Bayesian Set Estimation with Alternative Loss Functions: Optimality and Regret Analysis

Fulvio De Santis, Stefania Gubbiotti

Department of Statistical Sciences, Sapienza University of Rome, Rome, Italy
Email: fulvio.desantis@uniroma1.it, stefania.gubbiotti@uniroma1.it

Abstract

Decision-theoretic interval estimation requires the use of loss functions that, typically, take into account the size and the coverage of the sets. We here consider the class of monotone loss functions that, under quite general conditions, guarantee Bayesian optimality of highest posterior probability sets. We focus on three specific families of monotone losses, namely the linear, the exponential and the rational losses whose difference consists in the way the sizes of the sets are penalized. Within the standard yet important set-up of a normal model we propose: 1) an optimality analysis, to compare the solutions yielded by the alternative classes of losses; 2) a regret analysis, to evaluate the additional loss of standard non-optimal intervals of fixed credibility. The article uses an application to a clinical trial as an illustrative example.

Keywords

Bayesian Inference, Decision-Theoretic Approach, Highest Posterior Density Sets, Interval Estimation, Regret

1. Introduction

Together with point estimation and testing, set estimation is the most popular method for inference on the unknown parameter of a statistical model. In the decision-theoretic framework, set estimation has traditionally received less attention than the other two problems, both from a frequentist and a Bayesian point of view. Set estimators are in fact typically derived using non-decisional approaches. For instance, frequentist intervals are found using pivotal quantities or inverting acceptance regions of tests; Bayesian intervals, such as equal tails or highest posterior density sets, are obtained by exploiting some features of the posterior distribution of the parameter, with no explicit connections with a decision-theoretic context. Probably this is the consequence of the difficulty in choos-

ing a satisfactory loss function for this decision problem.

In the few studies on this topic, loss functions typically take into account two aspects of the candidate set estimators, *i.e.* their size (aptly penalized by a specific coefficient) and their ability to contain the true value of the parameter. The most widespread loss function is the so-called linear loss, which is a linear function of the size of the set (see Expression (2) of Section 1). Despite its simplicity, this function has been found to produce optimal sets (both frequentist and Bayesian) with paradoxical flaws even in a very standard problem such as set estimation of the normal mean with unknown variance: see [1] and [2]. In order to address these inconveniences, the Authors have then proposed to consider other losses that are increasing but non-linear functions of the sizes of the sets under evaluation. Among these, the exponential and the rational losses (see Expressions (3) of Section 2).

The analysis contained in the present article is closely related to that in [1] and [2]. Our contribution is intended to fill the following three gaps. 1) The approach of [1] and [2] is mainly theoretical. In the present article we explore the features of optimal actions and posterior expected losses induced by three different loss functions (linear, exponential and rational) in several numerical examples. Specifically, we assess the sensitivity of optimal actions with respect to both the choice of the loss function and of the prior distribution (informative vs non-informative priors). 2) As a second contribution, we propose a regret analysis: we evaluate the additional losses of standard highest posterior density sets of fixed posterior probability with respect to optimal sets under the three losses. 3) As far as we know, a formal decision-theoretic approach set estimation under different loss functions has never been applied to any real experimental context. We here provide an application to clinical trial data.

As we said, the literature on the subject is scant. In addition to the already mentioned motivating articles [1] and [2], general introductions to Bayesian decision-theoretic set estimation can be found, for instance, in [3] [4] [5] [6] [7]. An objective Bayesian decisional approach is proposed in [8]. The use of posterior regret for interval estimation has been lately considered in [9]. Evaluation of additional posterior loss of non-optimal actions for point estimation and testing has been examined by [10] and [11]. Considerations on the conflict between optimal and non-optimal Bayesian actions can be found in [12].

This article moves from a preliminary study in [13] for the Poisson model and is organized as follows. In Section 2 we introduce the basic elements of the Bayesian decision-theoretic approach to set estimation. Section 3 is about optimality analysis for interval estimation of a normal mean with known and unknown variance. Using both simulations and clinical trial data we discuss the effect of the values of the penalizing constant for the length of intervals and the impact of prior choices on features of optimal sets (length, posterior probability and posterior expected loss) using the three losses. In Section 4 we change perspective: from the search of optimal sets (whose length and credibility is not fixed in advance) we switch to a regret analysis that aims at evaluating the additional cost

of using intervals with fixed credibility levels rather than optimal sets. Using the clinical trial example again, we examine the effects of the following quantities on the posterior regret associated with the exponential and rational losses: the value of the penalizing constant, the sample size, the credibility level. Finally, Section 5 summarizes the main points come to light in the article. All elaborations are performed using the software R [14]: code is available upon request.

2. Methodology

Given a parametric model, $f_n(\cdot\,|\,\theta), \theta \in \Theta$, let $\pi(\theta)$ denote the prior distribution of θ, $\mathbf{x}_n$ an observed sample of size n and $\pi(\theta\,|\,\mathbf{x}_n)$ the corresponding posterior distribution. For simplicity, suppose that $\Theta \subseteq \mathbb{R}^1$ and that $\pi(\theta)$ is a probability density function. Assuming to be interested in set estimation of θ from a decision theoretic perspective (see [3] and [7]), let $\mathcal{C}$ be a class of subsets of Θ and $\mathbb{L}(\theta, C)$ the loss function for a generic set $C \in \mathcal{C}$. This approach prescribes one to select a set $C^\star$ that minimizes the posterior expected loss $\rho(C, \mathbf{x}_n)$ as C varies in $\mathcal{C}$, *i.e.*:

$$C^\star = \arg\min_{C \in \mathcal{C}} \rho(C, \mathbf{x}_n), \quad \text{where } \rho(C, \mathbf{x}_n) = \int_\Theta \mathbb{L}(\theta, C)\pi(\theta\,|\,\mathbf{x}_n)\mathrm{d}\theta.$$

The most widely used family of losses for set estimation is defined by setting

$$\mathbb{L}(\theta, C) = \mathbb{S}\big[L(C)\big] + \mathbb{I}_{\bar{C}}(\theta), \tag{1}$$

where the *size* $\mathbb{S}(\cdot)$ is an increasing function of $L(C)$—the Lebesgue measure of C—and $\mathbb{I}_{\bar{C}}(\cdot)$ is the indicator function of the set $\bar{C} = \Theta \setminus C$. The resulting posterior expected loss of $C \in \mathcal{C}$ is

$$\rho(C, \mathbf{x}_n) = \mathbb{S}\big[L(C)\big] + 1 - \mathbb{P}(C\,|\,\mathbf{x}_n),$$

which embodies a compromise between the size of C and its posterior probability of containing θ, denoted as $\mathbb{P}(C\,|\,\mathbf{x}_n)$. One important property of the class of monotone functions (see, for instance, [2]) is that, if θ is an absolutely continuous random variable (as we assume here), optimal actions are highest posterior density (HPD) sets defined as $C^\star = \{\theta \in \Theta : \pi(\theta\,|\,\mathbf{x}_n) \geq h\}, h \geq 0$. More specifically, we here assume that HPD sets are intervals $C^\star = \big[L^\star, U^\star\big]$.

Let $\mathcal{L}(C) = U - L$ be the length of a generic interval C. The simplest form of loss (1) for C is obtained by selecting

$$\mathbb{S}_\ell\big[\mathcal{L}(C)\big] = a\mathcal{L}(C), \quad a > 0, \tag{2}$$

as size function, which yields the class of *linear* loss functions, $a\mathcal{L}(C) + \mathbb{I}_{\bar{C}}(\theta)$, where a is a penalizing constant for interval lengths. Casella, Hwang and Robert in [1] and [2] show that, in the case of unbounded parameter space, optimal sets under the linear loss function may be dominated by unreasonable sets. For instance, in the case of the normal model $\mathrm{N}(\theta, \sigma^2)$ with unknown variance, the standard Student's *t*-interval for θ is dominated by a set that is empty as the sample variance is sufficiently large. They also show that (under mild conditions) these kinds of problems are avoided if both the components of (1) assume values

in $[0,1]$ or, more specifically, if $\mathbb{S}(\cdot)$ is a nonlinear and increasing function that ranges monotonically in the unit interval such that $\lim_{C \to \varnothing} \mathbb{S}(\mathcal{L}(C)) = 0$ and $\lim_{C \to \Theta} \mathbb{S}(\mathcal{L}(C)) = 1$. To resolve this paradox the Authors propose the following two nonlinear functions

$$\mathbb{S}_e\big[\mathcal{L}(C)\big] = 1 - e^{-\frac{a\mathcal{L}(C)^2}{2}} \quad \text{and} \quad \mathbb{S}_r\big[\mathcal{L}(C)\big] = \frac{a\mathcal{L}(C)}{a\mathcal{L}(C)+1}, \; a < 0 \qquad (3)$$

that result in the classes of *exponential* and *rational* loss functions. The posterior expected losses corresponding to the three size functions under examination in this article are then given by:

$$\rho_j(C, \mathbf{x}_n) = \mathbb{S}_j\big[\mathcal{L}(C)\big] + 1 - \mathbb{P}(C \mid \mathbf{x}_n), \quad j = \ell, e, r. \qquad (4)$$

Note that whereas $\rho_\ell(C, \mathbf{x}_n) \geq 0$, for all C, *i.e.* it is unbounded above, the values of $\rho_e(C, \mathbf{x}_n)$ and $\rho_r(C, \mathbf{x}_n)$ are upper bounded. As an example, **Figure 1** shows $\rho_j(C, \mathbf{x}_n)$, $j = \ell, e, r$ as a function of $\mathcal{L}(C)$: shaded areas represent the contributions of $\mathbb{S}_j\big[\mathcal{L}(C)\big]$ (dotted lines) and of $1 - \mathbb{P}(C \mid \mathbf{x}_n)$ (solid lines) respectively.

Posterior expected losses ρ_j allow us to follow two different decisional approaches:

(i) *Optimality analysis*, *i.e.* determination and comparison of optimal intervals

$$C_j^\star = \arg\min_{C \in \mathcal{C}} \rho_j(C, \mathbf{x}_n) \quad j = \ell, e, r;$$

(ii) *Regret analysis*, *i.e.* evaluation of standard intervals of given credibility level γ, denoted by C^γ, in terms of their *additional expected loss* (or *expected regret*) with respect to $C_j^\star$, quantified by

$$\Delta_j(C^\gamma) = \rho_j(C^\gamma, \mathbf{x}_n) - \rho_j(C_j^\star, \mathbf{x}_n) \quad j = \ell, e, r. \qquad (5)$$

Note for any arbitrary choice of $\gamma \in [0,1]$, $\Delta_j(C^\gamma) \geq 0$.

Approaches (i) and (ii) will be now specialized to the normal model in Sections 3 and 4 respectively.

3. Optimality Analysis

Let us assume that $X_i \mid \theta \sim \mathrm{N}(\theta, \sigma^2)$, with σ^2 known, and that $\theta \sim \mathrm{N}(\mu_0, \sigma^2/n_0)$. These assumptions imply that $\theta \mid \mathbf{x}_n \sim \mathrm{N}(\bar{\mu}, \sigma^2/\bar{n})$ where $\bar{\mu} = (n_0\mu_0 + n\bar{x}_n)/\bar{n}$ and $\bar{n} = n_0 + n$. In this case the class of HPD sets—which are also equal tails (ET) intervals—is $C_k = \bar{\mu} \pm k\sigma/\sqrt{\bar{n}}$, $k > 0$, and

$$\rho_j(C_k, x_n) = \mathbb{S}_j\big[\mathcal{L}(C_k)\big] + 1 - \big[\Phi(k) - \Phi(-k)\big], \qquad (6)$$

where $\Phi(\cdot)$ is the standard normal cdf and $\mathbb{S}_j\big[\mathcal{L}(C_k)\big]$, for $j = \ell, e, r$, are obtained by setting $\mathcal{L}(C_k) = 2k\sigma/\sqrt{\bar{n}}$ in (2) and (3). The non-informative case is simply obtained by setting $n_0 = 0$, which yields standard confidence intervals $\bar{x}_n \pm k\sigma/\sqrt{n}$, $k > 0$. Note that the minimizers $k_j^\star$ can be determined as follows:

(i) $k_\ell^\star = \begin{cases} -2\ln\big(a\sqrt{2\pi}\,\sigma/\sqrt{\bar{n}}\big) & \text{if } a < \big(\sqrt{2\pi}\,\sigma/\sqrt{\bar{n}}\big)^{-1} \\ 0 & \text{otherwise} \end{cases}$;
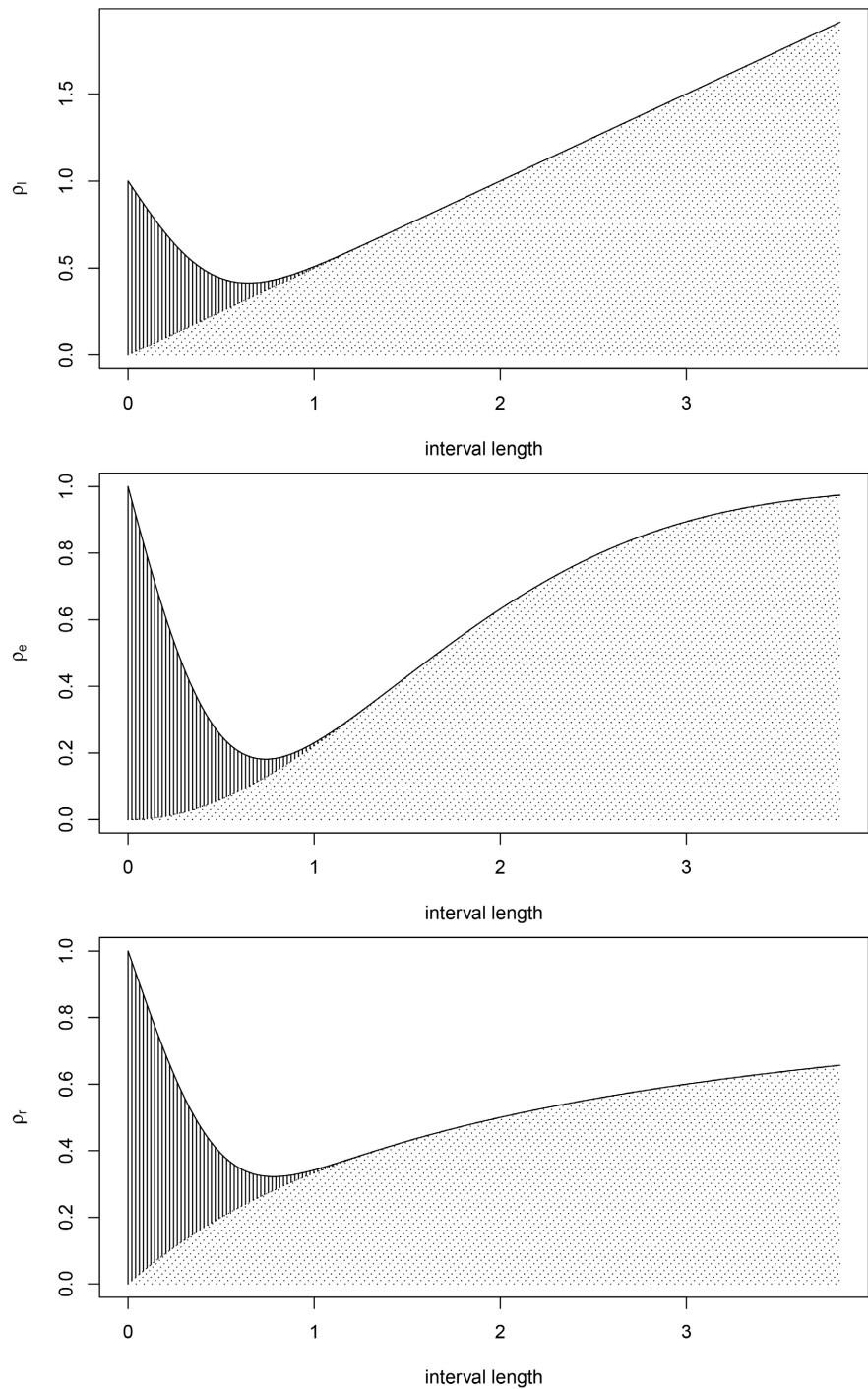
Figure 1. Posterior expected losses $\rho_j(C, \mathbf{x}_n)$, $\mathbb{S}_j[\mathcal{L}(C)]$ (dotted lines) and $1 - \mathbb{P}(C | \mathbf{x}_n)$ (solid lines), $j = \ell, e, r$, as functions of $\mathcal{L}(C)$.

(ii) $k_r^\star$ satisfies the equation $\phi[k_r^\star] = \dfrac{\bar{a}}{2[k_r^\star + \bar{a}]^2}$, where $\bar{a} = \dfrac{a\sqrt{n}}{2\sigma}$ and $\phi(\cdot)$ denotes the density function of a standard normal random variable;

(iii) $k_e^\star$ is determined numerically.

For the unknown variance case let us assume that $\theta\,|\,\sigma^2 \sim N\left(\mu_0, \sigma^2/n_0\right)$ and that $\sigma^2 \sim \text{Inv-Ga}\left(\nu_0/2, \text{rate} = \nu_0\sigma_0^2/2\right)$ or, equivalently, an inverse chi-square r.v. of parameters $\left(\nu_0, \text{rate} = \sigma_0^2\right)$. In this case standard conjugate analysis yields that $\theta\,|\,\mathbf{x}_n \sim t_{\bar{\nu}}\left(\bar{\mu}, \bar{\sigma}^2/\bar{n}\right)$, where $\bar{\nu} = \nu_0 + n$, $\bar{\mu}$ is unchanged w.r.t. the known variance case, $\bar{\nu}\bar{\sigma}^2 = \nu_0\sigma_0^2 + (n-1)S_n^2 + \dfrac{n_0 n}{n_0 + n}\left(\bar{x}_n - \mu_0\right)^2$, S_n^2 is the sample variance and $\bar{n} = n_0 + n$ (see Lesaffre and Lawson (2012), p. 86-87). HPD sets are then $C_k = \bar{\mu} \pm k\bar{\sigma}/\sqrt{\bar{n}}$, $k > 0$, and

$$\rho_j\left(C_k, \mathbf{x}_n\right) = \mathbb{S}_j\left[\mathcal{L}\left(C_k\right)\right] + 1 - \left[\mathbb{F}_{\bar{\nu}}\left(\bar{\mu} + k\bar{\sigma}/\sqrt{\bar{n}}\right) - \mathbb{F}_{\bar{\nu}}\left(\bar{\mu} - k\bar{\sigma}/\sqrt{\bar{n}}\right)\right], \quad (7)$$

where $\mathbb{F}_{\bar{\nu}}(\cdot)$ is the cdf of the t random variable with $\bar{\nu}$ degrees of freedom, location $\bar{\mu}$ and scale $\bar{\sigma}/\sqrt{\bar{n}}$ and $\mathbb{S}_j\left[\mathcal{L}\left(C_k\right)\right]$, for $j = \ell, e, r$, are obtained by setting $L\left(C_k\right) = 2k\bar{\sigma}/\sqrt{\bar{n}}$ in (2) and (3). Note that the non-informative case is retrieved for $n_0 = \nu_0 = 0$, i.e. using a flat prior for $\theta\,|\,\sigma^2$ and $\pi\left(\sigma^2\right) = 1/\sigma^2$. In this case HPD sets are $C_k = \bar{x}_n \pm kS_n/\sqrt{n}$ (see Lesaffre and Lawson (2012) p. 88). The minimizers $k_j^\star$ of (7) depend on the data and can be determined numerically, for $j = \ell, e, r$, or according to results similar to (i)-(iii). For instance, it can be checked that $k_r^\star$ satisfies the equation $\varphi_{\bar{\nu}}\left[k_r^\star\right] = \dfrac{\bar{a}}{2\left[k_r^\star + \bar{a}\right]^2}$, where

$\bar{a} = \dfrac{a\sqrt{n}}{2\bar{\sigma}}$ and $\varphi_{\bar{\nu}}(\cdot)$ denotes the density function of a $t_{\bar{\nu}}$ random variable with $\bar{\nu}$ degrees of freedom, location $\bar{\mu}$ and scale $\bar{\sigma}/\sqrt{\bar{n}}$. For futher details see [2].

3.1. Numerical Examples

For each loss function and for selected values of a we determine the optimal sets using the following procedure. We first consider a grid of values for k and determine the corresponding ET/HPD sets C_k and their expected loss $\rho_j\left(C_k, \mathbf{x}_n\right)$. Then we select $k_j^\star$ as the minimizer of $\rho_j\left(C_k, \mathbf{x}_n\right)$, for $j = \ell, e, r$. **Figure 2** shows the plots of $\rho_j\left(C_k, \mathbf{x}_n\right)$ as functions of k for Normal posteriors in the known variance case, under an informative prior ($n_0 = 5$, left panels) and a non-informative prior ($n_0 = 0$, right panels). For each value of a the optimal $k_j^\star$ is circled. According to the definition of the different size functions in (2) and (3), the impact of the penalizing coefficient a on interval lengths is not directly comparable. However, in general, as expected, the larger a the smaller $k_j^\star$, $j = \ell, e, r$. Among the three losses, $\rho_\ell\left(C_k, \mathbf{x}_n\right)$ is the most sensitive with respect to the values of a, which results in the largest range of values for $k_\ell^\star$. Note also that the optimal values $k_r^\star$ are the largest, yielding optimal intervals with posterior probability $\gamma_r^\star$ larger than 0.94 with smaller variability than those of the linear and of the exponential loss. For all the explored a values, the exponential loss function provides optimal sets with intermediate levels of length and posterior probability, and overall smaller values of the posterior expected loss. As expected, for each value of a, the curves ρ_j as functions of k are slightly but
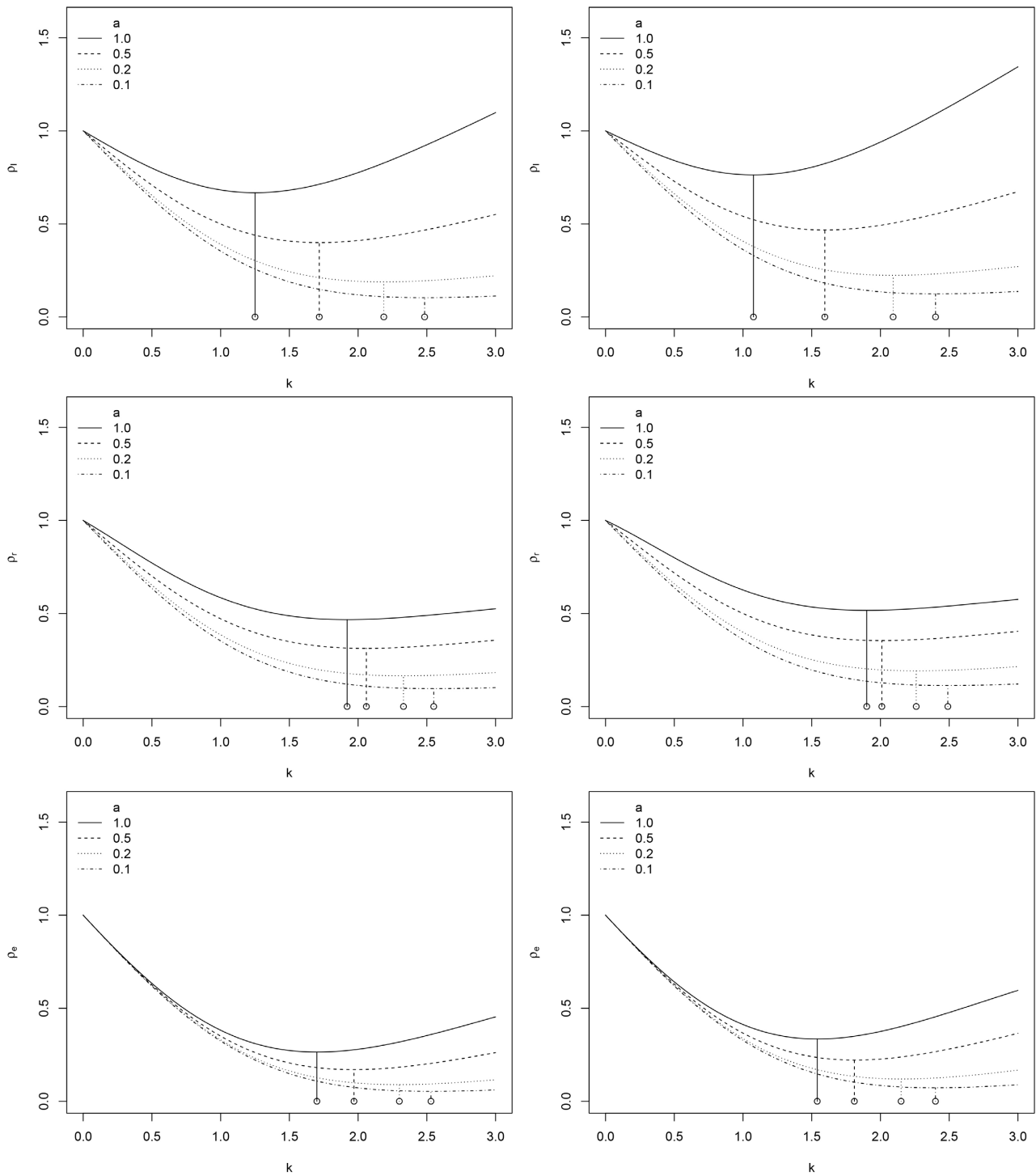
Figure 2. Posterior expected losses $\rho_j(C, \mathbf{x}_n)$, $j = \ell, e, r$, as functions of k for different values of a for Normal posteriors under an informative prior with $n_0 = 5$ (left column) and a non-informative prior with $n_0 = 0$ (right column). For each ρ_j, $j = \ell, e, r$, circles denote $k_j^\star$.

uniformly higher for the non-informative case (right panels). Interestingly, even though a selected $k_j^\star$ can be surprisingly smaller in the non-informative case, the corresponding value of $\rho_j^\star$ is always greater than in the informative case,

due to the larger posterior variance of θ. For instance, under the exponential loss function with $a = 1$, $k_e^\star$ and $\rho_e^\star$ are equal to 1.54 and 0.335 in the non-informative case and 1.70 and 0.264 in the informative case (see **Table 1**). The above comments are consistent with the numerical values of $k_j^\star$, $\mathcal{L}_j^\star$, $\gamma_j^\star$ and $\rho_j^\star$ reported in **Table 1** for the Normal model with known variance, both for the informative and the non-informative case.

Table 2 reports $k_j^\star$, $\mathcal{L}_j^\star$, $\gamma_j^\star$ and $\rho_j^\star$ for the unknown variance case assuming $\sigma_0^2 = 1$, (i) $n_0 = 0, \nu_0 = 100$ and (ii) $n_0 = 0, \nu_0 = 0$. Notice that the values of **Table 2** side (i) are very close to those of **Table 1** side (ii), since $\nu_0 = 100$ yields a prior for σ^2 highly concentrated on the variance parameter that is assumed to be known in the former case. As regards the comparison between the informative and the non-informative case, considerations similar to those made for the known variance case also apply to **Table 2** [compare (i) and (ii)].

3.2. Application to Clinical Data

In this section we consider an application from [15], in which an experimental drug (t), indicated for the treatment of iron deficiency anemia in adult patients with cronic kidney disease, is compared to a control treatment (c). The primary endpoint is the mean change in hemoglobin from baseline to day 35. Let $\theta = \theta_t - \theta_c$ be the difference between the expected changes in hemoglobin under the two treatments. Let $\bar{X}_n \sim \mathrm{N}\left(\theta, \sigma^2/n\right)$ be the difference between the sample mean changes in hemoglobin $\bar{X}_n^{(1)}$ and $\bar{X}_n^{(0)}$ under the two treatments, where

Table 1. Length, posterior probability, posterior expected loss for $C_{k_j^\star}$ and values of $k_j^\star$ under the three loss functions for selected values of *a*, for the Normal model with known variance $\sigma^2 = 1$ assuming (i) an informative prior for θ (e.g. $n_0 = 5$) and (ii) a non-informative prior for θ (e.g. $n_0 = 0$).

Loss	a	(i) $n_0 = 5$				(ii) $n_0 = 0$			
		$\mathcal{L}_j^\star$	$\gamma_j^\star$	$\rho_j^\star$	$k_j^\star$	$\mathcal{L}_j^\star$	$\gamma_j^\star$	$\rho_j^\star$	$k_j^\star$
linear	1.0	0.480	0.648	0.833	0.930	0.430	0.503	0.927	0.680
	0.5	0.775	0.866	0.521	1.500	0.860	0.826	0.604	1.360
	0.2	1.043	0.957	0.252	2.020	1.214	0.945	0.298	1.920
	0.1	1.208	0.981	0.140	2.340	1.423	0.976	0.167	2.250
rational	1.0	0.981	0.943	0.553	1.900	1.202	0.943	0.603	1.900
	0.5	1.028	0.953	0.719	1.990	1.297	0.960	0.762	2.050
	0.2	1.203	0.980	0.877	2.330	1.537	0.985	0.900	2.430
	0.1	1.368	0.992	0.940	2.650	1.733	0.994	0.952	2.740
exponential	1.0	0.733	0.844	0.391	1.420	0.791	0.789	0.480	1.250
	0.5	0.878	0.911	0.264	1.700	0.974	0.876	0.335	1.540
	0.2	1.059	0.960	0.146	2.050	1.202	0.943	0.192	1.900
	0.1	1.188	0.979	0.090	2.300	1.360	0.968	0.120	2.150

Table 2. Length, posterior probability, posterior expected loss for $C_{k_j^\star}$ and values of $k_j^\star$ under the three loss functions for selected values of a, for the Normal model with unknown variance, assuming $\sigma_0^2 = 1$, (i) $n_0 = 0, v_0 = 100$ and (ii) $n_0 = 0, v_0 = 0$.

Loss	a	(i) $n_0 = 0, v_0 = 100$				(ii) $n_0 = 0, v_0 = 0$			
		$\mathcal{L}_j^\star$	$\gamma_j^\star$	$\rho_j^\star$	$k_j^\star$	$\mathcal{L}_j^\star$	$\gamma_j^\star$	$\rho_j^\star$	$k_j^\star$
linear	1.0	0.428	0.502	0.926	0.680	0.420	0.500	0.920	0.700
	0.5	0.856	0.823	0.605	1.360	0.828	0.802	0.612	1.380
	0.2	1.215	0.944	0.299	1.930	1.206	0.928	0.313	2.010
	0.1	1.429	0.975	0.168	2.270	1.464	0.965	0.181	2.440
rational	1.0	1.202	0.941	0.605	1.910	1.194	0.925	0.619	1.990
	0.5	1.297	0.958	0.764	2.060	1.308	0.946	0.778	2.180
	0.2	1.542	0.984	0.901	2.450	1.650	0.980	0.912	2.750
	0.1	1.750	0.994	0.952	2.780	1.800	0.987	0.961	3.000
exponential	1.0	0.787	0.786	0.480	1.250	0.762	0.767	0.485	1.270
	0.5	0.976	0.876	0.336	1.550	0.948	0.855	0.346	1.580
	0.2	1.202	0.941	0.193	1.910	1.194	0.925	0.207	1.990
	0.1	1.366	0.968	0.121	2.170	1.380	0.956	0.135	2.300

$\sigma^2/n = \sigma_t^2/n_t + \sigma_c^2/n_c$ and $n = n_t + n_c$, with n_i the size of the arm $i = t, c$. In the example $n_t = 62$, $n_c = 64$, $\bar{x}_{n_t} = 0.87$ and $\bar{x}_{n_c} = 0.71$. Here, we assume that σ_t and σ_c are known and equal to the sample standard deviations reported in the original example, *i.e.* 1.14 and 1 respectively. Using the non-informative prior $\pi(\theta) = \text{const}$ for θ, the 95% credible interval is $(-0.215, 0.535)$.

Table 3 shows the optimal intervals under the three loss functions for $a = 1$ with different choices of the prior parameters: (A) $\mu_0 = 0, n_0 = 100$ (no difference between treatment effects); (B) $\mu_0 = 0.16, n_0 = 100$ (prior information perfectly matching sample data); (C) $\mu_0 = 0.32, n_0 = 100$ (optimistic prior mean); (D) $n_0 = 0$ (non-informative).

As μ_0 increases [from (A) to (C)], intervals bounds are shifted towards larger values, but the selected $k_j^\star$ is the same for each given loss function, thus yielding the same values of $\mathcal{L}_j^\star$, $\gamma_j^\star$ and $\rho_j^\star$, $j = \ell, e, r$. In the non-informative case (D), the values of $\rho_j^\star$ and $\mathcal{L}_j^\star$ are uniformly greater than in the previous cases, due to the larger posterior variance. The rational loss always yields the widest optimal intervals, with posterior probability close to the conventional level 95%.

Figure 3 shows the behavior of $\mathcal{L}_j^\star$, $\gamma_j^\star$, $\rho_j^\star$ and $k_j^\star$ as functions of the coefficient *a*. As *a* increases, $\mathcal{L}_j^\star$ and $\gamma_j^\star$ tend to decrease and the corresponding values of $\rho_j^\star$ tend to increase. As expected the linear loss (solid line) is the most sensitive to changes of the values of *a*. Due to remark (i) of Section 3, $C_{k_\ell^\star}$ is non trivial ($k_\ell^\star > 0$) for values of $a < \left(\sqrt{2\pi}\,\sigma/\sqrt{n}\right)^{-1}$, that is equal to 2.08 in this example. This motivates the presence of the cusp in the three plots. Note that in the second panel the curve representing $\gamma_r^\star$ shows values always very
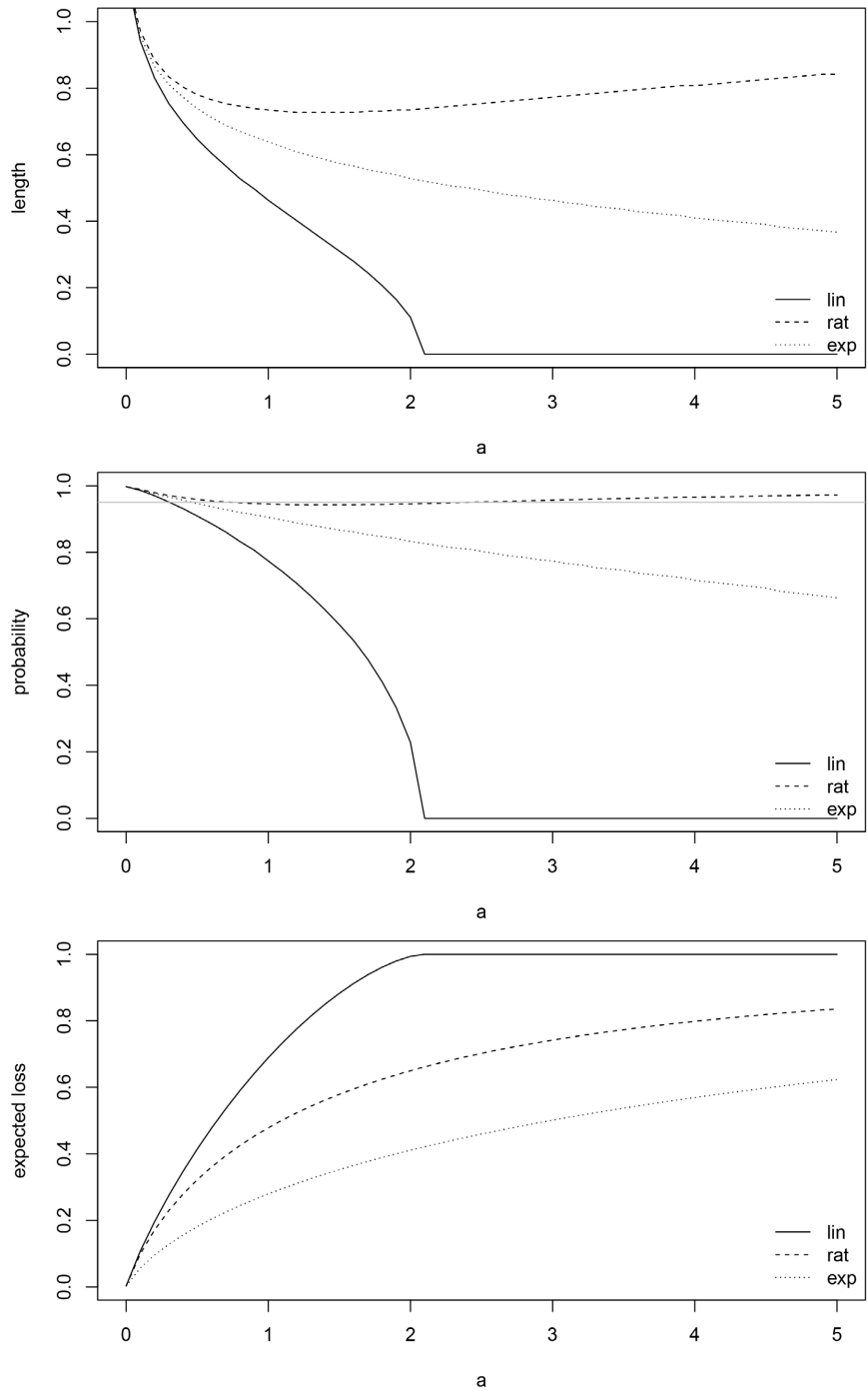
Figure 3. Length, posterior probability and posterior expected loss of the optimal set C^* as functions of *a*, under the linear (solid line), rational (dashed line) and exponential loss (dotted line), for the Normal model with known variance assuming $n_0 = 0$ (non-informative).

close to 0.95; whereas γ_e^* progressively reduces as *a* increases. Finally, inspection of the values of ρ_j^* allows an overall look at the sensitivity to *a* of the three losses (see third panel). Although a direct comparison of the absolute values of ρ_j^* is not meaningful, due to the different role of *a* in the three loss functions,

Table 3. Bounds, length, posterior probability, posterior expected loss for optimal sets $C_{k_j^\star}$ and values of $k_j^\star$ under the three loss functions with $a = 1$, for the Normal model with known variance, assuming four alternative priors: (A) $\mu_0 = 0, n_0 = 10$ (no difference between treatment effects); (B) $\mu_0 = 0.16, n_0 = 10$ (prior information perfectly matching sample data); (C) $\mu_0 = 0.32, n_0 = 10$ (optimistic prior mean); (D) $n_0 = 0$ (non-informative).

Loss	Prior	μ_0	n_0	$\mathcal{L}_j^\star$	$U_j^\star$	$L_j^\star$	$\gamma_j^\star$	$\rho_j^\star$	$k_j^\star$
linear	(A)	0	100	−0.115	0.293	0.408	0.847	0.561	1.430
	(B)	0.16	100	−0.044	0.364	0.408	0.847	0.561	1.430
	(C)	0.32	100	0.027	0.435	0.408	0.847	0.561	1.430
	(D)	-	0	−0.071	0.391	0.463	0.774	0.689	1.210
rational	(A)	0	100	−0.191	0.369	0.560	0.950	0.409	1.960
	(B)	0.16	100	−0.120	0.440	0.560	0.950	0.409	1.960
	(C)	0.32	100	−0.049	0.511	0.560	0.950	0.409	1.960
	(D)	-	0	−0.207	0.527	0.734	0.945	0.478	1.920
exponential	(A)	0	100	−0.181	0.359	0.540	0.941	0.194	1.890
	(B)	0.16	100	−0.110	0.430	0.540	0.941	0.194	1.890
	(C)	0.32	100	−0.039	0.501	0.540	0.941	0.194	1.890
	(D)	-	0	−0.159	0.479	0.639	0.905	0.280	1.670

the rate of increase of the three curves shows a greater degree of robustness of the exponential loss.

4. Regret Analysis

As discussed at the end of Section 2 one typically uses the suboptimal sets C^γ that guarantee minimal length in the class of fixed γ-credibility level intervals. For the normal model with known variance $C^\gamma = \bar{\mu} \pm z_{\frac{1+\gamma}{2}} \sigma/\sqrt{n}$. In this section we are interested in evaluating $\Delta_j(C^\gamma)$, the additional expected loss of sets C^γ given by Equation (5). Our goal is to quantify the cost of using the more pragmatic interval C^γ instead of the optimal set $C_{k_j^\star}$ under a given loss function. Here the comparison is restricted to the exponential and rational loss functions, due to the drawbacks of the linear loss previously discussed. The values of $\Delta_j(C^\gamma)$ are computed noting that $\rho_j(C^\gamma, \mathbf{x}_n) = \mathbb{S}_j\left[2z_{\frac{1+\gamma}{2}} \sigma/\sqrt{n}\right] + 1 - \gamma$ and determining the values of $\rho_j(C_{k_j^\star}, \mathbf{x}_n)$ numerically as discussed in the previous section.

We explore the behaviour of $\Delta_j(C^\gamma)$ with respect to *a*, *n* and γ for the application of Section 3.2. Moreover, we compare two alternative prior assumptions, *i.e.* the non-informative case (with $n_0 = 0$) and an informative case (with $n_0 = 100$). Note under the non-informative prior the γ-credible interval coincides with the frequentist γ-confidence interval $\bar{x}_n \pm z_{\frac{1+\gamma}{2}} \sigma/\sqrt{n}$.

Application to Clinical Data (Continued)

Figures 4-6 show the values of $\Delta_j(C^\gamma)$, $j = e, r$, as functions of a, n and γ respectively. Solid lines correspond to a prior sample size $n_0 = 100$, dashed lines to the non-informative case ($n_0 = 0$).

1) *Effect of a* (**Figure 4**)

a) As a varies the curves Δ_j have a minimum in $\tilde{a}_j$, which represents the value of the penalizing coefficient such that C^γ are the closest to the optimal sets $C_{k_j^\star}$. Note that $\Delta_j = 0$ when the value $\tilde{a}_j$ is such that C^γ is optimal under the loss function j.

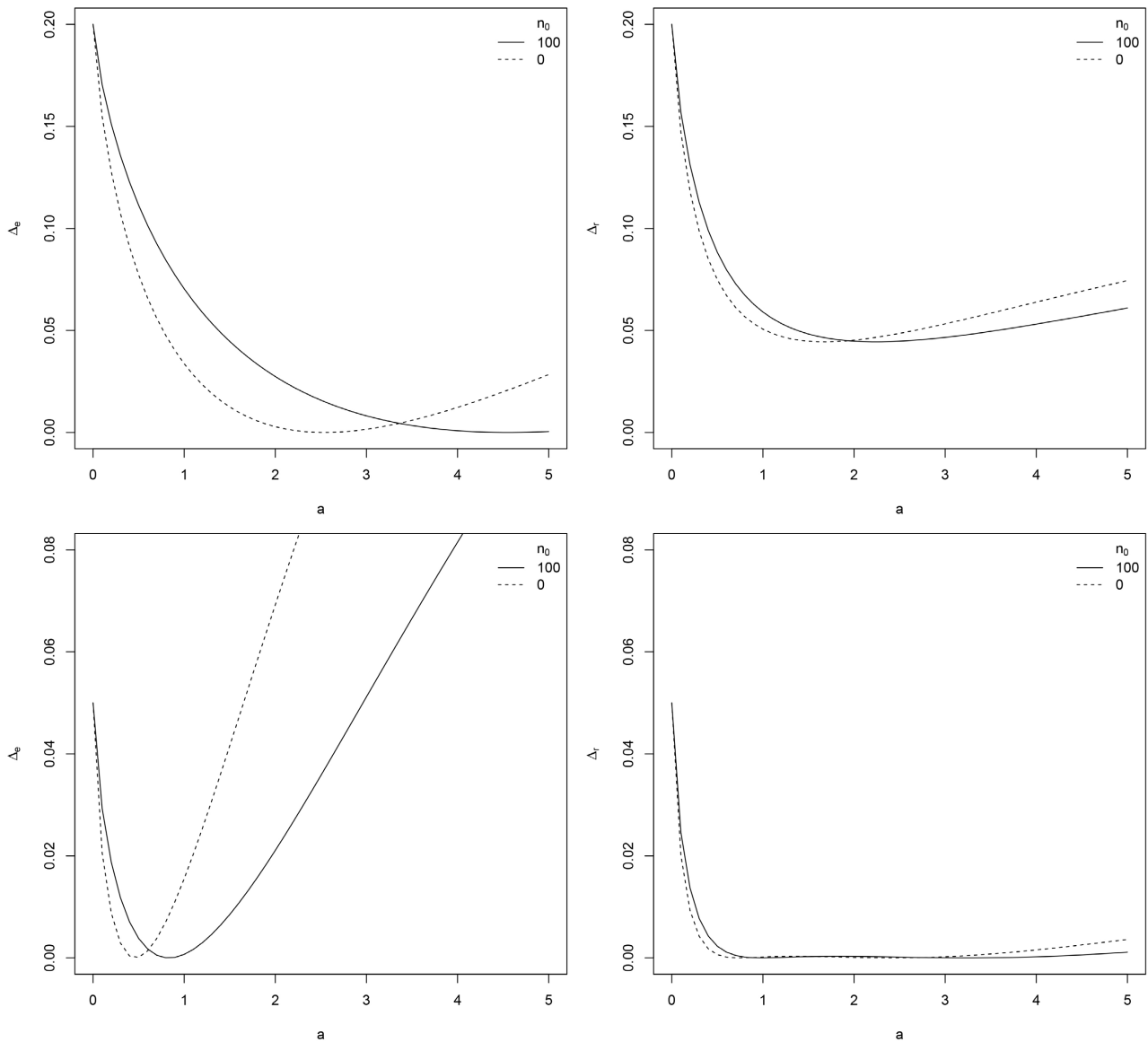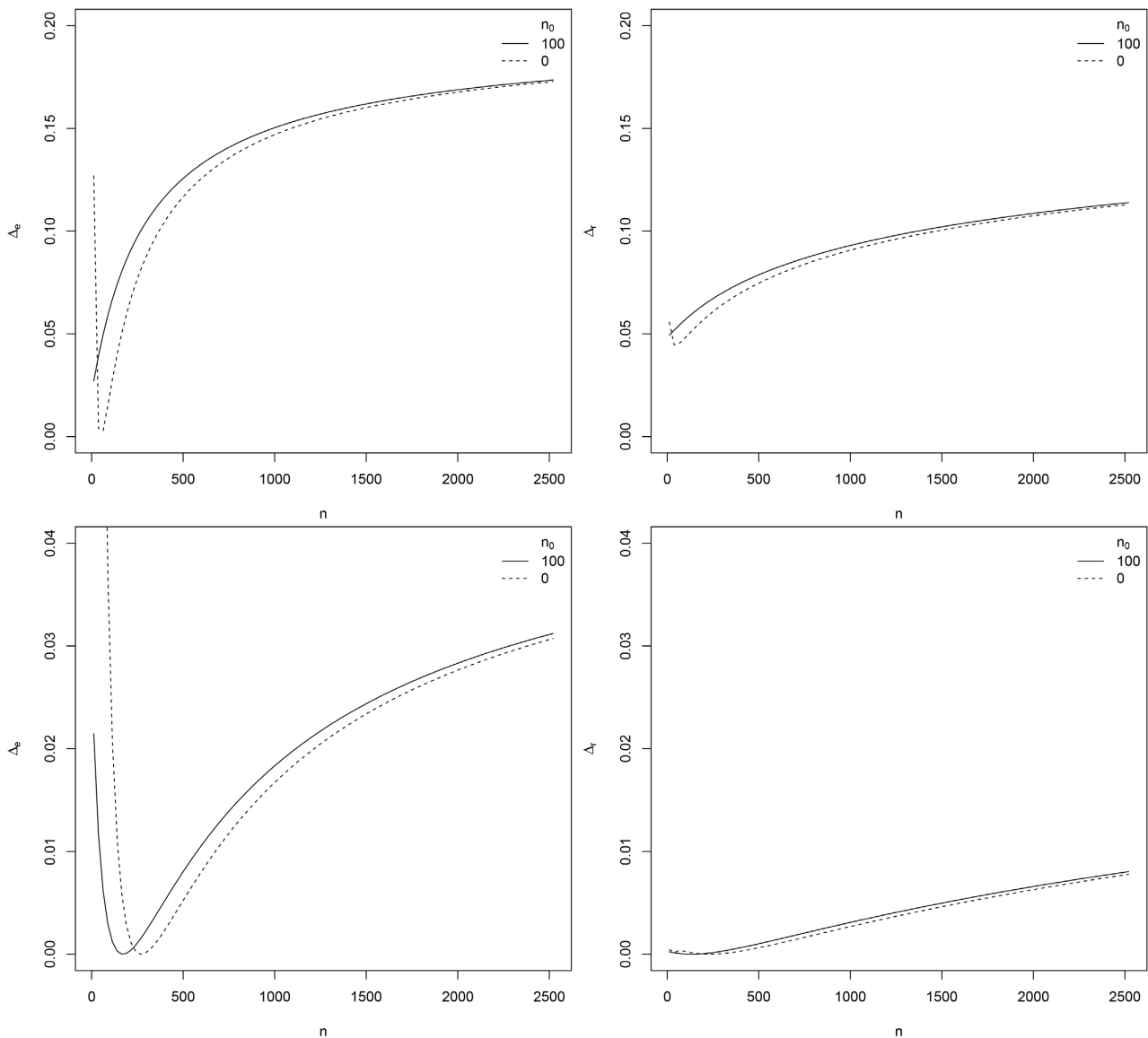b) In all panels but the last one (see next item), $\tilde{a}_j$ is smaller for $n_0 = 0$ than

Figure 4. Additional expected loss $\Delta_j(C^\gamma)$ with $\gamma = 0.80$ (top panels) and $\gamma = 0.95$ (bottom panels) as function of a, under the exponential (left column) and the rational loss (right column), for the Normal model with known variance assuming $n_0 = 100$ (solid line) and $n_0 = 0$ (dashed line) with $n = 126$.
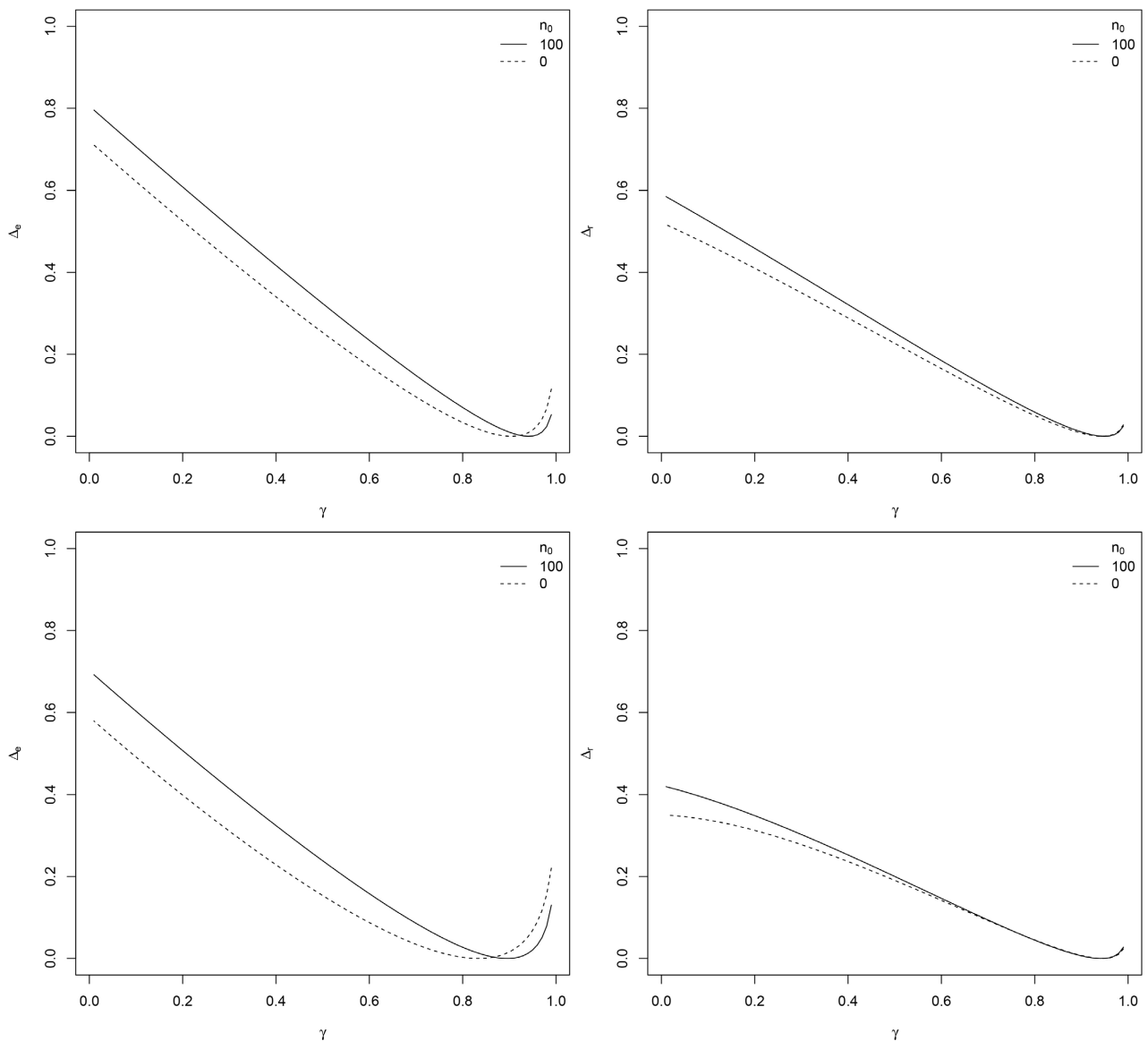
Figure 5. Additional expected loss $\Delta_j\left(C^\gamma\right)$ with $\gamma = 0.80$ (top panels) and $\gamma = 0.95$ (bottom panels) as functions of n, under the exponential (left column) and the rational loss (right column) with $a = 1$, for the Normal model with known variance assuming $n_0 = 100$ (solid line) and $n_0 = 0$ (dashed line).

for $n_0 = 100$. This means that, since the non-informative prior yields larger sets, the degree of penalization of interval length has to be smaller than it can be for larger values of n_0.

c) Under the rational loss function, for $\gamma = 0.95$ (last panel) Δ_r is substantially negligible for a large range of values of a. This is consistent with what we observed in the previous section (see for instance **Figure 3** second panel) since the optimal sets under the rational loss have credibility approximately equal to 0.95 for several values of a.

d) According to the choice of a either the standard confidence interval or the Bayesian credible interval can be preferred in terms of additional expected loss.

Figure 6. Additional expected loss $\Delta_j\left(C^\gamma\right)$ as function of γ, under the exponential (left column) and the rational loss (right column) with $a=1$ (top panels) $a=2$ (center panels) and $a=5$ (bottom panels) for the normal model with known variance assuming $n_0=100$ (solid line) and $n_0=0$ (dashed line).

2) *Effect of n* (**Figure 5**)

a) The range of Δ_r is much smaller than that of Δ_e. Values of Δ_j are particularly small when $\gamma=0.95$ (bottom panels).

b) In all the plots the solid and dashed curves Δ_j as functions of n have a minimum point $\tilde{n}_j$. This means that there exist a value of the sample size such that C^γ is optimal. Such a value $\tilde{n}_j$ is smaller for $n_0=100$ than for $n_0=0$ since the informative prior implies shorter γ-credible sets than the non-informative one.

c) For values larger than $\tilde{n}_j$ the curves Δ_j increase. In particular the higher steepness of solid curves is due to the shorter length of intervals under the in-

formative prior for each given sample size. Moreover, solid and dashed curves eventually tend to coincide because the effect of n_0 becomes more and more negligible with respect to larger and larger values of *n*.

d) For $\gamma = 0.8$ (top panels) there are not values of *n* such that $\Delta_j = 0$, *i.e.* sets C^γ are never optimal.

3) *Effect of* γ (**Figure 6**)

a) In all panels the maximum value of Δ_j is obtained for $\gamma = 0$ (when C^γ reduces to $\bar{\mu}$).

b) For increasing values of γ the curves Δ_j decrease and reach their minimum at $\gamma_j^\star$, *i.e.* Δ_j increases with $\left| \gamma - \gamma_j^\star \right|$.

c) The smaller the value of γ the larger the discrepancy between the lengths of $C_{k_j^\star}$ and C^γ and this discrepancy is stronger for informative priors that yield shorter intervals with respect to non-informative priors. The opposite is observed for large values of γ when the exponential loss is used; whereas the rational loss seems to be more insensitive to the prior at least for values of $\gamma > 0.5$.

d) Depending on the chosen value for γ, either the standard confidence interval or the Bayesian credible interval can have the smallest regret. However, for large values of γ —typically chosen in the practice—regret is smaller for the informative Bayesian interval.

5. Conclusions

We can now attempt to summarize some indications drawn from the numerical example and the clinical data application of Sections 3 and 4.

As regards the optimality analysis the main points are the following.

1) As shown by **Table 1** and **Table 2**, the value of the penalizing coefficient *a* is critical: there are no general guidelines for its choice that is, however, highly influential in determining optimal intervals. This is also true in general in decisional analysis, for instance in testing problems when generalized 0 - 1 losses are used.

2) Non-informative distributions imply larger values of $\mathcal{L}_j$ and, as a consequence, of ρ_j than those obtained with informative priors, as one can argue by looking at **Table 3**.

3) **Figure 2** and **Figure 3** allow us to sketch a comparison among the behaviour with respect to *a* of the posterior expected losses induced by the three loss functions: the linear loss is the most sensitive; the rational loss yields larger sets and values of $\gamma_r^\star$ close to 0.95; the values of ρ_e are the most robust with respect to *a*.

In the regret analysis of Section 4 we have explored the impact of *a*, *n* and γ. Here are the main comments.

1) The value $\tilde{a}_j$, that minimizes the additional loss of C^γ, quantifies the degree of penalization such that C^γ is as close as possible to be optimal. In this sense it can be useful to support the interpretation and the choice of *a*.

2) As shown in **Figure 4**, according to the choice of a either the standard confidence interval or the Bayesian credible interval can be preferred in terms of additional expected loss.

3) **Figure 5** suggests that, as n varies, the range of Δ_r is smaller than that of Δ_e. Furthermore it hints that it is possible to find values of n such that C^γ is optimal and sample size beyond which the weight of the prior on Δ_j becomes more and more negligible.

4) In terms of additional expected loss, it is not granted that the Bayesian credible interval has to be preferred with respect to the standard confidence interval: it depends on the chosen value for γ. Nevertheless, in the practice large values of γ are typically fixed. According to **Figure 6** for large values of γ the informative Bayesian interval have (slightly) smaller regret than confidence intervals.

As pointed out in [2] "it is difficult to treat the set estimation problem in a decision-theoretic way" because of some limitations related to the available loss functions and the difficulty of selecting an appropriate loss function. Nevertheless the possibility of balancing size and credibility is still appealing. In this paper we have explored the features of some loss functions that are alternative to the linear loss function and that produce sensible sets for normal models. Bearing in mind that the distribution of the data plays an important role, we hope to extend the present analysis to other models, in the spirit of [13].

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Casella, G., Hwang, J.T.G. and Robert, C. (1993) A Paradox in Decision-Theoretic Interval Estimation. *Statistica Sinica*, **3**, 141-155.

[2] Casella, G., Hwang, J.T.G. and Robert, C. (1994) Loss Function for Set Estimation, In: Berger, J.O. and Gupta, S.S., Eds., *Statistical Decision Theory and Related Topics V*, Springer-Verlag, New York, 237-252. https://doi.org/10.1007/978-1-4612-2618-5_18

[3] Berger, J.O. (1985) Statistical Decision Theory and Bayesian Analysis. Springer-Verlag, Chicago. https://doi.org/10.1007/978-1-4757-4286-2

[4] French, S. and Rios Insua, D. (2000) Statistical Decision Theory, Kendall's Library of Statistics. Wiley, Hoboken.

[5] Pratt, J.W., Raiffa, H. and Schlaifer, R. (2008) Introduction to Statistical Decision Theory. MIT Press, Cambridge.

[6] Parmigiani, G. and Inoue, L. (2009) Decision Theory: Principles and Approaches. Wiley Series in Probability and Statistics. Wiley, Hoboken. https://doi.org/10.1002/9780470746684

[7] Robert, C. (2007) The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation. Springer Science & Business Media, Berlin.

[8] Bernardo, J.M. (2005) Intrinsic Credible Regions: An Objective Bayesian Approach

to Interval Estimation. *Test*, **14**, 317-384. https://doi.org/10.1007/BF02595408

[9] Rice, K. and Ye, L. (2022) Expressing Regret: A Unified View of Credible Intervals. *The American Statistician*, **76**, 248-256. https://doi.org/10.1080/00031305.2022.2039764

[10] De Santis, F. and Gubbiotti, S. (2016) A Decision-Theoretic Approach to Sample Size Determination under Several Priors. *Applied Stochastic Models in Business and Industry*, **33**, 282-295. https://doi.org/10.1002/asmb.2211

[11] De Santis, F. and Gubbiotti, S. (2021) On the Predictive Performance of a Non-Optimal Action in Hypothesis Testing. *Statistical Methods & Applications*, **30**, 689-709. https://doi.org/10.1007/s10260-020-00539-1

[12] Brutti, P., De Santis, F. and Gubbiotti, S. (2014) Predictive Measures of the Conflict between Frequentist and Bayesian Estimators. *Journal of Statistical Planning and Inference*, **148**, 111-122. https://doi.org/10.1016/j.jspi.2013.12.009

[13] De Santis, F. and Gubbiotti, S. (2021) Optimal Credible Intervals under Alternative Loss Functions. In: *Book of Short Papers SIS* 2021, *Statistical Methods & Applications*, Pearson, Milano, 1618-1623.

[14] R Core Team (2022) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org

[15] Gamalo, M.A., Wu, R. and Tiwari, R.C. (2016) Bayesian Approach to Non-Inferiority Trials for Normal Means. *Statistical Methods in Medical Research*, **25**, 221-240. https://doi.org/10.1177/0962280212448723