-
0.75	2	5	12	1	1	1	-

Table 3: Optimal sample sizes, given $\mu_D = 0.8$, $n_D = 10$, $k_L = 0.75$, comparing the criterion based on the predictive expectation of the lower bound, the corresponding single-prior criterion, both with an informative and a non informative prior density.

We now compare these sample sizes, based on π_U, to those determined using $\pi_0 \in \Gamma_B$, which assigns weight $n_0 > 2$ to μ_A. Recall that in this paper we only considered unimodal densities and that the assumption of unimodality implies that prior sample sizes, which express the degree of uncertainty assigned to prior means, are strictly larger than 2. In general, it depends on the value of μ_A whether $n_E^*(\pi_0) < n_E^*(\pi_U)$ or $n_E^*(\pi_0) > n_E^*(\pi_U)$ Intuitively, since we are assuming that $\mu_A \leq \mu_D$, $\mu_D > k_L$, we expect that for large values of μ_A, it is convenient to assign as much weight as possible to μ_A, i.e. we expect $n_E^*(\pi_0) < n_E^*(\pi_U)$. Conversely, for small values of μ_A we expect that $n_E^*(\pi_U) < n_E^*(\pi_0)$. As expected, it can be easily checked that

$$n_E^*(\pi_0) < n_E^*(\pi_U) \quad \text{if} \quad \mu_A > \frac{(2k_L - 1) - \beta_0 k_L}{(2k_L - 1) - \beta_0}$$

Summarizing we have that, for sufficiently small values of μ_A

$$n_E^*(\pi_U) \leq n_E^*(\pi_0) \leq n_E^*(\Gamma_B),$$

whereas for sufficiently large values of μ_A

$$n_E^*(\pi_0) \leq n_E^*(\Gamma_B) \leq n_E^*(\pi_U).$$

The above considerations are numerically exemplified in Table 3.

From the opposite point of view, one can consider the expectation criterion for the upper bound of the posterior mean, assuming a scenario in which $\mu_A \geq \mu_D, \mu_D < k_U$. Results in terms of optimal sample sizes are reported in Table 4. It is straightforward to see that the sample size increases with the value of the analysis prior mean and, for each given value of μ_A, decreases when the class of priors is smaller and smaller, until it collapses in a single prior distribution.

μ_A	(a) $\beta_A^L = 5$ $\beta_A^U = 7$	(b) $\beta_A^L = 4$ $\beta_A^U = 8$	(c) $\beta_A^L = 3$ $\beta_A^U = 9$	(d) $\beta_0 = 5$	(e) $\beta_0 = 6$	(f) $\beta_0 = 7$	(g) $\alpha_U = 1$ $\beta_U = 1$
0.25	2	4	6	1	1	1	-
0.30	12	14	17	7	8	10	-
0.35	23	27	31	15	18	21	-
0.40	36	41	47	25	30	35	-
0.45	52	59	67	36	43	50	-
0.50	71	81	91	50	60	70	70
0.55	94	107	121	66	80	93	-
0.60	123	141	158	87	105	122	-
0.65	161	183	206	114	137	160	-
0.70	211	241	271	150	180	210	-
0.75	281	321	361	200	240	280	-
0.80	386	441	496	275	330	385	-

Table 4: Optimal sample sizes, given $\mu_D = 0.2$, $n_D = 10$, $k_U = 0.25$, comparing the criterion based on the predictive expectation of the upper bound, the corresponding single-prior criterion, both with an informative and a non informative prior density.

6 Application to a clinical trial

Example: ***Drug.*** In the present section we consider the set up of an example described in [11]. Let us suppose a drug has an unknown true response rate $\theta$and that previous experience with similar compounds has suggested that response rates between 0.2 and 0.6 could be feasible, with an expectation around 0.4. Through a sort of 'method of moments', this pre-experimental information is translated into a Beta prior of parameters $(9.2, 13.8)$ (see [11] for details). It is convenient to think of this prior distribution as that which would have arisen had we started with a uniform prior $Beta(\theta; 1, 1)$ and then observed 8.2 successes out of 21 $(= (9.2 - 1) + (13.8 - 1))$ patients. As shown in Figure 7 the $Beta(\theta; 9.2, 13.8)$ density (continuous line) well represents the prior assumptions: here we take this density as a single analysis prior. Moreover we consider a class of restricted conjugate Beta priors with the same prior mean $\mu_A = 0.4$ and with $(\beta_A^L = 5, \beta_A^U = 20)$. The densities corresponding to the lower and upper bound in this class Γ_B, are superimposed in Figure 7 (dotted line and dashed-dotted line respectively), together with the uniform prior (dashed line in the plot) that, as mentioned before, we take as a non-informative prior (see again [11] for discussion on this point).

Let us focus on the sample size determination problem taking into account the different criteria we introduced in this paper. For illustrative purposes we consider a design prior distribution $\pi_D(\theta) = Beta(\theta; 2, 2)$, which is centred on the prior mean $\mu_D = 0.5$. Under this scenario we have $\mu_D > \mu_A$, which refers again to the first scenario of Section 3.2 and actually excludes the criteria involving the upper bound. The study, indeed, aims at showing evidence in favour of a large value of the treatment response rate θ; therefore it is reasonable to focus on the criteria in which a sufficiently large value of the (expected) lower bound is required in the condition for selecting the optimal sample size.

Figure 8 represents the four quantities e_n, $p_{n,r}$, e_n^L and p_{n,r_L}^L as a function of n. In order to make the comparison easier, we have "starting from the end" in the application of the SSD criteria: we have chosen the different parameters involved in the four criteria in such a way that the optimal sample size turns out to be comparable in the four cases. In particular we obtain $n_E^* = 58, n_P^* = 57, n_E^{*,L} = 57, n_P^{*,L} = 53$, setting respectively $k = 0.05$ for the predictive expectation, $r_0 = 0.1$ and $\epsilon = 0.1$ for the predictive probability, $k_L = r_L = 0.45$ and $\epsilon_L = 0.52$. This means that an optimal sample size of 58 allows one to obtain an expected value of the posterior range smaller than 0.05, and simultaneously, to keep as small as 0.1 the predictive probability that this range is larger than 0.1. Similarly the expected lower bound exceeds a

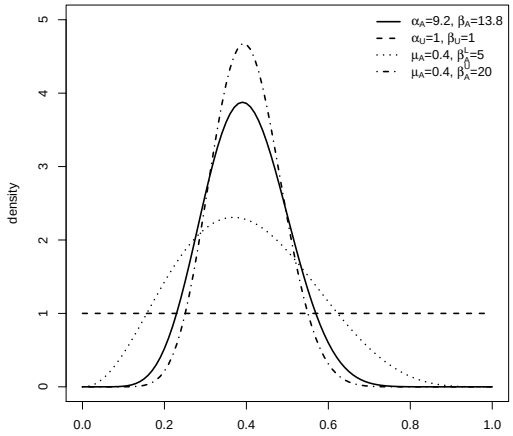

Figure 7: Example: Drug. Analysis prior distributions.

threshold 0.45 for a sample size equal to 57. To obtain a comparable value of the sample size using the predictive probability distribution of the lower bound we need to fix the probability cutoff at a level $\epsilon_L = 0.52$. Moreover, using the lower bound criterion with a single prior $Beta(9.2, 13.8)$, we obtain an optimal sample size equal to 24, which is, as expected, smaller than the one corresponding to the class Γ_B with $(\beta_A^L = 5, \beta_A^U = 20)$.

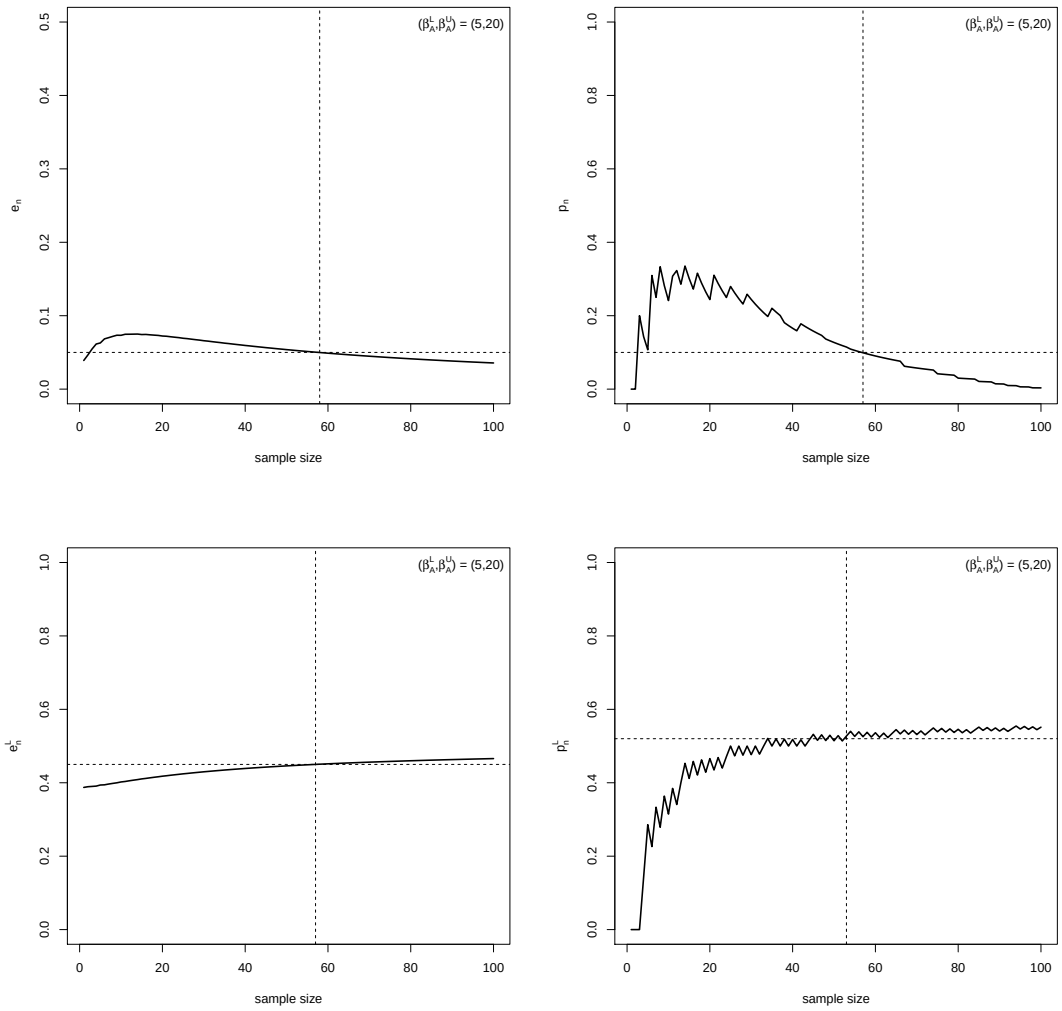

Figure 8: Example: Drug. e_n, $p_{n,r}$, e_n^L and p_{n,r_L}^L with respect to the sample size.

References

[1] C J Adcock. Sample size determination: a review. *Statistician*, 46:261–283, 1997.

[2] J O Berger. *The robust Bayesian viewpoint (with discussion).* Robustness of Bayesian Analysis. J. Kadane, ed. Amsterdam: North-Holland, 1984.

[3] J O Berger. Robust Bayesian analysis: sensitivity to the prior. *The Journal of Statistical Planning and Inference*, (25):303–328, 1990.

[4] P. Brutti and F. De Santis. Avoiding the range of equivalence in clinical trials: Robust Bayesian sample size determination for credible intervals. *The Journal of Statistical Planning and Inference*, 138:1577–1591, 2008.

[5] P. Brutti, F. De Santis, and S. Gubbiotti. Robust Bayesian sample size determination in clinical trials. *Statistics in Medicine*, 27(13):2290–2306, 2008.

[6] K. Chaloner and I. Verdinelli. Bayesian experimental design: a review. *Statistical Science*, 10:237–308, 1995.

[7] A DasGupta and S Mukhopadhyay. Uniform and subuniform posterior robustness: the sample size problem. *The Journal of Statistical Planning and Inference*, 40:189–200, 1994.

[8] F. De Santis. Sample size determination for robust Bayesian analysis. *Journal of the American Statistical Association*, 101(473):278–291, 2006.

[9] R. Etzioni and J. B. Kadane. Optimal experimental design for another's analysis. *Journal of the American Statistical Association*, 88(424):1404–1411, 1993.

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

[11] D. J. Spiegelhalter, K.R. Abrams, and J. P. Myles. *Bayesian approaches to clinical trials and health-care evaluation.* Wiley, 2004.

[12] F. Wang and A. E. Gelfand. A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. *Statistical Science*, 17(2):193–208, 2002.