Objective Bayes Estimation and Hypothesis Testing: the reference-intrinsic approach

Miguel A. Juárez

University of Warwick

M.A.Juarez@Warwick.ac.uk

Abstract

Conventional frequentist solutions to point estimation and hypothesis testing typically need ad hoc modifications when dealing with non-regular models, and may prove to be misleading. The decision oriented objective Bayesian approach to point estimation using conventional loss functions produces non-invariant solutions, and conventional Bayes factors suffer from Jeffreys-Lindley-Bartlett paradox. In this paper we illustrate how the use of the intrinsic discrepancy combined with reference analysis produce solutions to both point estimation and precise hypothesis testing, which are shown to overcome these difficulties. Specifically, we illustrate the methodology with some non-regular examples. The solutions obtained are compared with some previous results.

Keywords: Bayesian reference criterion, intrinsic divergence, intrinsic estimator, logarithmic discrepancy, reference prior.

1 Introduction

Two of the most widely studied problems in statistics textbooks are point estimation and precise hypothesis testing. From a frequentist standpoint, maximum likelihood is the most exploited technique in point estimation and a chief ingredient in sharp hypothesis testing. Though it is assumed that asymptotic properties of maximum likelihood estimators (MLE) usually hold, –e.g. asymptotic normality– this is not necessarily the case when the sampling distribution is not regular, and frequently ad-hoc modifications are needed, reducing the generality of the procedure. Moreover, a number of drawbacks have been exposed about the use of p-values in sharp hypothesis testing (see e.g. Edwards *et al.*, 1963; Berger and Selke, 1987; Selke *et al.*, 2001).

From a decision-theoretic Bayesian standpoint, any of these problems may be posed as $\{\mathcal{M}, \pi, \ell, \mathcal{A}\}$, where $\mathcal{M} = \{p(\boldsymbol{x} \mid \boldsymbol{\psi}), \boldsymbol{x} \in \mathcal{X}, \boldsymbol{\psi} \in \Psi\}$ is the sampling model for the observable $\boldsymbol{x}$; $\pi(\boldsymbol{\psi})$ is a probability density function (pdf) describing the decision maker's prior beliefs about the parameter $\boldsymbol{\psi}$; $\mathcal{A}$ is the space of possible actions; and $\ell(\boldsymbol{\psi}, a)$ is a loss function that measures the consequences of deciding to act according to $a \in \mathcal{A}$, when the true value of the parameter is $\boldsymbol{\psi}$. Bayes decision rule, that which minimises the posterior expected loss, is then optimal for the specific problem at hand. However, in some circumstances such as scientific communication or public decision making, there is a need for a solution that lets the data speak for itself or that introduces as little subjective information as possible. It may also be the case that the decision maker wants to assess the impact that different subjective inputs (both on the prior and the loss functions) have on the final decision, and thus, a benchmark solution is needed. Objective Bayes methods are aimed to provide such solutions.

From an objective Bayesian perspective, it is well known that the irrespective use of "flat" priors may be misleading. Furthermore, Jeffreys' prior may not be defined when the assumed model is non-regular. Likewise, though point estimators derived from "automatic" loss functions –quadratic, zero-one and linear– may be viewed as acceptable location measures of the quantity of interest, their lack of invariance under bijective transformations (with exception of the one-dimensional median) may be suspicious for scientific purposes, where a specific parametrisation is generally chosen for convenience.

Regarding sharp hypothesis testing, the use of improper priors may lead to indeterminate answers. The commonly used tool to overcome this problem, conventional Bayes Factors (Jeffreys, 1961, §5.2), assumes that the prior has a point mass on the null value. This setting leads to what is come to be known as the Jeffreys-Lindley-Bartlett (JLB) paradox (Bartlett, 1957; Lindley, 1957). Although there has been some attempts to overcome these disadvantages (Berger and Pericchi, 2001; O'Hagan, 1995; Robert and Caron, 1996), the resulting factors do not necessarily correspond to any prior, thus being open to criticism.

This paper illustrates how we can merge the use of the reference algorithm (Berger and Bernardo, 1992; Bernardo, 1979) to derive non-informative priors, with the *intrinsic discrepancy* (Bernardo and Rueda, 2002; Bernardo and Juárez, 2003) as loss function to obtain an objective answer for both problems. Objective in the precise sense of only depending on the data and the sampling model. The paper is organised as follows: in Section 2 the general framework is described, in Section 3 some non-regular models will be analysed using the proposed methodology and a few final remarks are presented in Section 4.

2 The Reference-Intrinsic Methodology

Assume it is agreed that the probabilistic behaviour of an observable, $\boldsymbol{x}$, is well described by the model $\mathcal{M} = \{p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\lambda}), \boldsymbol{x} \in X, \boldsymbol{\theta} \in \Theta, \boldsymbol{\lambda} \in \Lambda\}$, and that one is interested in testing whether the (null) hypothesis $H_0 \equiv \{\boldsymbol{\theta} = \boldsymbol{\theta}_0\}$ is compatible with the data. Bernardo (1999) argues that this situation may be posed as a decision problem where the loss function is conveniently described by a proper scoring rule, derived from the Kullback-Leibler (directed) divergence from the assumed model to the simplified model induced by the null, $\mathcal{M}_0 = p(\boldsymbol{x} \mid \boldsymbol{\theta}_0, \boldsymbol{\lambda})$, i.e.

$$k(\boldsymbol{\theta}_0 \mid \boldsymbol{\theta}, \boldsymbol{\lambda}) = \inf_{\boldsymbol{\lambda}_0 \in \boldsymbol{\Lambda}} \int_{\boldsymbol{X}} p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\lambda}) \log \frac{p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\lambda})}{p(\boldsymbol{x} \mid \boldsymbol{\theta}_0, \boldsymbol{\lambda}_0)}\, d\boldsymbol{x}.$$

There is a large literature about the use of the Kullback-Leibler divergence in statistics starting with the works of Kullback (1968) and Kullback and Leibler (1951). Robert (1996) regards it as an intrinsic loss in order to address point estimation within a decision-theoretical framework and Gutiérrez-Peña (1992) derives some properties when the assumed distribution is a member of the exponential family. Bernardo (1982, 1985); Bernardo and Bayarri (1985); Ferrándiz (1985) and Rueda (1992) make use of it in particular hypothesis testing settings.

Noting that the quantity of interest in a decision problem is that which enters the loss function, $k = k\left(\boldsymbol{\theta}_0 \mid \boldsymbol{\theta}, \boldsymbol{\lambda}\right)$ is then taken as the quantity of interest for which the reference posterior, $\pi_k(k \mid \boldsymbol{x})$, is derived. Consequently, the posterior expected value of the logarithmic divergence

$$d(\boldsymbol{\theta}_0, \boldsymbol{x}) = \int k\, \pi_k(k \mid \boldsymbol{x})\, \mathrm{d}k,$$

gives a measure of the discrepancy from the true, full model to the simplified one, given by the data.

Recently, Bernardo and Rueda (2002) propose a more general form for the loss function, namely

Definition 1 (Intrinsic discrepancy).
Assume that $\mathcal{M} = \{p\left(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\lambda}\right), \boldsymbol{x} \in \mathcal{X},\ \boldsymbol{\theta} \in \Theta, \boldsymbol{\lambda} \in \Lambda\}$ *is a probability model that describes the random behaviour of the observable quantity* $\boldsymbol{x}$*. The* intrinsic discrepancy (loss) *of using the simplified model, with fixed* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$*, instead of the full model is given by*

$$\delta\left(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\theta}_0\right) = \min\Big\{k\left(\boldsymbol{\theta}_0 \mid \boldsymbol{\theta}, \boldsymbol{\lambda}\right), k\left(\boldsymbol{\theta}, \boldsymbol{\lambda} \mid \boldsymbol{\theta}_0\right)\Big\},$$

where

$$k\left(\boldsymbol{\theta}_0 \mid \boldsymbol{\theta}, \boldsymbol{\lambda}\right) = \min_{\lambda_0 \in \Lambda} \int_X f(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\lambda}) \log \frac{f(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\lambda})}{f(\boldsymbol{x} \mid \boldsymbol{\theta}_0, \boldsymbol{\lambda}_0)}\, \mathrm{d}\boldsymbol{x},$$

and

$$k(\boldsymbol{\theta}, \boldsymbol{\lambda} \mid \boldsymbol{\theta}_0) = \min_{\lambda_0 \in \Lambda} \int_X f(\boldsymbol{x} \mid \boldsymbol{\theta}_0, \boldsymbol{\lambda}_0) \log \frac{f(\boldsymbol{x} \mid \boldsymbol{\theta}_0, \boldsymbol{\lambda}_0)}{f(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\lambda})} \, \mathrm{d}\boldsymbol{x}.$$

The intrinsic discrepancy has a number of appealing properties: it is symmetric, non-negative and vanishes only if $f(\boldsymbol{x} \mid \boldsymbol{\omega}, \gamma) = f(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\lambda})$ a.e. Is additive for conditional independent quantities, in the sense that if $\boldsymbol{z} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ are independent given the parameters, then $\delta_{\boldsymbol{z}}(\boldsymbol{\theta}; \boldsymbol{\theta}_0) = \sum_{i=1}^{n} \delta_{\boldsymbol{x}_i}(\boldsymbol{\theta}; \boldsymbol{\theta}_0)$. Is invariant under one-to-one transformations of both the data and the parameter of interest and it is also invariant under the choice of the nuisance parameter. Moreover, if $f_1(\boldsymbol{x} \mid \boldsymbol{\psi})$ and $f_2(\boldsymbol{x} \mid \boldsymbol{\phi})$ have nested supports so that $f_1(\boldsymbol{x} \mid \psi) > 0$ iff $\boldsymbol{x} \in \mathcal{X}_1(\Psi)$, $f_2(\boldsymbol{x} \mid \boldsymbol{\phi}) > 0$ iff $\boldsymbol{x} \in \mathcal{X}_2(\Phi)$ and either $\mathcal{X}_1(\Psi) \subset \mathcal{X}_2(\Phi)$ or $\mathcal{X}_2(\Phi) \subset \mathcal{X}_1(\Psi)$, the intrinsic divergence is still well defined and reduces to the logarithmic discrepancy, viz. $\delta(\boldsymbol{\phi}; \boldsymbol{\psi}) = k(\boldsymbol{\phi} \mid \boldsymbol{\psi})$ when $\mathcal{X}_1(\Psi) \subset \mathcal{X}_2(\Phi)$ and $\delta(\boldsymbol{\phi}; \boldsymbol{\psi}) = k(\boldsymbol{\psi} \mid \boldsymbol{\phi})$ when $\mathcal{X}_2(\Phi) \subset \mathcal{X}_1(\Psi)$ (for a thorough discussion, see Bernardo and Rueda, 2002; Juárez, 2004).

As mentioned above, the posterior expected intrinsic discrepancy, hereafter *the intrinsic statistic*,

$$d(\boldsymbol{\theta}_0 \mid \boldsymbol{x}) = \int_\Theta \int_\Lambda \delta(\boldsymbol{\theta}, \boldsymbol{\omega}; \boldsymbol{\theta}_0) \, \pi_\delta(\boldsymbol{\theta}, \boldsymbol{\lambda} \mid \boldsymbol{x}) \, \mathrm{d}\boldsymbol{\theta} \, \mathrm{d}\boldsymbol{\lambda}, \tag{1}$$

is a measure, in natural information units (nits; bits, if using $\log_2$), of the evidence against using $p(\boldsymbol{x} \mid \boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{\lambda})$, provided by the data $\boldsymbol{x}$. It is a monotonic test statistic for the (null) hypothesis, $H_0 \equiv \{\boldsymbol{\theta} = \boldsymbol{\theta}_0\}$ and thus induces the decision rule, hereafter the Bayesian reference criterion (BRC):

$$\text{Reject } \boldsymbol{\theta} = \boldsymbol{\theta}_0 \text{ iff } d(\boldsymbol{\theta}_0 \mid \boldsymbol{x}) > d^*.$$

In order to calibrate the threshold value, d^*, Bernardo and Rueda (2002) argue that the intrinsic statistic can be interpreted as the expected value of the log likelihood ratio against using the simplified model, given the data. Hence, values of d^* around 2.5 would imply a ratio of $e^{2.5} \approx 12$, providing mild evidence against the null; while values around 5 ($e^5 \approx 150$) can be regarded as strong evidence against H_0; values of $d^* \geq 7.5$ ($e^{7.5} \approx 1800$) can be safely used to reject the null.

As a natural consequence of the decision-theoretic approach taken here, the best approximation $p(\boldsymbol{x} \mid \boldsymbol{\theta}^*, \boldsymbol{\lambda})$ to the full model $p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\lambda})$, given the data, is given by that $\boldsymbol{\theta}^*$ which makes the intrinsic discrepancy loss as small as possible. Thus, it is natural to define (Bernardo and Juárez, 2003)

Definition 2 (Intrinsic estimator).

The intrinsic estimator, $\boldsymbol{\theta}^*$, *of* $\boldsymbol{\theta}$ *is the minimizer of the intrinsic statistic, i.e.*

$$\boldsymbol{\theta}^* = \boldsymbol{\theta}^*(\boldsymbol{x}) = \min_{\boldsymbol{\theta}_0 \in \Theta} d(\boldsymbol{\theta}_0 \mid \boldsymbol{x}).$$

Both the expected intrinsic discrepancy and the intrinsic estimator share a number of attractive properties (Bernardo and Rueda, 2002; Bernardo and Juárez, 2003): they are invariant under monotonic transformations of $\boldsymbol{\theta}$ and $\boldsymbol{x}$, and to choice of the nuisance parameter, are consistent with sufficient statistics, avoid the marginalisation paradoxes (Dawid *et al.*, 1973) and the expected intrinsic discrepancy avoids the JLB paradox. Moreover, in contrast with the vast majority of the methods proposed to deal with these problems, its application does not depend on the asymptotic behaviour of the model.

3 Examples

In this next section, we will restraint our attention to some simple non-regular models, applying the methodology described above and then comparing the results with some of the most common answers in the literature. In order to clarify the concepts introduced above, we will analyse in first place a one parameter, non-regular model.

3.1 Laplace Model

The Double-Exponential (Laplace) distribution, De $(x \mid 1, \theta)$, with pdf

$$p(x \mid \theta) = \frac{1}{2}\exp\Big[-|x-\theta|\Big],\ x \in \mathbb{R},\ \theta \in \mathbb{R},$$

does not belong to the exponential family and, thus, does not admit sufficient statistics of fixed dimension. In addition, the MLE is not necessarily unique. In fact, for $m = 1, 2, \ldots$

$$\hat{\theta} = \begin{cases} x_{((n+1)/2)} & n = 2m \\ \text{any point } \in \left\{x_{(n/2)}, x_{(n/2+1)}\right\} & n = 2m+1 \end{cases}.$$

On the other hand, the Kullback-Leibler divergence for one observation from this model,

$$\begin{aligned} k(\theta_2 \mid \theta_1) &= \int \exp[-|x-\theta_1|] \log \frac{\exp[-|x-\theta_1|]}{\exp[-|x-\theta_2|]} \, \mathrm{d}x \\ &= |\theta_1-\theta_2| + \exp\Big[-|\theta_1-\theta_2|\Big] - 1, \end{aligned}$$

turns out to be symmetric, from which

$$\delta(\theta;\theta_0) = |\theta-\theta_0| + e^{-|\theta-\theta_0|} - 1\,.$$

The intrinsic discrepancy is a piecewise one-to-one function of the parameter (Figure 1), permitting the use the reference prior when θ is the parameter of interest to calculate the intrinsic statistic. It is worth noticing that this model does not meet the regularity conditions needed to calculate Fisher matrix (see e.g. Schervish, 1995, p. 111), and thus Jeffreys prior is not defined for it.

By standard properties of reference priors it is easy to prove that $\pi(\theta) \propto 1$, and therefore

$$\pi(\theta \mid \boldsymbol{x}) \propto \exp\left[-\sum_{i=1}^{n} |x_i - \theta|\right].$$

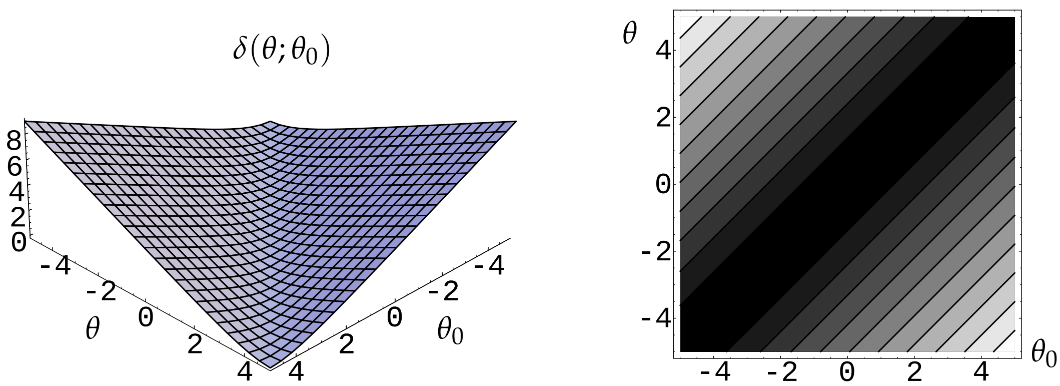


Figure 1. *The intrinsic discrepancy (left) and its contour plots for the Laplace model.*

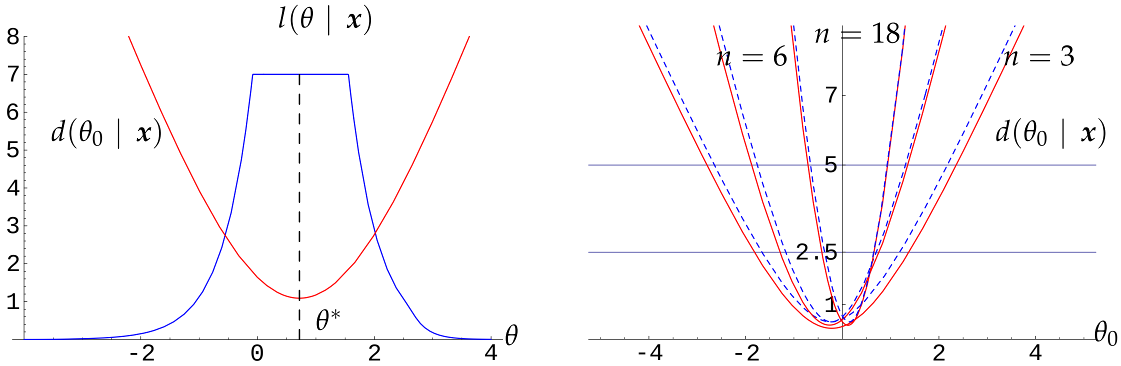
Given that the intrinsic discrepancy is a convex function of θ_0, the intrinsic estimator,

$$\theta^* = \arg\min_{\theta_0 \in \mathbb{R}} \int \delta(\theta;\theta_0)\, \pi(\theta \mid \boldsymbol{x})\, \mathrm{d}\theta,$$

is unique (DeGroot and Rao, 1963), and can be readily calculated by numerical methods. This is illustrated in Figure 2 (a). In this case, $\boldsymbol{x} = \{2.62, -0.08, 1.56, -3.14\}$, is a simulated sample and the non-uniqueness of the MLE is apparent from the shape of the likelihood function; however, $\theta^* = 0.374$.

The intrinsic statistic can be calculated by numerical methods and the approximation $d(\theta_0 \mid \boldsymbol{x}) \approx \delta(\check{\theta};\theta_0) + 1/2$, with $\check{\theta}$ any consistent estimator of θ (Juárez, 2004), works well even for moderated sample sizes, as illustrated in Figure 2 (b).

As stated in the Introduction, from the subjective point of view the Bayes rule is optimal for the problem at hand. Nonetheless, as the objective solution does not envisage any specific use, it may be sensible to evaluate the performance of the intrinsic solutions under repeated, homogenous sampling. To this end we simulated ten thousand data sets for each sample size, with $\theta = 0$. Then the p-value of the BRC was estimated as the relative number of times that $H_0 \equiv \{\theta = 0\}$ was rejected, using the threshold value $d^* = 2.5$. Regarding the intrinsic estimator, we calculated the relative number of times that it was closer to the true value of the parameter than the MLE, $\hat{\theta}$–with the convention that when n is even, the MLE is the mid point of $\{x_{(n/2)}, x_{(n/2+1)}\}$; also, the sample mean value and standard deviation



(a) Likelihood and intrinsic statistic for simulated sample of size $n = 4$.

(b) Intrisic statistic and the proposed approximation.

Figure 2. *Laplace Model. Depicted in the left pane is the (scaled) likelihood and the intrinsic statistic for the simulated sample of size $n = 4$, also the intrinsic estimator is pointed out. The right pane depicts the intrinsic statistic (solid) and the proposed approximation (dashed) calculated for simulated samples of different sizes and $\theta = 0$. The consistent estimator used for the approximation is $\check{\theta} = \bar{x}$.*

were calculated. The results are summarised in Table 1.

Table 1. *Estimated* p-*values for the* BRC *in the Laplace model, using the threshold value $d^* = 2.5$, calculated from 10,000 simulated repetitions for each sample size with $\theta = 0$. Also, the relative number of times (%) that θ^* was closer than $\hat{\theta}$ to the true value of the parameter, and the mean and (std. dev.) of both estimators for each sample size.*

n	p-value	%	θ^*		$\hat{\theta}$	
4	0.064	50.4	-2.78×10^{-4}	(0.633)	-1.05×10^{-3}	(0.638)
15	0.059	52.9	3.29×10^{-3}	(0.299)	2.72×10^{-3}	(0.314)
50	0.057	52.5	8.39×10^{-4}	(0.155)	1.24×10^{-3}	(0.159)
100	0.054	52.1	-1.34×10^{-3}	(0.106)	-8.37×10^{-4}	(0.108)

The second column exemplifies how the intrinsic statistic is an absolute measure of the weight of the evidence, brought by the data, against using the simplified model obtained when letting $\theta = \theta_0$. This is in contrast with conventional p-values that must be calibrated according to sample size and the dimension of the parameter. Summary statistics for both point estimators seem rather similar, suggesting that sampling properties of the intrinsic estimator are quite in line with those of the (conventional) MLE; θ^* appears to be consistently closer to the true value of the parameter, though.

3.2 Pareto and related models

One of the most appealing characteristics of the reference-intrinsic methodology is that both the intrinsic statistic and the intrinsic estimator are invariant under invertible transformations, they are also invariant under one-to-one transformations of the data and under the choice of the nuisance parameter. Here we will illustrate this point.

There is a vast literature concerning Pareto-like distributions, commencing with the work of Pareto (1897). It has been used to model a number quantities, including income, wealth, business size, city size, insurance excess loss, ozone levels in upper atmosphere (see e.g. Arnold, 1983; Arnold and Press, 1983, 1989). In addition to its many fields of application, a relevant feature of the classical Pareto pdf,

$$\mathrm{Pa}\,(x \mid \alpha, \beta) = \alpha\, \beta^{\alpha} x^{-(\alpha+1)}, \qquad x \geq \beta, \quad \alpha, \beta > 0\,,$$

is its relationship with other widely used models. For instance, let $x \sim \mathrm{Pa}\,(x \mid \alpha, \beta)$, then

1.- If $y = x^{-1}$, then $y \sim \mathrm{Ip}\,(y \mid \alpha, \beta_{ip})$, an inverted Pareto distribution; where $\beta_{ip} = \beta^{-1}$ and

$$\mathrm{Ip}\,(y \mid \alpha, \beta) = \alpha\, \beta^{-\alpha} y^{\alpha-1}, \qquad y \leq \beta, \quad \alpha, \beta > 0\,.$$

In particular, $\mathrm{Ip}\,(x \mid 1, \beta)$ is a uniform distribution $\mathrm{Un}\,(x \mid 0, \beta)$.

2.- If $y = -\log x$, then $y \sim \mathrm{Le}\,(y \mid \alpha, \beta_{le})$, a location-Exponential distribution; where $\beta_{le} = -\log \beta$ and

$$\mathrm{Le}\,(y \mid \alpha, \beta) = \alpha\, \exp\Big[-\alpha(y-\beta)\Big], \qquad y \geq \beta, \quad \alpha > 0,\, \beta \in \mathbb{R}\,.$$

The latter is also a Weibull distribution, $\mathrm{Wei}\,(y \mid 1, \alpha, \beta_{le})$, with

$$\mathrm{Wei}\,(y \mid \gamma, \alpha, \beta) = \gamma\, \alpha^{\gamma} (y-\beta)^{\gamma-1} \exp\Big[-\alpha^{\gamma}(y-\beta)^{\gamma}\Big]\,,$$

which also has a large history of different applications, particularly in reliability analysis (see e.g. Nelson, 1982).

In the above distributions, the shape parameter α remains the same, while the location parameter β_i, $i = ip, le$, is just a one-to-one transformation of the original location parameter β. Thus, from an objective viewpoint, it seems sensible that inferences about the parameters conducted with any of the transformed models should yield the same decisions as those drawn for the original one. This characteristic is apparent for the reference-intrinsic methodology, since for the Pareto model

$$\delta_{pa}(\alpha,\beta;\alpha_0,\beta_0) = \begin{cases} \log\dfrac{\alpha}{\alpha_0} + \dfrac{\alpha_0}{\alpha} - 1 + \alpha_0 \log\dfrac{\beta}{\beta_0} & \beta \geq \beta_0 \\ \log\dfrac{\alpha_0}{\alpha} + \dfrac{\alpha}{\alpha_0} - 1 + \alpha \log\dfrac{\beta_0}{\beta} & \beta \leq \beta_0 \end{cases}.$$

On the other hand, for the inverted Pareto model

$$\delta_{ip}(\alpha,\beta';\alpha_0,\beta'_0) = \begin{cases} \log\dfrac{\alpha}{\alpha_0} + \dfrac{\alpha_0}{\alpha} - 1 + \alpha_0 \log\dfrac{\beta'_0}{\beta'} & \beta' \leq \beta'_0 \\ \log\dfrac{\alpha_0}{\alpha} + \dfrac{\alpha}{\alpha_0} - 1 + \alpha \log\dfrac{\beta'}{\beta'_0} & \beta' \geq \beta'_0 \end{cases}.$$

Evidently, if we let $\beta' = 1/\beta$, both loss functions are identical. For the Weibull model,

$$\delta_{we}(\alpha,\beta'';\alpha_0,\beta''_0) = \begin{cases} \log\dfrac{\alpha}{\alpha_0} + \dfrac{\alpha_0}{\alpha} - 1 + \alpha_0(\beta'' - \beta''_0) & \beta'' \geq \beta''_0 \\ \log\dfrac{\alpha_0}{\alpha} + \dfrac{\alpha}{\alpha_0} - 1 + \alpha(\beta''_0 - \beta'') & \beta'' \leq \beta''_0 \end{cases};$$

and one can check that when $\beta'' = \log\beta' = -\log\beta$, the three functions are the same.

Recall that in this setting, the intrinsic discrepancy is the quantity of interest, and each of the above loss functions is a one-to-one transformation of any other. Given that the reference posterior is invariant under these transformations, it follows that the intrinsic statistic and thereof the intrinsic estimator are also invariant. We can then choose a model to work with and then apply the appropriate transformation to the final result. Let us choose the inverted Pareto model.

Suppose we have iid observations, $z = \{x_1, \ldots, x_n\}$ from an $\mathrm{Ip}\,(x_i \mid \alpha, \beta)$ distribution, the likelihood function is

$$L(\alpha, \beta) = \alpha^n \, \beta^{-n\alpha} \, t_1^{n\alpha - n}; \qquad \beta \geq t_2,$$

where $\{t_1, t_2\} = \left\{\prod x_i^{1/n}, x_{(n)}\right\}$ are jointly sufficient statistics for α and β. The MLE s

$$\hat{\beta} = t_2, \qquad\qquad \hat{\alpha} = \left(\log \frac{t_2}{t_1}\right)^{-1},$$

are (conditionally) independent, given the true parameter values (Malik, 1970) and

$$\hat{\beta} \mid \alpha, \beta \sim \mathrm{Ip}\,(\hat{\beta} \mid n\,\alpha, \beta)$$

and

$$\hat{\alpha} \mid \alpha \sim \mathrm{Ig}\,(\hat{\alpha} \mid n, n\,\alpha)\,,$$

where $\mathrm{Ig}\,(x \mid a, b)$ stands for an inverted Gamma distribution with parameters $\{a, b\}$.

3.2.1 The shape parameter

In order to apply the definitions in Section 2, we first assume that the shape parameter α is the parameter of interest. The intrinsic discrepancy is

$$\delta(\alpha; \alpha_0) = n \begin{cases} -\log\theta + \theta - 1 & \theta < 1 \\ \log\theta + \theta^{-1} - 1 & \theta \geq 1 \end{cases}$$

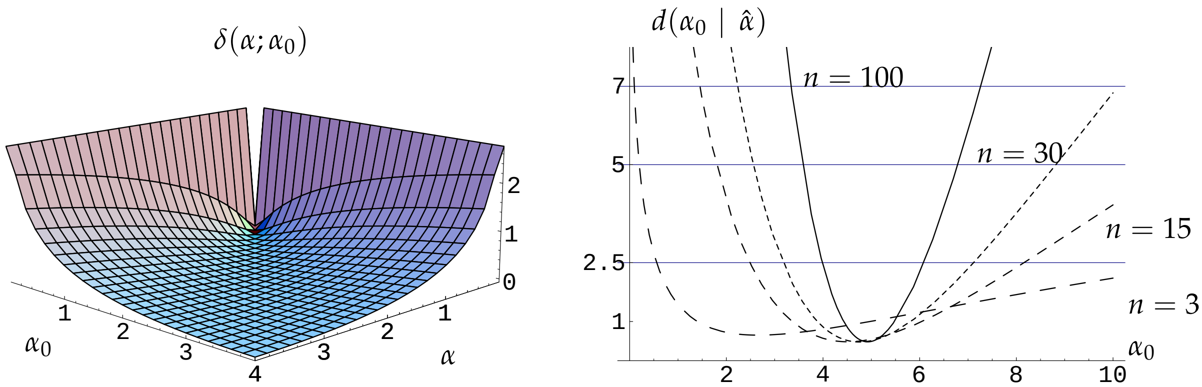
where $\theta = \alpha/\alpha_0$. As $\delta(\theta)$ does not depend on the nuisance parameter, β, and is a piecewise one-to-one function of α, we can derive the reference posterior for the ordered parametrization $\{\alpha, \beta\}$. This is found to be $\pi(\alpha, \beta) \propto (\alpha\,\beta)^{-1}$, and yields a (marginal) Gamma reference posterior, $\text{Ga}\left(\alpha \mid n-1, n\,\hat{\alpha}^{-1}\right)$, with pdf

$$\text{Ga}\left(\alpha \mid n-1, n\,\hat{\alpha}^{-1}\right) \propto \alpha^{n-2} \exp\left[-\frac{n}{\hat{\alpha}}\,\alpha\right].$$

The intrinsic statistic, for $n > 2$,

$$d(\alpha_0 \mid \hat{\alpha}) = \int_0^{\infty} \delta(\alpha; \alpha_0)\, \text{Ga}\left(\alpha \mid n-1, n\,\hat{\alpha}^{-1}\right)\, \mathrm{d}\alpha\,,$$

can be easily handled by numerical methods. Its typical behaviour, along with the intrinsic discrepancy, is depicted in Figure 3.



(a) Intrinsic discrepancy. (b) Intrinsic statistic.

Figure 3. *The intrinsic discrepancy –on the left– and the intrinsic statistic calculated for different sample sizes and $\hat{\alpha} = 5$, in the Inverted Pareto model.*

An analytic approximation to the intrinsic statistic (due to the asymptotic normality of the reference posterior), that works well even for moderated sample sizes is given by Juárez (2004) as

$$d(\alpha_0 \mid \hat{\alpha}) \approx \delta(\hat{\alpha}; \alpha_0) + \frac{1}{2};$$

while for the intrinsic estimator we can use

$$\alpha^* \approx \left(1 - \frac{3}{2n}\right)\hat{\alpha}.$$

3.2.2 The location parameter

Consider now the location parameter β. The intrinsic discrepancy is

$$\delta(\beta, \alpha; \beta_0) = \begin{cases} \log(1-\theta) & \theta < 0 \\ \theta & \theta > 0 \end{cases},$$

where $\theta = \alpha \log(\beta/\beta_0)$. Recalling the invariance under the choice of the nuisance parameter, we may use the reference prior for the ordered parametrization $\{\theta, \alpha\}$, $\pi(\theta, \alpha) \propto \alpha^{-1}$, or in terms of the original parametrization, $\pi(\beta, \alpha) \propto \beta^{-1}$; yielding the joint reference posterior, for $n > 2$,

$$\pi(\beta, \alpha \mid t_1, t_2) = \frac{n^{n+1}}{\hat{\alpha}^n \Gamma[n]} \alpha^n \beta^{-(n\alpha+1)} t_1^{n\alpha}; \qquad \alpha > 0, \beta \geq t_2.$$

The intrinsic statistic,

$$d(\beta_0 \mid z) = \int_0^\infty \left[\int_{t_2}^{\beta_0} \log\left(1 - \alpha \log \frac{\beta}{\beta_0}\right) \pi(\beta, \alpha \mid z)\, \mathrm{d}\beta \;+ \int_{\beta_0}^\infty \left(\alpha \log \frac{\beta}{\beta_0}\right) \pi(\beta, \alpha \mid z)\, \mathrm{d}\beta \right] \mathrm{d}\alpha,$$

as well as the intrinsic estimator, can be computed numerically. Nevertheless, we might obtain an analytical approximation by replacing α in the former equation by its MLE, $\hat{\alpha}$ and carrying out the one dimensional integration, obtaining

$$d(\beta_0 \mid z) \approx t\Big[1 + n\, e^{-n} \big(\mathrm{Ei}(n) - \mathrm{Ei}(n - \log t)\big)\Big] + n \log\big(1 + \frac{1}{n} \log t\big),$$

where $\mathrm{Ei}(x) = -\int_{-x}^\infty u^{-1} e^{-u}\, \mathrm{d}u$, and $t = \big(\hat{\beta}/\beta_0\big)^{n\hat{\alpha}}$ is the usual test statistic arising from the generalised likelihood ratio. Both the proposed approximation and the exact value of the intrinsic statistic, calculated for some simulated data, are depicted in Figure 4.

A simple way to obtain an analytic approximation to the intrinsic estimator,

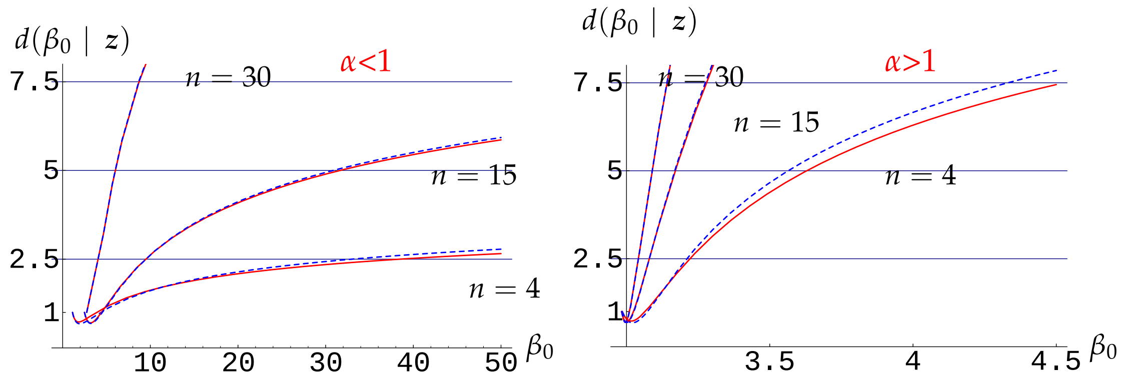


Figure 4. *The intrinsic statistic and the proposed approximation (dashed). In the left pane the behaviour for $\alpha < 1$ and in the right pane for $\alpha > 1$; both are calculated from simulated data of different sizes with $\beta = 3$.*

is plugging $\hat{\alpha}$ into the conditional reference posterior $\pi(\beta \mid \alpha, z) = \text{Pa}\,(\beta \mid n\,\alpha, t_2)$, and into the intrinsic divergence and then carry out the one-dimensional integration-optimisation , resulting in

$$\beta^*_{apr} = 2^{\frac{1}{n\,\hat{\alpha}}}\,\hat{\beta}.$$

3.2.3 Comparisons

Due to the asymmetry generated by the shape parameter, it seems reasonable that, for the same sample size, we should have less precise estimates of β when $\alpha < 1$ than when $\alpha > 1$. Table 2 presents the results obtained from simulated data for a small sample size and two values for the shape parameter. The behaviour of the intrinsic estimator seems to capture the increased variability brought by the different values of shape parameter better than the MLE. Also, the estimated p-value for the BRC using the threshold value $d^* = 2.5$ is presented.

Table 2. *Mean value and standard deviation of the intrinsic estimator, the proposed approximation and the maximum likelihood estimator of the location parameter of the Inverted Pareto model, calculated from 5000 simulated samples holding $\beta = 3$ and $n = 4$.*

	$\alpha = 1/5$			$\alpha = 5$		
	β^*	β^*_{apr}	$\hat{\beta}$	β^*	β^*_{apr}	$\hat{\beta}$
Mean	2.70	2.84	1.34	2.94	2.93	2.86
Std. dev.	2.3	2.5	0.89	0.14	0.14	0.13
p-value	0.09			0.09		

CRiSM Paper No. 05-14, www.warwick.ac.uk/go/crism

For the sake of comparison, assume now that the shape parameter is known and, without loss of generality, $\alpha = 1$. It is readily seen that then $x \sim \text{Un}\,(x \mid 0, \beta)$ and, therefore, $-\log x = y \sim \text{Le}\,(y \mid 1, \beta')$. Each of the correspondent intrinsic discrepancies,

$$\delta_x(\beta; \beta_0) = \left| \log \frac{\beta}{\beta_0} \right|$$

and

$$\delta_y(\beta'; \beta'_0) = |\beta' - \beta'_0|\,,$$

is a piecewise one-to-one function of the implied parameter. Thus, we can use $\pi(\beta) \propto \beta^{-1}$ and $\pi(\beta') \propto 1$, to derive the appropriate reference posteriors, which are $\text{Pa}\,(\beta \mid n, t)$ and $\text{Le}\,(\beta' \mid n, t')$, respectively, where $t = \max\{x_1, \ldots, x_n\}$ and $t' = -\log t = \min\{y_1, \ldots, y_n\}$.

The calculation of the intrinsic statistics is straightforward,

$$\begin{aligned} d(\beta_0 \mid t) &= \int_t^\infty \left| \log \frac{\beta}{\beta_0} \right| \text{Pa}\,(\beta \mid n, t)\; \mathrm{d}\beta \\ &= 2\tau - \log \tau - 1 \end{aligned}$$

and

$$\begin{aligned} d(\beta_0 \mid t') &= \int_{-\infty}^{t'} |\beta' - \beta'_0|\; \text{Le}\,(\beta' \mid n, t')\; \mathrm{d}\beta' \\ &= 2\tau' - \log \tau' - 1\,, \end{aligned}$$

where $\tau = \tau(\beta_0, t) = (t/\beta_0)^n$ and $\tau' = \tau(\beta_0, t) = \exp[n(\beta'_0 - t')]$ are the test statistics derived from the generalised likelihood ratio. The invariance of the intrinsic statistic is apparent, it is straightforward to verify that $d(\beta'_0 \mid t') = d(-\log \beta_0, -\log t)$. Consequently, the intrinsic estimator is also invariant, thus we have $\beta^*(t) = 2^{1/n} t$ and $\beta'^*(t') = -\log \beta^* = t' - \log 2^{1/n}$.

Interestingly, under homogeneous repeated sampling, both τ and τ' follow a $\text{Un}\,(y \mid 0, 1)$ distribution under the null $H_0 \equiv \{\beta = \theta_0\ (\beta' = \beta'_0)\}$. Similarly, from the Bayesian viewpoint, $\tau_B = (t/\theta)^n$ and $\tau'_B = \exp[n(\beta' - t')]$ follow the same Uniform distribution, yielding the correspondence between the frequentist p-values and the Bayesian Reference Criterion, exhibited in Table 3.

Table 3. *Some* p-*values,* $P[d > d^\star \mid H_0]$*, associated to the corresponding threshold values,* d^**, from the* BRC *for the Uniform model.*

$d^\star$	$P[d > d^\star \mid H_0]$	$d^\star$	$P[d > d^\star \mid H_0]$
1	0.203	**5**	**0.00278**
2	0.054	6	0.00097
2.5	**0.031**	7	0.00038
3	0.019	8	0.00018
4	0.007	9	0.00008

The risk function $R_\beta(c)$ in the Uniform model, for estimators of the form $\tilde\theta = c\,t$, with $c > 1$, under the intrinsic discrepancy loss function is

$$\begin{aligned} R_\theta(c) &= \int_0^\theta n \left|\log \frac{\theta}{c\,t}\right| n\,\theta^{-n} t^{n-1}\,\mathrm{d}t \\ &= 2\,c^{-n} + n\log c - 1. \end{aligned}$$

Straightforward calculations show that $R_\theta(c)$ attains a global minimum at $c = 2^{1/n}$, i.e. the intrinsic estimator is the only admissible estimator under the intrinsic discrepancy loss. This same holds for β'^* in the location-Exponential model for estimators within the class $\mathcal{C} = \{\tilde\beta \,:\, \tilde\beta = t + c, c < 0\}$.

The Location-Exponential model is used by Berger and Pericchi (2001) to illustrate the use of the intrinsic Bayes factors (IBF) for non-regular cases. There, the authors compute the arithmetic and the median intrinsic Bayes factors (AIBF and MIBF, respectively) as

$$\text{AIBF} = B\,\frac{1}{n}\sum_{i=1}^{n}\left[\exp(y_i - \beta'_0) - 1\right]^{-1}$$

and

$$\text{MIBF} = B\,\left[\exp(\mathrm{Med}[\boldsymbol{y}] - \beta'_0) - 1\right]^{-1},$$

where $B = n^{-1}\{\exp[n(t' - \beta'_0)] - 1\}$. They also argue that the fractional Bayes factor (O'Hagan, 1997),

$$\text{FBF} = B\, b\, n\, \left\{\exp\left[b\, n\, (t' - \beta'_0)\right] - 1\right\}^{-1},$$

is clearly unreasonable, given that FBF > 1 for any $0 < b < 1$.

Even though both IBF's are defined in this case, their behaviour under homogeneous repeated sampling is awkward. To illustrate this we simulated 100,000 sets of different sizes from Le $(x \mid 1, -0.1)$ and then computed the relative number of times that the (null) hypothesis $H_0 \equiv \{\beta = -0.1\}$ was rejected. As we can see from Table 4, p-values arising from alternative IBF s vary widely with sample size, while those computed from the BRC and the frequentist test behave as expected.

Table 4. *Estimated* p*-values corresponding to comparable test sizes for the* BRC *($d^* = 3$), the frequentist test ($\alpha = 0.05$) and the* IBF*'s ($B \geq 20$) calculated from simulated values (100,000 replications) from the Location-Exponential model for several sample sizes and* $\beta = -0.1$

n	AIBF	MIBF	BRC	FREQ
3	0.998	0.889	0.018	0.049
10	0.997	0.675	0.018	0.050
25	0.938	0.448	0.019	0.049
100	0.525	0.165	0.018	0.050
1000	0.097	0.019	0.019	0.049

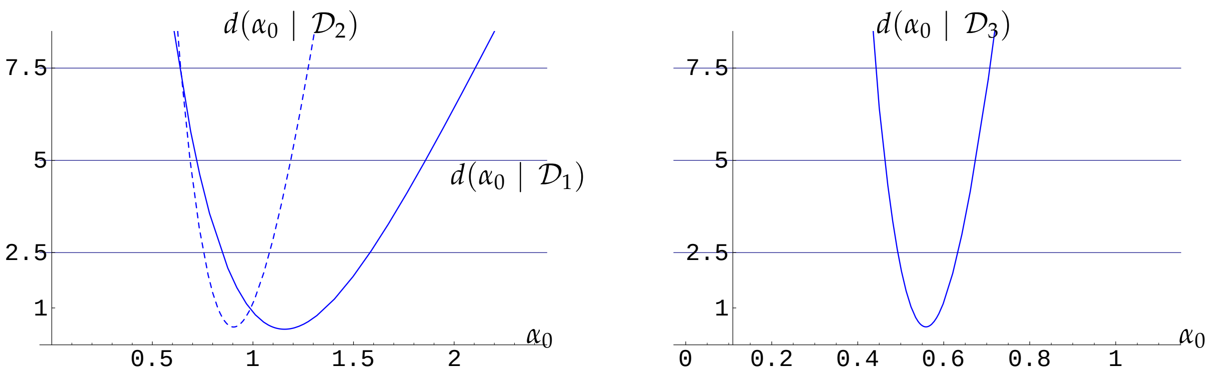
Another anomaly is that it is possible to reach conflicting decisions from alternative IBF for small samples. For instance, consider the simulated data set from a Un $(x_i \mid 0, 1)$, $\boldsymbol{x} = \{0.01, 0.05, 0.50, 0.99\}$, and the transformed data when making $y = -\log x$. Table 5 shows the alternative Bayes factors and the intrinsic statistic calculated for testing the (null) hypothesis $H_0 \equiv \{\beta = 1\ (\beta' = 0)\}$. Given the invariance properties of the intrinsic statistic, clearly $d(1 \mid t) = d(0 \mid t')$ and thus decisions derived from both data sets are the same, for they convey exactly the same amount of information about the parameter. In contrast, different decisions can be reached for alternative IBF s.

Table 5. *Alternative intrinsic Bayes factors and the intrinsic statistic to test the hypothesis* $\beta = 1$ $(\beta' = 0)$*, calculated for the simulated data, from a* $\text{Un}\,(x \mid 0, \beta)$*,* $\boldsymbol{x} = \{0.01, 0.05, 0.5, 0.99\}$ *and for the transformed data* $y = -\log x$*, where* $\beta' = -\log \beta$*.*

	AIBF_{01}	MIBF_{01}	d
β	5.05	13.97	0.98
β'	3.89	519.21	0.98

3.2.4 USA metropolitan areas

As mentioned at the beginning of the section, Pareto distributions are often used to model city population sizes. Zipf (1949) formally established that, within a given country, the size of the k largest cities is inversely proportional to its rank; this regularity implies that the distribution of the population size of these cities, s, is $\text{Pa}\,(s \mid \alpha, \beta)$, with shape parameter equal to one. To verify Zipf's law, we analyse the 276 USA metropolitan areas (MA) population, based on the 2000 census (`www.census.gov/main/www/cen2000.html`). Three nested data sets were considered: data set $\mathcal{D}_1$ contains the 50 largest MA's; data set $\mathcal{D}_2$ contains the 135 largest MA's; and data set $\mathcal{D}_3$ is the whole sample. The intrinsic estimators for each case are $\alpha_1^* = 1.2$, $\alpha_2^* = 0.914$ and $\alpha_3^* = 0.562$; intrinsic statistics for each data set are depicted in Figure 5. From these we can state that Zipf's law holds for the first two data sets, since $d(1 \mid \mathcal{D}_1) = 0.88$ and $d(1 \mid \mathcal{D}_2) = 1.26$; while $d(1 \mid \mathcal{D}_3) = 39.15$, providing overwhelming evidence against $\alpha = 1$, for the whole of USA MA's (for a thorough discussion on this phenomenon see Eeckhout, 2004).



(a) The 50 and the 135 largest USA MA's

(b) All USA MA's

Figure 5. *The intrinsic statistic for the three USA metropolitan areas data sets analysed.*

3.3 The change-point problem

The change-point problem has an log history, dating back to Page (1955, 1957), and has been addressed in different ways by Carlin *et al.* (1992); Hinkley (1970); Smith (1975), i.a. In general, the problem is to be able to discriminate in a sequence of independent observations, $\boldsymbol{x} = \{x_1, \ldots, x_n\}$, if all members are drawn from a common distribution, $p(x)$, or if there exists a point, r, for which the first r observations come from $p_1(x)$ and the rest from $p_2(x)$.

This problem may be addressed in two alternative ways:

Retrospective Consider the sequence of observations, $\boldsymbol{x} = \{x_1, \ldots, x_n\}$, as a realisation of a concrete process. Determine if there exists a change point, $1 \leq r < n$ (see, e.g. Chernoff and Zacks, 1964; Beibel, 1996 and references therein). Or

Sequential Consider the sequence $\boldsymbol{x} = \{x_1, x_2, \ldots, x_r, x_{r+1}, \ldots\}$. Determine, as soon as possible, if a change has occurred at point r (see, e.g. Shiryayev, 1963; Lorden, 1971 and references therein).

3.3.1 Page's artificial data

Here we focus on the retrospective approach and analyse the artificial data of Page (1957). Thus, we will assume that $p_1(x) = \mathrm{N}(x \mid 5, 1)$ and $p_2(x) = \mathrm{N}(x \mid 6, 1)$, so the likelihood function is

$$L(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \begin{cases} \prod_{i=1}^{r} \mathrm{N}(x \mid 5, 1) \prod_{j=r+1}^{n} \mathrm{N}(x \mid 6, 1) & 1 \leq r < n \\ \prod_{i=1}^{n} \mathrm{N}(x \mid 5, 1) & r = n \end{cases} . \tag{2}$$

For this simple setting, the intrinsic discrepancy is the linear loss function,

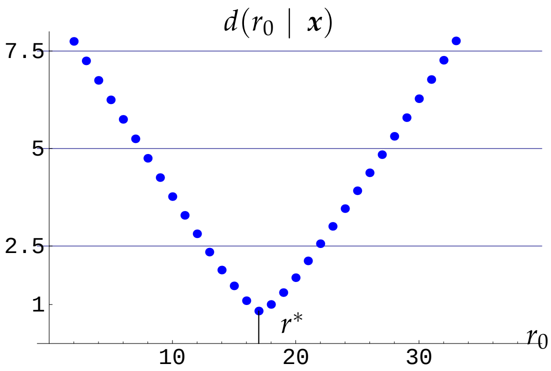
$$\delta(r; r_0) = \frac{1}{2} |r - r_0| .$$

Hence, the intrinsic estimator is just the posterior median. As $\delta(r; r_0)$ is a piecewise one-to-one function of r, we may use the reference prior $\pi(r) = (n-1)^{-1}$ to derive the appropriate posterior, which is proportional to the likelihood.

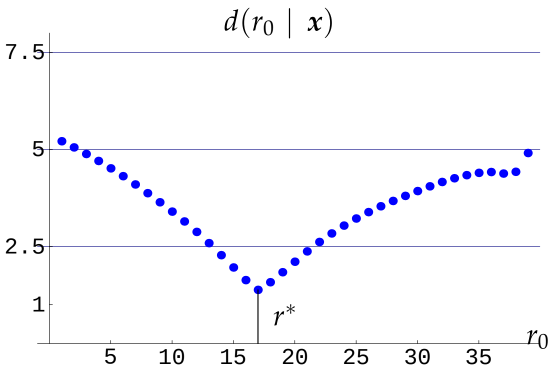
If both means are unknown, i.e. $p_1(x) = \mathrm{N}(x \mid \mu, 1)$ and $p_2(x) = \mathrm{N}(x \mid \eta, 1)$, we get

$$\delta(r, \mu, \eta; r_0) = \frac{1}{2}(\mu - \eta)^2 |r - r_0| \begin{cases} \dfrac{r}{r_0} & r \leq r_0 \\ \dfrac{n-r}{n-r_0} & r \geq r_0 \end{cases},$$

from where it is possible to derive the appropriate reference posterior and then calculate the intrinsic statistic and estimator. Figure 6 depicts the intrinsic statistic for both scenarios. We can see that $r^* = 17$, irrespective of the known means assumption, coinciding with Page's analysis. If the means are known, we could say that the change took place within the 13th and the 21st observations, and we could be quite sure that no change took place outside the 7th and 27th observations. On the other hand, if the means are unknown, we could say that a change occurred between the 15th and 21st observations and were almost sure that no change took place before the second one. In any case, the true change point, at the 21st observation, is effectively detected.



(a) Known means.



(b) Unknown means.

Figure 6. *The intrinsic statistic for the change-point problem, calculated for Page's data.*

3.3.2 River Nile data

Consider the measurements of the annual volume of discharge from the Nile River to Aswan from 1871–1970, first analysed by Cobb (1978) and further examined from a nonparametric perspective by Carlstein (1988) and Dümbgen (1991). Assuming that $p_1(x) = \mathrm{N}(x \mid \mu, \lambda)$ and that $p_2(x) = \mathrm{N}(x \mid \eta, \lambda)$, with a common unknown precision, λ, the intrinsic discrepancy and a corresponding reference prior are

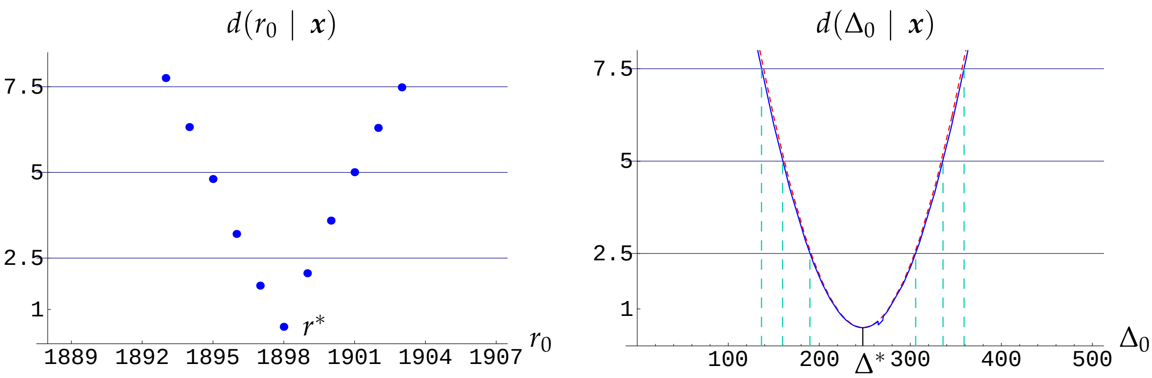
$$\delta(r, \theta; r_0) = \frac{n}{2} \begin{cases} \log\left[1 + \dfrac{r(r_0 - r)}{r\, r_0}\, \theta^2\right] & r \leq r_0 \\ \log\left[1 + \dfrac{(n-r)(r-r_0)}{n(n-r_0)}\, \theta^2\right] & r \geq r_0 \end{cases}$$

and

$$\pi(r, \theta) \propto \lambda^{-1}\left(1 + \frac{r(n-r)}{2n^2}\, \theta^2\right)^{-1/2},$$

respectively, where $\theta = \lambda^{1/2}(\mu - \eta)$ is the standardised distance between the two means.

In this case, $r^* = 1898$, which coincides with the three analyses mentioned above. In addition, the intrinsic change-point regions corresponding to the proposed threshold values are $R_{2.5} = \{1897, 1899\}$, $R_5 = \{1895, 1900\}$ and $R_{7.5} = \{1894, 1903\}$. If we carry the analysis one step further, and perform inference on the difference of means, $\Delta = \mu - \eta$, conditional on $r = 1898$, the intrinsic estimator of the change in the mean (which is also the MLE, see Juárez, 2004) is $\Delta^* = 247.78$ and the corresponding non-rejection regions are $R_{2.5} = \{189.69, 305.87\}$, $R_5 = \{159.65, 335.91\}$ and $R_{7.5} = \{136.50, 359.05\}$. These results are summarised in Figure 7.



(a) Change-point. (b) Difference in means.

Figure 7. *River Nile data. In the left pane, the intrinsic statistic for the change-point. The right pane depicts the intrinsic statistic for the difference in means, conditional on* $r = r^* = 1898$.

We have already mentioned that FBF does not perform well in non-regular problems. The IBF requires a training sample to convert the initial improper prior into a proper one. However, no minimal training sample exists for the (retrospective) change-point problem, thus an ad-hoc modification is in needed (Moreno *et al.*, 2003). In contrast, the reference-intrinsic methodology needs no modification and renders sensible results.

4 Conclusions

The reference-intrinsic methodology provides objective Bayesian decision rules for the precise hypothesis testing and point estimation problems, objective in the precise sense of depending on the data and the sampling model alone. In addition to this, the point stressed in this paper is that the presented method needs no modification, regardless the regularity conditions of the sampling model and the dimension of the parameter.

When testing sharp hypotheses, the use of conventional Bayes factors relies on a particular prior with a point mass on the null, leading to the so called JLB paradox. Alternative Bayes factors are still open to criticism. We have seen how fractional Bayes factors fail when dealing with non-regular models and it can be shown that other problems arise in the presence of nuisance parameters. As for intrinsic Bayes factors, it may be the case (as the AIBF in the exponential location model) that the resulting factor comes from no prior, thus being not really Bayesian; or that one of the Bayes Factors (the MIBF in that same example) be biased in favor of one of the hypothesis. Moreover, it may also happen (like the AIBF) that the resulting Bayes Factor depends on the whole sample, even when sufficient statistics are available, thus violating the sufficiency principle. We have also seen that the intrinsic statistic is typically a one-to-one function of the test statistic derived from the generalised likelihood ratio, providing a link between the frequentist test and the BRC; indeed, when this is the case, the BRC may be seen as a way to calibrate the p-values.

One of the most compelling features of the intrinsic estimator, shared by the MLE, is its invariance under monotonic transformations even when the dimension of the parametric space is greater than one, a characteristic not shared by the most frequently used objective Bayesian point estimators. Additionally, as illustrated in the examples, it is a consistent estimator of the parameter and, hence, agrees asymptotically with the MLE–when this exists–, accounting for the increase in uncertainty when nuisance parameters are present in the sampling model, while typically exhibiting compelling properties under repeated, homogeneous sampling.

References

Arnold, B. C. (1983), *Pareto Distributions*, USA: International Co-operative Publishing House.

Arnold, B. C. and Press, S. J. (1983), Bayesian inference for Pareto populations, *Journal of Econometrics*, **21**, 287–306.

Arnold, B. C. and Press, S. J. (1989), Bayesian estimation and prediction for Pareto data, *J. Amer. Statist. Assoc.*, **84**, 1079–1084.

Bartlett, M. S. (1957), Comment on "A statistical paradox" by D. V. Lindley, *Biometrika*, **44**, 533–534.

Beibel, M. (1996), A note on Ritosv's Bayes approach to the minimax property of the cusum procedure, *The Annals of Statistics*, **24**, 1804–1812.

Berger, J. O. and Bernardo, J. M. (1992), On the development of reference priors, *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.), Oxford: University Press, pp. 35–60.

Berger, J. O. and Pericchi, L. R. (2001), Objective Bayesian methods for model selection: Introduction and comparison, *Model Selection*, *Lecture Notes*, vol. 38 (P. Lahiri, ed.), Institute of Mathematical Statistics, pp. 135–207, (with discussion).

Berger, J. O. and Selke, T. (1987), Testing a point null hypothesis: The irreconcilability of p-values and evidence, *J. Amer. Statist. Assoc.*, **82**, 112–139, (with discussion).

Bernardo, J. M. (1979), Reference posterior distributions for Bayesian inference, *J. Roy. Statist. Soc. B*, **41**, 113–147.

Bernardo, J. M. (1982), Contraste de modelos probabilísticos desde una perspectiva bayesiana, *Trabajos de Estadística*, **33**, 16–30.

Bernardo, J. M. (1985), Análisis bayesiano de los contrastes de hipótesis paramétricos, *Trabajos de Estadística*, **36**, 45–54.

Bernardo, J. M. (1999), Nested hypothesis testing: The Bayesian Reference Criterion, *Bayesian Statistics 6* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.), Oxford: University Press, pp. 101–130.

Bernardo, J. M. and Bayarri, M. J. (1985), Bayesian model criticism, *Model Choice* (J. P. Florens, M. Mouchart, J. P. Raoult and L. Simar, eds.), Bruxelles: Pub. Fac. Univ. Saint Louis, pp. 43–59.

Bernardo, J. M. and Juárez, M. A. (2003), Intrinsic estimation, *Bayesian Statistics 7* (J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West, eds.), Oxford: University Press, pp. 465–475.

Bernardo, J. M. and Rueda, R. (2002), Bayesian hypothesis testing: A reference approach, *International Statistical Review*, **70**, 351–372.

Carlin, B. P., Gelfand, A. E. and Smith, A. F. M. (1992), Hierarchical Bayesian analysis of changepoint problems, *Appl. Statist.*, **41**, 389–405.

Carlstein, E. (1988), Nonparametric change-point estimation, *The Annals of Statistics*, **16**, 188–197.

Chernoff, H. and Zacks, S. (1964), Estimating the current mean of a Normal distribution which is subjected to changes in time, *The Annals of Statistics*, **35**, 999–1018.

Cobb, G. W. (1978), The problem of the Nile: Conditional solution to a changepoint problem, *Biometrika*, **65**, 243–251.

Dawid, A. P., Stone, M. and Zidek, J. V. (1973), Marginalization paradoxes in Bayesian and structural inference, *J. Roy. Statist. Soc. B*, **35**, 189–223, (with discussion).

DeGroot, M. H. and Rao, M. M. (1963), Bayes estimation with convex loss, *The ann. of Math. Stat.*, **34**, 839–846.

Dümbgen, L. (1991), The asymptotic behavior of some nonparametric change-point estimators, *The Annals of Statistics*, **19**, 1471–1495.

Edwards, W., Lindman, H. and Savage, L. J. (1963), Bayesian statistical inference for psychological research, *Psychological Review*, **70**, 193–242.

Eeckhout, J. (2004), Gibrat's law for (all) cities, *American Economic Review*, **94**, 1429–1451.

Ferrándiz, J. R. (1985), Bayesian inference on Mahalanobis distance: an alternative approach to Bayesian model testing, *Bayesian Statistics 2* (J. M. Bernardo, M. H. D. Groot, D. V. Lindley and A. F. M. Smith, eds.), Amsterdam: North-Holland, pp. 645–654.

Gutiérrez-Peña, E. (1992), Expected logarithmic divergence for exponential families, *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.), Oxford university press, pp. 669–674.

Hinkley, D. V. (1970), Inference about the change-point in a sequence of random variables, *Biometrika*, **57**, 1–17.

Jeffreys, H. (1961), *Theory of Probability*, 3rd ed., Oxford: University Press.

Juárez, M. A. (2004), *Objective Bayesian methods for estimation and hypothesis testing*, Unpublished PhD thesis, Universitat de Valencia.

Kullback, S. (1968), *Information Theory and Statistics*, New York: Dover.

Kullback, S. and Leibler, R. A. (1951), On information and sufficiency, *Ann. Math. Stat.*, **22**, 79–86.

Lindley, D. V. (1957), A statistical paradox, *Biometrika*, **44**, 187–192.

Lorden, G. (1971), Procedures for reacting to a change in distribution, *Ann. Math. Statis.*, **42**, 1897–1908.

Malik, H. J. (1970), Estimation of the parameters of the Pareto distribution, *Metrika*, **15**, 126–132.

Moreno, E., Casella, G. and García-Ferrer, A. (2003), ObjectiveBayesian analysis of the changepoint problem, *Working Paper 06-03*, Universidad Autónoma de Madrid.

Nelson, W. (1982), *Applied life data analysis*, New York: Wiley.

O'Hagan, A. (1995), Fractional Bayes Factors for model comparison, *J. Roy. Statist. Soc. B*, **57**, 99–138.

O'Hagan, A. (1997), Properties of intrinsic and fractional Bayes factors, *Test*, **6**, 101–118.

Page, E. S. (1955), A test for a change in a parameter occurring at an unknown point, *Biometrika*, **42**, 523–527.

Page, E. S. (1957), On problems in which a change in a parameter occurs at an unknown point, *Biometrika*, **44**, 248–252.

Pareto, V. (1897), *Cours d'economie Politique, II*, Lausanne: F. Rouge.

Robert, C. P. (1996), Intrinsic losses, *Theory and Decisions*, **40**, 191–214.

Robert, C. P. and Caron, N. (1996), Noninformative Bayesian testing and neutral Bayes factors, *Test*, **5**, 411–437.

Rueda, R. (1992), A Bayesian alternative to parametric hypothesis testing, *Test*, **1**, 61–67.

Schervish, M. J. (1995), *Theory of Statistics*, New York: Springer.

Selke, T., Bayarri, M. J. and Berger, J. O. (2001), Calibration of p-values for testing precise null hypotheses, *The American Statistician*, **55**, 62–71.

Shiryayev, A. N. (1963), On optimum methods in quickest detection problems, *Theory Probab. Appl.*, **8**, 22–46.

Smith, A. F. M. (1975), A Bayesian approach to inference about a change-point in a sequence of random variables, *Biometrika*, **62**, 407–416.

Zipf, G. K. (1949), *Human behavior and the principle of least effort*, Cambridge, MA: Addison-Wesley.