Objective Bayesian Methods for Estimation and Hypothesis Testing

SUMMARY

Miguel Ángel Juárez Hermosillo[1]

Advisor: Dr. José Miguel Bernardo Herranz

[1] This PhD thesis was done thanks to the support of Conacyt, México

Chapter 1

METHODS

This chapter briefly describes some of the most often encountered methods in the literature for estimating and testing.

Point Estimation

From a frequentist perspective, the methods of maximum likelihood (MLE), moments, least squares and unbiased estimation are described. Some shortcomings of these methods are highlighted, as not existence or not uniqueness, lack of invariance, lack of definition in the boundaries of the parameter space and lack of dependance on sufficient statistics, among others.

From a Bayesian standpoint, the problem is described as a decision one, where the action space is the parameter space; and the Bayes estimator is that rule which minimises the posterior expected loss. To obtain an objective solution, the reference algorithm (Berger and Bernardo, 1992*a*; Bernardo, 1979*b*), to derive an objective prior, is adopted; but the absence of a similar technique to derive an objective loss function is pointed out. Furthermore, it is argued that invariance of the Bayes rule is a main feature for an objective estimator and that this property is not exhibited by the bulk of conventional Bayes estimators.

Hypothesis Testing

The frequentist Neyman and Pearson (1933) (NP) test size and the Fisherian p-value are described. Well known criticisms to both methodologies are stated; for instance, the important amount of information that NP method potentially leaves aside, the need of calibrating p-values accordingly to sample and dimension sizes, or the arbitrariness in the selection of a test statistic.

Conventional Bayes factors are argued to be the result of a specific decision theoretic setup, with a no-continuous prior and a $0 - K_i$ loss function; which need not to be objective.

Alternative solutions, such as fractional Bayes factors (FBF) (O'Hagan, 1997) and intrinsic Bayes factors (IBF) (Berger and Pericchi, 2001) are also described. Objections such as the FBF not being useful in no-regular problems, or the exposure of the IBF to the Jeffreys-Lindley-Bartlett paradox and its lack of dependance on sufficient statistics are mentioned.

Chapter 2

INTRINSIC DISCREPANCY

Some measures of discrepancy between two probability densities are investigated in the first part of the chapter. In the second part we explore the properties of the *intrinsic losses* as introduced by Robert (1996). In the last part an objective *intrinsic discrepancy* (Bernardo and Rueda, 2002) is defined, its properties explored and it is advocated to be an objective loss, proper for point estimation and hypothesis testing problems.

Measures of divergence

In first place we follow the classification of Ali and Silvey (1966), and consider that a measure of divergence between two probability distributions, $p_i(\boldsymbol{x})$, $i=1,2$, should be based on the likelihood ratio $\phi(\boldsymbol{x}) = p_1(\boldsymbol{x})/p_2(\boldsymbol{x})$. Then, we study the following measures:

Kullback-Leibler divergence Also known as logarithmic divergence or directed divergence,

$$k(p_j \mid p_i) = \int p_i(\boldsymbol{x}) \log \frac{p_i(\boldsymbol{x})}{p_j(\boldsymbol{x})} \, \mathrm{d}\boldsymbol{x} \, .$$

If the distributions belong to a parametric family of models, indexed by a parameter: $p_1(\boldsymbol{x}) = \{p(\boldsymbol{x} \mid \boldsymbol{\theta}), \boldsymbol{x} \in \mathscr{X}_1, \boldsymbol{\theta} \in \Theta\}$, $p_2(\boldsymbol{x}) = \{p(\boldsymbol{x} \mid \boldsymbol{\psi}), \boldsymbol{x} \in \mathscr{X}_2, \boldsymbol{\psi} \in \Psi\}$, the KL divergence can be written as

$$k(\boldsymbol{\psi} \mid \boldsymbol{\theta}) = \int_{\mathscr{X}_1} p(\boldsymbol{x} \mid \boldsymbol{\theta}) \log \frac{p(\boldsymbol{x} \mid \boldsymbol{\theta})}{p(\boldsymbol{x} \mid \boldsymbol{\psi})} \, \mathrm{d}\boldsymbol{x}$$
$$k(\boldsymbol{\theta} \mid \boldsymbol{\psi}) = \int_{\mathscr{X}_2} p(\boldsymbol{x} \mid \boldsymbol{\psi}) \log \frac{p(\boldsymbol{x} \mid \boldsymbol{\psi})}{p(\boldsymbol{x} \mid \boldsymbol{\theta})} \, \mathrm{d}\boldsymbol{x} \, .$$

The KL divergence is

- Positive and bounded from below; i.e. $k(p_j \mid p_i) \geq 0$ and $k(p_j \mid p_i) = 0$ iff $p_i(\boldsymbol{x}) = p_j(\boldsymbol{x})$, a.e.

- Additive for conditionally independent observations. If $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ are independent observations from either $p_i(\boldsymbol{x})$ or $p_j(\boldsymbol{x})$, then $k_{\boldsymbol{x}}(p_j \mid p_i) = \sum_{l=1}^{n} k_{x_l}(p_j \mid p_i)$.
- Compatible with sufficient statistics. If $\boldsymbol{t} = \boldsymbol{t}(\boldsymbol{x})$ is a transformation of the original data, then $k_{\boldsymbol{x}}(p_i \mid p_j) \geq k_{\boldsymbol{t}}(p_i \mid p_j)$ with equality iff $\boldsymbol{t}$ is sufficient.
- Invariant under one to one transformations. Let $\{p(\boldsymbol{x} \mid \boldsymbol{\theta}),\ \boldsymbol{x} \in \mathscr{X}(\boldsymbol{\theta}),\ \boldsymbol{\theta} \in \Theta\}$ be a model and let $\boldsymbol{\phi} = \boldsymbol{\phi}(\boldsymbol{\theta})$ be a one-to-one transformation, then

$$k_{\boldsymbol{x}}(\boldsymbol{\phi}_j \mid \boldsymbol{\phi}_i) = k_{\boldsymbol{x}}\left(\boldsymbol{\theta}^{-1}(\boldsymbol{\phi}_j) \mid \boldsymbol{\theta}^{-1}(\boldsymbol{\phi}_i)\right).$$

$\mathscr{J}$**-divergence** This symmetrized version of the KL divergence,

$$\mathscr{J}(p_1, p_2) = \frac{1}{2} \int \left(p_1(\boldsymbol{x}) - p_2(\boldsymbol{x})\right) \log \frac{p_1(\boldsymbol{x})}{p_2(\boldsymbol{x})}\, \mathrm{d}\boldsymbol{x},$$

was advocated by Jeffreys (1939/1961, p. 179) to derive objective priors, mainly because its invariance properties. Besides being symmetric, it shares all the properties of the KL divergence.

L_m**-norm** This measure,

$$L_m(p_1, p_2) = \int \left| p_1^{\frac{1}{m}}(\boldsymbol{x}) - p_2^{\frac{1}{m}}(\boldsymbol{x}) \right|^m \mathrm{d}\boldsymbol{x},$$

was also considered by Jeffreys to derive objective priors. He also demonstrates (Jeffreys, 1939/1961, p. 180) that when $m = 2$, and under regularity conditions, $\mathscr{J}(\boldsymbol{\theta}, \boldsymbol{\theta} + \Delta\boldsymbol{\theta}) \approx 4\, L_2(\boldsymbol{\theta}, +\Delta\boldsymbol{\theta})$. Unlike KL and $\mathscr{J}$ divergences, $L_2(p_1, p_2)/2$ is a distance, also known as Hellinger distance, and Robert (1996) proposes it as an intrinsic loss.

Chernoff's divergence The Chernoff (1952) measure,

$$\mathscr{C}(p_1, p_2) = \max_{0 \leq t \leq 1} -\log \psi(t)$$

where

$$\psi(t) = \int p_1(\boldsymbol{x}) \left(\frac{p_2(\boldsymbol{x})}{p_1(\boldsymbol{x})}\right)^t \mathrm{d}\boldsymbol{x},$$

and the related measure of Bhattacharyya (1943), that results when holding fixed $t = 1/2$, $\mathscr{B}(p_1, p_2) = -\log \psi(1/2)$, are additive for independent observations but Chernoff's measure requires them to be identically distributed.

In a similar fashion, Rényi (1965, 1976, p. 583) defines the *informational loss of order*

$\alpha > 0$, of substituting p_1 by p_2, when p_1 is correct as

$$\mathcal{R}_\alpha(p_2 \mid p_1) = \begin{cases} (\alpha-1)^{-1} \log \left[\int p_1(\boldsymbol{x}) \left(\frac{p_2(\boldsymbol{x})}{p_1(\boldsymbol{x})} \right)^\alpha \mathrm{d}\boldsymbol{x} \right] & \alpha \neq 1 \\ k(p_2 \mid p_1) & \alpha = 1 \end{cases}.$$

Intrinsic losses

Robert (1996) argues for the need for an objective loss function that depends only on the sampling distribution $p(\boldsymbol{x} \mid \boldsymbol{\theta})$ and coines the concept of *intrinsic loss*, considering two candidates: the KL divergence and the Hellinger distance. Unlike the KL divergence, the Hellinger distance is symmetric and well defined even for no-regular models; however it is not additive for conditionally independent observations. These shortcomings are addressed in the next section.

Intrinsic discrepancy between two distributions

Following Bernardo and Rueda (2002),

Definition 2.1 (Intrinsic discrepancy).
The *intrinsic discrepancy*, $\delta(p_1, p_2)$, between two densities $p_1(\boldsymbol{x}), \boldsymbol{x} \in \mathcal{X}_1$ y $p_2(\boldsymbol{x}), \boldsymbol{x} \in \mathcal{X}_2$ is given by

$$\delta(p_1, p_2) = \min\Big\{ k\big(p_2(\boldsymbol{x}) \mid p_1(\boldsymbol{x})\big), k\big(p_1(\boldsymbol{x}) \mid p_2(\boldsymbol{x})\big) \Big\}.$$

If two families of densities,

$$M_1 \equiv \{p_1(\boldsymbol{x} \mid \boldsymbol{\phi}), \boldsymbol{x} \in \mathcal{X}_1(\boldsymbol{\phi}), \boldsymbol{\phi} \in \Phi\} \text{ and } M_2 \equiv \{p_2(\boldsymbol{x} \mid \boldsymbol{\psi}), \boldsymbol{x} \in \mathcal{X}_2(\boldsymbol{\phi}), \boldsymbol{\psi} \in \Psi\},$$

are considered, the intrinsic discrepancy is

$$\delta^*(M_1, M_2) = \min_{\substack{\boldsymbol{\phi} \in \Phi \\ \boldsymbol{\psi} \in \Psi}} \delta\Big(p_1(\boldsymbol{x} \mid \boldsymbol{\phi}), p_2(\boldsymbol{x} \mid \boldsymbol{\psi})\Big).$$

Proposition 2.1 (Properties of the intrinsic discrepancy).
Let $\delta(p_1, p_2)$ as in Definition 2.1, then

(i) The intrinsic discrepancy $\delta(p_1, p_2) \geq 0$, with equality iff $p_1(\boldsymbol{x}) = p_2(\boldsymbol{x})$ a.e.

(ii) The intrinsic discrepancy is invariant under monotone transformations of the data. Hence, if $\boldsymbol{y} = \boldsymbol{y}(\boldsymbol{x})$ is a one-to-one monotone transformation, then $\delta_{\boldsymbol{y}}(p_1, p_2) = \delta_{\boldsymbol{x}}(p_1, p_2)$.

(iii) *The intrinsic discrepancy is additive for conditionally independent observations. Hence, if* $\boldsymbol{x} = \{x_1,\ldots,x_n\}$ *is a random sample from either* $p_1(\boldsymbol{x})$ *or* $p_2(\boldsymbol{x})$*, then* $\delta_{\boldsymbol{n}}(p_1,p_2) = n\,\delta_x(p_1,p_2)$.

(iv) *If both densities are members of a parametric family,* $p\left(\boldsymbol{x}\mid\varphi\right)$*, such that* $p_1(\boldsymbol{x}) = p\left(\boldsymbol{x}\mid\varphi_1\right)$ *and* $p_2(\boldsymbol{x}) = p\left(\boldsymbol{x}\mid\varphi_2\right)$*; then, the intrinsic discrepancy,* $\delta(p_1,p_2) = \delta(\varphi_1,\varphi_2)$*, is invariant under one-to-one re-parameterizations. Thus, if* $\boldsymbol{\psi} = \boldsymbol{\psi}(\varphi)$ *is one-to-one, then* $\delta\left(p\left(\boldsymbol{x}\mid\boldsymbol{\psi}_1\right), p\left(\boldsymbol{x}\mid\boldsymbol{\psi}_2\right)\right) = \delta\left(\varphi(\boldsymbol{\psi}_1), \varphi(\boldsymbol{\psi}_2)\right)$.

(v) *The intrinsic discrepancy is a measure of the minimum amount of information –in nits– which the observation* $\boldsymbol{x}\in\mathscr{X}$ *is expected to provide in order to discriminate between the models* $p_1(\boldsymbol{x})$ *and* $p_2(\boldsymbol{x})$.

(vi) *The intrinsic discrepancy is symmetric; i.e.* $\delta(p_1,p_2) = \delta(p_2,p_1)$.

(vii) *The intrinsic discrepancy is defined for densities with nested supports. Precisely,* $\delta(p_i,p_j) = k(p_j\mid p_i)$ *if* $\mathscr{X}_i\subset\mathscr{X}_j$.

Intrinsic discrepancy loss

The intrinsic discrepancy is proposed as an appropriate objective loss function for point estimation and hypothesis testing.

Definition 2.2 (Intrinsic discrepancy loss).
Assume that an adequate description of the probabilistic behaviour of the random quantity $\boldsymbol{x}$, is given by the model, $\{p(\boldsymbol{x}\mid\boldsymbol{\theta},\boldsymbol{\lambda}),\ \boldsymbol{x}\in\mathscr{X},\ \boldsymbol{\theta}\in\Theta, \boldsymbol{\lambda}\in\Lambda\}$. The intrinsic discrepancy (loss) of substituting the whole model with the restricted one, obtained when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, is the intrinsic discrepancy between $p(\boldsymbol{x}\mid\boldsymbol{\theta},\boldsymbol{\lambda})$ and the family of densities $\{p(\boldsymbol{x}\mid\boldsymbol{\theta}_0,\boldsymbol{\lambda}),\boldsymbol{\lambda}\in\Lambda\}$, i.e.

$$\delta^*(\boldsymbol{\theta},\boldsymbol{\lambda};\boldsymbol{\theta}_0) = \inf_{\boldsymbol{\lambda}_0\in\Lambda}\delta(\boldsymbol{\theta},\boldsymbol{\lambda};\boldsymbol{\theta}_0,\boldsymbol{\lambda}_0).$$

Where

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

is the minimum of the directed divergences.

Proposition 2.2 (Properties of the intrinsic discrepancy).
Consider the model $\{p(\boldsymbol{x}\mid\boldsymbol{\theta},\boldsymbol{\lambda}),\ \boldsymbol{x}\in\mathscr{X},\ \boldsymbol{\theta}\in\theta,\ \boldsymbol{\lambda}\in\Lambda\}$ *and assume that* $\boldsymbol{\theta}$ *is the parameter of interest. The intrinsic discrepancy,* $\delta^*(\boldsymbol{\theta},\boldsymbol{\lambda};\boldsymbol{\theta}_0)$ *is*

(i) *Invariant under one-to-one transformations of the data; i.e. if* $\boldsymbol{y} = \boldsymbol{y}(\boldsymbol{x})$ *is a monotone transformation, then* $\delta_{\boldsymbol{y}}(\boldsymbol{\theta}, \boldsymbol{\theta}_0; \boldsymbol{\lambda}) = \delta_{\boldsymbol{x}}(\boldsymbol{\theta}, \boldsymbol{\theta}_0; \boldsymbol{\lambda})$.

(ii) *Compatible with sufficient statistics. If* $\boldsymbol{t} = \boldsymbol{t}(\boldsymbol{x})$ *is sufficient for* $p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\lambda})$, *then* $\delta^*_{\boldsymbol{t}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\theta}_0) = \delta^*_{\boldsymbol{x}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\theta}_0)$.

(iii) *Additive in the sense that if data* $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ *are conditionally independent, then* $\delta^*_{\boldsymbol{x}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\theta}_0) = \sum \delta^*_{x_i}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\theta}_0)$. *Moreover, if* $\boldsymbol{x}$ *are iid., then* $\delta^*_{\boldsymbol{x}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\theta}_0) = n\, \delta^*_{x_i}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\theta}_0)$.

(iv) *Invariant under the choice of the nuisance. In fact, if* $\boldsymbol{\omega} = \boldsymbol{\omega}(\boldsymbol{\lambda})$ *is a monotone re-parametrization of* $\boldsymbol{\lambda}$, *then* $\delta^*(\boldsymbol{\theta}, \boldsymbol{\omega}; \boldsymbol{\theta}_0) = \delta^*\big(\boldsymbol{\theta}, \boldsymbol{\lambda}(\boldsymbol{\omega}); \boldsymbol{\theta}_0\big)$.

Some results are derived to aid in the computation of the intrinsic discrepancy

Proposition 2.3 (Intrinsic discrepancy in a regular model).
Let $\{p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\lambda}),\ \boldsymbol{x} \in \mathscr{X},\ \boldsymbol{\theta} \in \Theta,\ \boldsymbol{\lambda} \in \Lambda\}$ *be a probability model that meets the regularity conditions. Then*

$$\begin{aligned}\delta^*(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\theta}_0) &= \inf_{\boldsymbol{\lambda}_0 \in \Lambda} \delta(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\theta}_0, \boldsymbol{\lambda}_0) \\ &= \min\left\{ \inf_{\boldsymbol{\lambda}_0 \in \Lambda} k(\boldsymbol{\theta}, \lambda \mid \boldsymbol{\theta}_0, \boldsymbol{\lambda}_0),\ \inf_{\boldsymbol{\lambda}_0 \in \Lambda} k(\boldsymbol{\theta}_0, \boldsymbol{\lambda}_0 \mid \boldsymbol{\theta}, \boldsymbol{\lambda}) \right\}.\end{aligned}$$

Corolary 2.1.
Let $\{p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\lambda}),\ \boldsymbol{x} \in \mathscr{X},\ \boldsymbol{\theta} \in \Theta,\ \boldsymbol{\lambda} \in \Lambda\}$ *be a probabilistic model from the exponential family. Then,*

$$\begin{aligned}\delta^*(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\theta}_0) &= \inf_{\boldsymbol{\lambda}_0 \in \Lambda} \delta(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\theta}_0, \boldsymbol{\lambda}_0) \\ &= \min\left\{ \inf_{\boldsymbol{\lambda}_0 \in \Lambda} k(\boldsymbol{\theta}, \lambda \mid \boldsymbol{\theta}_0, \boldsymbol{\lambda}_0),\ \inf_{\boldsymbol{\lambda}_0 \in \Lambda} k(\boldsymbol{\theta}_0, \boldsymbol{\lambda}_0 \mid \boldsymbol{\theta}, \boldsymbol{\lambda}) \right\},\end{aligned}$$

with

$$\begin{aligned}k(\boldsymbol{\psi}_j \mid \boldsymbol{\psi}_i) &= \int p\left(\boldsymbol{x} \mid \boldsymbol{\psi}_i\right) \log \frac{p\left(\boldsymbol{x} \mid \boldsymbol{\psi}_i\right)}{p\left(\boldsymbol{x} \mid \boldsymbol{\psi}_j\right)}\, \mathrm{d}\boldsymbol{x} \\ &= M(\boldsymbol{\psi}_i) - M(\boldsymbol{\psi}_j) + \left(\boldsymbol{\psi}_j^t - \boldsymbol{\psi}_i^t\right) \nabla M(\boldsymbol{\psi}_i),\end{aligned}$$

where $\boldsymbol{\psi}_k = \{\boldsymbol{\theta}_k, \boldsymbol{\lambda}_k\}$, $k = 0, 1$, $M(\boldsymbol{\psi}) = \log a(\boldsymbol{\psi})$ *and* $\nabla M(\boldsymbol{\psi}) = \partial M(\boldsymbol{\psi}) / \partial \boldsymbol{\psi}$.

Under some conditions, the intrinsic discrepancy is convex.

Proposition 2.4 (Convexity of the intrinsic discrepancy).
Let $\{p(\boldsymbol{x} \mid \boldsymbol{\theta}), \boldsymbol{x} \in \mathscr{X}, \boldsymbol{\theta} \in \Theta\}$ *be a probability model. The intrinsic discrepancy,* $\delta(\boldsymbol{\theta}; \boldsymbol{\theta}_0)$*, is convex iff the log-likelihood ratio is convex in* $\boldsymbol{\theta}$.

Finally, recalling that in a decision problem the parameter of interest is that which enters the loss function, define (Bernardo and Rueda, 2002)

Definition 2.3 (Intrinsic statistic).
Let $\{p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\lambda}), \boldsymbol{x} \in \mathscr{X}, \boldsymbol{\theta} \in \Theta, \boldsymbol{\lambda} \in \Lambda\}$ be a parametric model that adequately describes the probabilistic behaviour of the random quantity $\boldsymbol{x}$ and let $\delta^*(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\theta}_0)$ as in Definition 2.2. We say that $d(\boldsymbol{\theta}_0 \mid \boldsymbol{x})$ is the *posterior expected intrinsic discrepancy* or *intrinsic statistic* if

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

where $\pi_{\delta^*}(\boldsymbol{\theta}, \boldsymbol{\lambda} \mid \boldsymbol{x})$ is the reference posterior for the parameters of the model $p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\lambda})$ when $\delta^*(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\theta}_0)$ is the parameter of interest.

Chapter 3

INTRINSIC ESTIMATION AND HYPOTHESIS TESTING

In this chapter the intrinsic statistic is applied to the problems of point estimation and hypothesis testing. The entailed definitions of the intrinsic estimator (Bernardo and Juárez, 2003) and the Bayesian reference criterion (Bernardo, 1999) are presented and their properties analysed. Both concepts are implemented in a number of basic statistical models.

Intrinsic estimation

The intrinsic statistic, $d(\boldsymbol{\theta}_0 \mid \boldsymbol{x})$, is a measure (in natural informational units) of the strength of the evidence (conveyed by the data) against using $p(\boldsymbol{x} \mid \boldsymbol{\theta}_0, \boldsymbol{\lambda})$ as a proxy for $p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\lambda})$. Evidently, the best proxy is attained at the value which yields the minimum loss; thus is natural to define (Bernardo and Juárez, 2003).

Definition 3.1 (Intrinsic estimator).
Let $\mathscr{M} \equiv \{p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\lambda}), \boldsymbol{x} \in \mathscr{X}, \boldsymbol{\theta} \in \Theta, \boldsymbol{\lambda} \in \Lambda\}$ be a parametric model which adequately describes the probabilistic behaviour of the random quantity $\boldsymbol{x}$. We call the *intrinsic estimator*, $\boldsymbol{\theta}^* = \boldsymbol{\theta}^*(\boldsymbol{x})$, of the parameter $\boldsymbol{\theta}$ to that value that minimises the intrinsic statistic. Thence,

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

The intrinsic estimator is well defined, regardless of the parameter vector dimension, and, it exists and is unique under mild conditions, pertaining the convexity of the intrinsic discrepancy.

Proposition 3.1 (Uniqueness of the intrinsic estimator).
The intrinsic estimator, $\boldsymbol{\theta}^*(\boldsymbol{x})$*, exists and is unique if the parameter space is strictly convex.*

Furthermore, the intrinsic estimator possess a number of nice properties.

Proposition 3.2 (Properties of the intrinsic estimator).
Derived from the intrinsic statistic, the intrinsic estimator

(i) Is Invariant under monotone transformations

(ii) Is compatible with sufficient statistics

(iii) Is invariant under the choice of the nuisance parameter.

(iv) Is invariant under monotone transformations of the data.

As expected, under regularity conditions, the intrinsic estimator and the MLE are asymptotically equivalent.

Proposition 3.3 (Asymptotic behaviour of the intrinsic estimator).
Consider a random sample, $\boldsymbol{z} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$*, from a parametric regular model,*

$$\left\{p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\lambda}),\, \boldsymbol{x} \in \mathscr{X},\, \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^k, \boldsymbol{\lambda} \in \Lambda\right\}.$$

Then, the MLE, $\hat{\boldsymbol{\theta}}$*, is an asymptotic approximation to the intrinsic estimator; i.e. for sufficiently large* n*,* $\boldsymbol{\theta}^* \approx \hat{\boldsymbol{\theta}}$*.*

Intrinsic testing

The intrinsic statistic (Bernardo and Rueda, 2002) is a measure –in nits– of the expected posterior amount of information required to recover the model, which has been assumed correct, from its closest approximation within the class of models, $M_0 \equiv \{p(\boldsymbol{x} \mid \boldsymbol{\theta}_0, \boldsymbol{\lambda}), \boldsymbol{\lambda} \in \Lambda\}$; it is a measure of the strength of the evidence provided by the data against M_0. It is a test statistic for the (null) hypothesis $H_0 \equiv \{\boldsymbol{\theta} = \boldsymbol{\theta}_0\}$. Thus, H_0 must be rejected iff $d(\boldsymbol{\theta}_0 \mid \boldsymbol{x}) > d^*$, for some threshold value d^*.

Definition 3.2 (Bayesian reference criterion).
Assume that the parametric model, $M \equiv \{p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\lambda}),\, \boldsymbol{x} \in \mathscr{X},\, \boldsymbol{\theta} \in \Theta, \boldsymbol{\lambda} \in \Lambda\}$, is an adequate description of the probabilistic behaviour of the random quantity $\boldsymbol{x}$, and consider the value $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ amongst those which continue being possible after observing $\boldsymbol{x}$. To decide if $p(\boldsymbol{x} \mid \boldsymbol{\theta}_0, \boldsymbol{\lambda})$ can be used as an acceptable proxy for $p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\lambda})$, use the Bayesian reference criterion (BRC)

i) Calculate the intrinsic discrepancy,

$$\delta^*(\boldsymbol{\theta},\boldsymbol{\lambda};\boldsymbol{\theta}_0)=\min_{\boldsymbol{\lambda}_0\in\Lambda}\delta(\boldsymbol{\theta},\boldsymbol{\lambda};\boldsymbol{\theta}_0,\boldsymbol{\lambda}_0).$$

ii) Calculate the intrinsic statistic,

$$d(\boldsymbol{\theta}_0\mid\boldsymbol{x})=\iint\delta^*(\boldsymbol{\theta},\boldsymbol{\lambda};\boldsymbol{\theta}_0)\pi_\delta(\boldsymbol{\theta},\boldsymbol{\lambda}\mid\boldsymbol{x})\,\mathrm{d}\boldsymbol{\lambda}\,\mathrm{d}\boldsymbol{\theta};$$

where $\pi_\delta(\boldsymbol{\theta},\boldsymbol{\lambda}\mid\boldsymbol{x})$, is the reference posterior when $\delta(\boldsymbol{\theta},\boldsymbol{\lambda};\boldsymbol{\theta}_0)$ is the parameter of interest.

iii) If a formal decision is required, reject $H_0\equiv\{\boldsymbol{\theta}=\boldsymbol{\theta}_0\}$ iff $d(\boldsymbol{\theta}_0\mid\boldsymbol{x})\ >\ d^*$, form some threshold value d^*; which, for scientific communication might be reported as: $d^*\approx 1$, no evidence against H_0; $d^*\approx 2.5$, mild evidence against H_0; and $d^*>5$, data provide strong evidence against the null.

Intrinsic hypothesis testing inherits the properties of the intrinsic statistic.

Proposition 3.4 (Properties of the intrinsic statistic).
Inherited from the properties of the intrinsic discrepancy and the reference posterior, the intrinsic statistic, $d(\boldsymbol{\theta}_0\mid\boldsymbol{x})$, is:

(i) Invariant under monotone transformations of the parameter of interest,

(ii) Compatible with sufficient statistics.

(iii) Invariant under the choice of the nuisance parameter.

(iv) Invariant under monotone transformations of the data.

A simple, powerful approximation to the intrinsic statistic is derived, for regular models.

Proposition 3.5 (Asymptotical approximation under regularity conditions).
Let $\{p(\boldsymbol{x}\mid\theta,\boldsymbol{\lambda}),\ \boldsymbol{x}\in\mathscr{X},\ \theta\in\Theta\subset\Re,\boldsymbol{\lambda}\in\Lambda\}$ be a parametric model such that the (marginal) posterior distribution of $\boldsymbol{\theta}$ is regular. If the intrinsic discrepancy, $\delta^*(\theta,\boldsymbol{\lambda};\theta_0)=\delta^*(\theta;\theta_0)$, then the intrinsic statistic can be well approximated by

$$d(\theta_0\mid\boldsymbol{x})\approx\delta^*(\tilde{\theta};\theta_0)+\frac{1}{2},$$

where $\tilde{\theta}$ is the mode of the asymptotic posterior of $\boldsymbol{\theta}$.

The second part of the chapter is devoted to the implementation of these concepts and results to a list of models of extensive use in the literature.

Chapter 4

EVALUATION AND COMPARISONS

From a subjectivistic standpoint, a Bayes rule is derived for the specific problem at hand and is optimal for it; thus a comparison among different Bayes rules, unless directed towards a sensibility analysis, is senseless. From an objective viewpoint, however, as the Bayes rule is derived for a generic use, with no specific aim in mind, comparison between alternatives is sensible. In this chapter comparisons with frequentist and Bayesian alternatives and evaluation under homogenous conditions are performed.

Point estimation

Under mild conditions, the intrinsic estimator is admissible (under intrinsic discrepancy loss)

Corolary 4.1.
Consider the continuous parametric model, $\{p(\boldsymbol{x} \mid \boldsymbol{\theta}), \boldsymbol{x} \in \mathcal{X}, \boldsymbol{\theta} \in \Theta\}$ and assume that the parameter space, $\Theta \subset \mathbb{R}^p$, is convex. Then, the intrinsic estimator is admissible.

In particular, if $p(\boldsymbol{x} \mid \boldsymbol{\theta})$ belongs to the exponential family, the intrinsic estimator, $\boldsymbol{\theta}^*$, is admissible.

In the other hand, under regularity conditions the risk of the intrinsic estimator is asymptotically constant.

Proposition 4.1 (Asymptotic risk under regularity conditions).
Suppose that $\boldsymbol{\theta}^*(\boldsymbol{x})$ is the intrinsic estimator of the parameter $\boldsymbol{\theta}$, which indexes the regular model $\{p(\boldsymbol{x} \mid \boldsymbol{\theta}), \boldsymbol{x} \in \mathcal{X}, \boldsymbol{\theta} \in \Theta\}$. Under these conditions, the risk of the intrinsic estimator, under intrinsic discrepancy loss, $R_{\boldsymbol{\theta}^*}(\boldsymbol{\theta})$, is asymptotically constant and equal to $1/2$.

Hypothesis testing

Unlike point estimation, frequentist and Bayesian solutions to precise hypothesis testing usually differ. Indeed, p-values can be criticised for *i*) an arbitrary selection of the text statistic; *ii*) not being a measure of the evidence against the null; *iii*) having arbitrary threshold values; *iv*) overestimating the significance; *v*) potentially lead to controversial answers and *vi*) not being a general procedure.

Nevertheless, it is possible to reach some agreement between the BRC and the test derived from the generalised likelihood ratio.

Proposition 4.2 (Asymptotic equivalence of the BRC).
Consider the regular model $\{p(\boldsymbol{x} \mid \theta), \boldsymbol{x} \in \mathcal{X}, \theta \in \Theta \subset \mathbb{R}\}$*, and assume the programme of Definition 3.2 is used to test the (null) hypothesis* $H_0 \equiv \{\theta = \theta_0\}$*. Then, asymptotically* H_0 *will be rejected with a* p*-value of* α *iff* $\alpha = 2\,\Phi(\sqrt{2\,d^* - 1})$*, where* $\Phi(z)$ *is the area to the right of* z *under a standard normal curve.*

BIBLIOGRAPHY

Ali, S.M. and Silvey, S.D. (1966). A general class of coefficients of divergence of one distribution to another. *J. Roy. Statist. Soc. B*, **28**, 131–142.

Alonso, M. (1958). *Enciclopedia del idioma: diccionario histórico y moderno de la lengua española*, vol. 2. Madrid: Aguilar.

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

Bayarri, M.J. (1981). Inferencia bayesiana sobre el coeficiente de correlación en una población normal bivariante. *Trabajos de Estadística e Investigación Operativa*, **32**, 18–31.

Bayarri, M.J. and DeGroot, M.H. (1992). Difficulties and ambiguities in the definition of a likelihood function. *J. It. Statist. Soc.*, **1**, 1–15.

Bayarri, M.J., DeGroot, M.H. and Kadane, J.B. (1988). What is the likelihood function? S.S. Gupta and J.O. Berger, eds., *Statistical Decision Theory and Related Topics IV*. Berlin: Springer, pp. 3–27.

Beran, R. (1977). Minimum Hellinger distance estimators for parametric models. *The Annals of Statistics*, **5**, 445–463.

Berger, J.O. (1985). *Statistical decision theory and Bayesian analysis*. New York: Springer-Verlag, 2nd ed.

Berger, J.O. and Bernardo, J.M. (1989). Estimating a product of means: Bayesian analysis with reference priors. *J. Amer. Statist. Assoc.*, **84**, 200–207.

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

Berger, J.O. and Bernardo, J.M. (1992*b*). Ordered group reference priors with applications to a multinomial problem. *Biometrika*, **79**, 25–37.

Berger, J.O. and Bernardo, J.M. (1992*c*). Reference priors in a variance components problem. P. Goel and N. Iyengar, eds., *Bayesian Analysis in Statistics and Econometrics IV*. Berlin:Springer, pp. 323–340.

Berger, J.O., Bernardo, J.M. and Mendoza, M. (1989). On priors that maximize expected information. J.P. Klein and J.C. Lee, eds., *Recent developments in statistics and their applications*. Seoul: Freedom Academy, pp. 1–20.

Berger, J.O., Boukai, B. and Wang, Y. (1997). Unified frequentist and bayesian testing of a precise hypothesis. *Statistical Science*, **3**, 133–160.

Berger, J.O. and Delampady, M. (1987). Testing precise hypothesis. *Statistical Science*, **2**, 317–352.

Berger, J.O. and Pericchi, L. (1996*a*). The intrinsic Bayes Factor for linear models. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, eds., *Bayesian Statistics 5*. Oxford: University Press, pp. 24–44.

Berger, J.O. and Pericchi, L. (1996*b*). The intrinsic Bayes Factor for model selection and prediction. *J. Amer. Statist. Assoc.*, **91**, 109–122.

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

Berger, J.O., Pericchi, L. and Varshavsky, J. (1998). Bayes Factors and marginal distributions in invariant situations. *Sankya*, **60**, 307–321.

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. (includes discussion).

Berger, J.O. and Strawderman, W.E. (1993). Choice of hierarchical priors: admissibility in estimation of Normal means. *Techinacl report 93-34C*, Purdue University.

Bernardo, J.M. (1979*a*). Expected loss as expected utility. *The Annals of Statistics*, **7**, 686–690.

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

Bernardo, J.M. (1981). Reference decisions. *Symposia Mathematica*, **25**, 85–94.

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. (1997). Non-informative priors do not exist. *Journal of statistical planning and inference*, **65**, 159–189. (includes discussion).

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

Bernardo, J.M. (2003). Bayesian statistics. R. Viertl, ed., *Encyclopedia of Life Support Systems (EOLSS). Probability and Statistics*. Oxford, UK: UNESCO, p. (www.eolss.net).

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

Bernardo, J.M. and Juárez, M.A. (2003). Intrinsic estimation. J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith and M. West, eds., *Bayesian Statistics 7*. Oxford: University Press, pp. 465–476.

Bernardo, J.M. and Ramón, J.M. (1998). An introduction to Bayesian reference analysis: inference on the ratio of multinomial parameters. *The Statistician*, **47**, 101–135.

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

Bernardo, J.M. and Rueda, R. (2003). A bayesian reference approach to variable selection. Invited paper to the 4th International Workshop on Objective Bayesian Methodology.

Bernardo, J.M. and Smith, A.F. (1994). *Bayesian Theory*. Chichester: John Wiley.

Bernoulli, D. (1769). Dijudicatio maxime probabilis plurium observationum discrepantium atque verisimillima inductio inde formada. E.S. Pearson and M. Kendall, eds., *Studies in the History of Statistics and Probability*, vol. 1. UK: Griffin, vol. 1, pp. 157–167. Reprinted in 1970.

Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math Soc.*, **35**, 99–109.

Birnbaum, A. (1962). On the foundations of statistical inference. *J. Amer. Statist. Assoc.*, **57**, 269–326. (incluye discusión).

Boos, D.D. (1981). Minimum distance estimation for location and goodness of fit. *J. Amer. Statist. Assoc.*, **76**, 663–670.

Borel, E. (1924/1964). A propos of a treatise on probability. H.E. Kyburg and H.E. Smokler, eds., *Studies in Subjective Probability*. New York: Dover, pp. 45–60.

Box, G.E.P. and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*. Reading: Addison-Wesley.

Brewster, J.F. and Zidek, J.V. (1974). Improving on equivariant estimators. *The Annals of Statistics*, **2**, 21–38.

Brown, L.D. (1968). Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters. *Ann. Math. Stat.*, **39**, 29–48.

Brown, L.D. (1990). Comment on: Developments in decision-theoretic variance estimation. *Statistical Science*, **5**, 103–106.

Chernoff, H. (1952). Measure of asymptotic efficiency for test a hypothesis based on the sum of observations. *Ann. Math. Stat.*, **23**, 493–507.

Clarke, B. and Wasserman, L. (1993). Non-informative priors and nuisance parameters. *J. Amer. Statist. Assoc.*, **88**, 1427–1432.

Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*. London: Chapman and Hall.

Cox, D.R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *J. Roy. Statist. Soc. B*, **49**, 1–39. (includes discussion).

Csiszár, I. (1975). I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, **3**, 146–158.

Csörgő, S. and Faraway, J. (1996). On the estimation of a normal variance. *Statistics and Decisions*, **14**, 23–34.

Dalal, S.R. and Hall, W.J. (1983). Approximating priors by mixtures of natural conjugate priors. *J. Roy. Statist. Soc. B*, **45**, 278–286.

Dale, A.I. (1991). *A History of Inverse Probability*. New York: Springer, segunda ed.

Datta, G.S. and Ghosh, M. (1995). Some remarks on noninformative priors. *J. Amer. Statist. Assoc.*, **90**, 1357–1363.

Datta, G.S. and Ghosh, M. (1996). On the invariance of noninformative priors. *The Annals of Statistics*, **24**, 141–159.

Dawid, A.P. (1983). Invariant prior distributions. S. Kotz, N.L. Johnson and C.B. Read, eds., *Encyclopedia of Statistical Sciences*, vol. 4. New York: Wiley, vol. 4, pp. 228–236.

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. (includes discussion).

DeGroot, M.H. (1970). *Optimal Statistical Decisions.* New York: McGraw-Hill.

DeGroot, M.H. (1973). Doing what comes naturally: Interpreting a tail area as a posterior probability or as a likelihood ratio. *J. Amer. Statist. Assoc.*, **68**, 966–969.

DeGroot, M.H. (1982). Comment on "Lindley's paradox". *J. Amer. Statist. Assoc.*, **77**, 336–339.

DeGroot, M.H. (1987). *Probability and Statistics.* Reading, MA: Addison-Wesley.

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

Edwards, A.W.F. (1972). *Likelihood.* UK: Cambridge University Press.

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

Efron, B. and Morris, C. (1973). Combining possibly related estimation problems. *J. Roy. Statist. Soc. B*, **35**, 379–421. (includes discussion).

Ferguson, T.S. (1967). *Mathematical Statistics: a Decision Theoretic Approach.* New York: Academic Press.

Ferguson, T.S. (1976). Development of the decision model. D.B. Owen, ed., *On the History of Probability and Statistics.* New York: Marcel Dekker, pp. 335–346.

Fernández, C. and Steel, M. (1999). Reference priors for the general location-scale model. *Statistics & Probability Letters*, **43**, 377–384.

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

de Finetti, B. (1937/1980). Foresight: Its logical laws, its subjective sources. H.E. Kyburg and H.E. Smokler, eds., *Studies in Subjective Probability*. New York: Krieger, pp. 53–118.

de Finetti, B. (1974). *Theory of Probability*, vol. 1. Chichester: Wiley.

de Finetti, B. (1975). *Theory of Probability*, vol. 2. Chichester: Wiley.

Fisher, R.A. (1915). Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. *Biometrika*, **10**, 507–521.

Fisher, R.A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron*, **1**, 3–32.

Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics. *Phil. Tran. Roy. Soc. London A*, **222**, 309–368. Reprinted in *Breakthroughs in Statistics* **1** (S. Kotz y N. L. Johnson, eds.). Berlin: Springer, 1991, 11–44.

Fisher, R.A. (1925). Theory of statistical estimation. *Proc. Camb. Phil. Soc.*, **22**, 700–725. Reprinted in *Contributions to mathematical statistics*, New York: Wiley, 1950, paper 11.

Fisher, R.A. (1956). *Statistical Methods and Scientific Inference*. Edinburgh: Oliver and Boyd.

French, S. and Ríos-Insua, D. (2000). *Statistical Decision Theory, Kendall's Advanced Theory of Statistics*, vol. 9. London: Edward Arnold.

García-Donato, G. (2003). *Factores Bayes y Factores Bayes convencionales: algunos aspectos relevantes*. Phd thesis, Universidad de Valencia.

Gauss, C.F. (1887). *Abhandlungen zur Methode der kleinsten Quadrate*. Berlin: Stnakiewicz.

Geisser, S. (2003). Models of parametric statistical inference. Unpublished book.

George, E. and McCulloch, R. (1993). Variable selection via Gibbs sampling. *J. Amer. Statist. Assoc.*, **88**, 881–889.

Ghosh, J.K. and Mukerjee, R. (1992). Non-informative priors. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, eds., *Bayesian Statistics 4*. Oxford: University Press, pp. 195–210. (includes discussion).

Gibbons, J.D. and Pratt, J.W. (1975). P-values: Interpretation and methodology. *The american statistician*, **29**, 20–25.

Gradshteyn, I.S. and Ryzhik, J.M. (1994). *Table of Integrals, Sums and Products*. London: Academic Press.

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

Halmos, P.R. and Savage, L.J. (1949). Applications of the Radom-Nikodym theorem to the theory of sufficient statistics. *The Annals of Statistics*, **20**, 225–241.

Healy, J.R. (1969). Rao's paradox concerning multivariate tests of significance. *Biometrics*, **25**, 411–413.

Hill, B.M. (1974). On coherence, inadmissibility and inference about many parameters in the theory of least squares. S.E. Fienberg and A. Zellner, eds., *Studies in Bayesian Econometrics and Statistics*. Netherlands: North-Holland, pp. 245–274.

Hill, B.M. (1994). On stenian shrinkage estimators: the finite/infinte problem and formalism in Probability and Statistics. P.R. Freeman and A.F.M. Smith, eds., *Aspects of uncertainty*. Chichester: Wiley, pp. 223–266.

Hwang, J.T., Casella, G., Robert, C., Wells, M. and Farrell, R. (1992). Estimation of accuracy in testing. *The Annals of Statistics*, **20**, 490–509.

James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Stat. Prob.*, vol. 1. Berkeley: U. California Press, vol. 1, pp. 361–380.

Jaynes, E.T. (1980). Marginalization and prior probabilities. A. Zellner, ed., *Bayesian analysis in econometrics and statistics*. Florida: Krieger, pp. 43–88. (includes discussion).

Jaynes, E.T. (1996). *Probability theory: the logic of science*. E-book. Available at `http://bayes.wustl.edu/etj/pub/book.pdf.tar.gz`.

Jeffreys, H. (1939/1961). *Theory of Probability*. Oxford: University Press, 3rd ed.

Jeffreys, H. (1980). Some general points in probability theory. A. Zellner, ed., *Bayesian analysis in econometrics and statistics*. Florida: Kieger, pp. 451–453.

Johnson, D.H. and Sinanović, S. (2001). Symmetrizing the Kullback-Leibler distance. Available at `http://www-ece.rice.edu/~dhj/cv.html#publications`.

Kalbfleisch, J.D. and Sprot, D.A. (1970). Applications of likelihood methods to models involving a large number of parameters. *J. Roy. Statist. Soc. B*, **32**, 175–208. (includes discussion).

Kass, R.E. and Raftery, A.E. (1995). Bayes Factors. *J. Amer. Statist. Assoc.*, **90**, 773–795.

Kass, R.E. and Wasserman, L. (1986). The selection of prior distributions by formal rules. *J. Amer. Statist. Assoc.*, **91**, 1343–1370. Contiene una extensa bibliografía comentada.

Kolmogorov, A.N. (1693). On the approximation of distributions of sums of independent summands by infinitely divisible distributions. *Sankhyā*, **23**, 159–174.

Koopman, B. (1936). On distributions admitting sufficient statistics. *Trans. Amer. Math. Soc.*, **39**, 399–409.

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

Kullback, S. (1983*a*). Fisher information. S. Kotz and N.L. Johnson, eds., *Encyclopedia od statistical sciences Vol. 3.* New York: Wiley, pp. 115–118.

Kullback, S. (1983*b*). Kullback information. S. Kotz and N.L. Johnson, eds., *Encyclopedia od statistical sciences Vol. 4.* New York: Wiley, pp. 421–425.

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

Laplace, P.S. (1812). *Théorie analytique des probabilitiés.* Paris: Académie Française.

Lavine, M. and Schervish, M.J. (1999). Bayes factors: What they are and what they are not. *The American Statistician*, **53**, 119–122.

LeCam, L. (1955). An extension of Wald's theory of statistical decision functions. *Ann. Math. Stat.*, **26**, 69–81.

LeCam, L. (1986). *Asymptotic methods in statistical decision theory.* New York: Springer-Verlag.

Lehmann, E.L. (1947). On families of admissible tests. *Ann. Math. Stats.*, **18**, 97–104.

Lehmann, E.L. (1950). Some principles of the theory of hypothesis testing. *Ann. Math. Stats.*, **21**, 1–26.

Lehmann, E.L. (1986). *Testing Statistical Hypotheses.* New York: John Wiley, 2nd ed.

Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation.* New York: John Wiley, 2nd ed.

Lindley, D.V. (1956). On a measure of the information provided by an experiment. *Ann. Math. Stat.*, **27**, 986–1005.

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

Lindley, D.V. (1958). Fiducial distributions and Bayes' theorem. *J. Roy. Statist. Soc. B*, **20**, 102–07.

Lindley, D.V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint, Vol. 2.* Cambridge: University Press.

Lindley, D.V. (1982). Bayesian inference. S. Kotz, N.L. Johnson and C.B. Read, eds., *Encyclopedia of statistical Sciences*, vol. 1. New York: John Wiley, vol. 1, pp. 197–204.

Lindley, D.V. (1985). *Making Decisions.* London: John Wiley, 2nd ed.

Lindley, D.V. (2000). The philosophy of statistics. *The Statistician*, **40**, 293–337.

Lindley, D.V. and Novick, M.R. (1981). The role of exchangeability in inference. *The Annals of Statistics*, **9**, 45–58.

Lindley, D.V. and Phillips, L.D. (1976). Inference for a Bernoulli process (a Bayesian view). *Americian Statistician*, **30**, 112–119.

Liseo, B. (1993). Elimination of nuisance parameters with reference priors. *Boimetrika*, **80**, 295–304.

Luce, R.D. and Raiffa, H. (1957). *Games and Decisions.* New York: Wiley.

Maatta, J.M. and Casella, G. (1990). Developments in decision-theoretic variance estimation. *Statistical Science*, **5**, 90–120. (includes discussion).

McCullagh, P. and Tibshirani, R. (1990). A simple method for the adjustment of profile likelihood. *J. Roy. Statist. Soc. B*, **52**, 325–344.

Mendoza, M. (1994). Asymptotic Normality under transformations. a result with Bayesian applications. *Test*, **3**, 173–180.

von Neumann, J. and Morgenstern, O. (1944/1953). *Theory of games and economic behaviour.* Princeton: University Press, 3rd ed.

Neyman, J. (1935). Sur un teorema concerte le cosidette statistiche sufficienti. *Giorn. Ist. Ital.*, **6**, 320–334.

Neyman, J. (1976). The emergence of mathematical statistics. D.B. Owen, ed., *On the History of Probability and Statistics.* New York: Marcel Dekker, pp. 147–194.

Neyman, J. and Pearson, E.S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philos. Trans. Roy. Soc. A*, **231**, 289–337.

O'Hagan, A. (1994). *Bayesian Inference, Kendall's Advanced Theory of Statistics*, vol. 2B. London: Edward Arnold.

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.

Parsian, A. and Nematollahi, N. (1996). Estimation of scale parameters under entropy loss function. *Journal of Statistical Planning and Inference*, **52**, 77–91.

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Phil. Trans. Roy. Soc. London A*, **185**, 1–40.

Pérez, S. (2005). Métodos bayesianos objetivos de comparación de medias. PhD Thesis, Universidad de Valencia (to be submitted).

Pitman, E.J.G. (1936). Sufficient statistics and intrinsic accuracy. *Proc. Cambridge Philos. Soc.*, **32**, 567–579.

Pitman, E.J.G. (1979). *Some Basic Theory for Statistical Inference.* London: Chapman and HAll.

Pratt, J.W., Raiffa, H. and Schlaifer, R. (1996). *Introduction to Statistical Decision Theory.* Cambridge, Mass.: MIT Press.

Ramsey, F.P. (1926/1980). Truth and probability. H.E. Kyburg and H.E. Smokler, eds., *Studies in Subjective Probability.* New York: Krieger, pp. 23–52.

Rao, C.R. (1952). *Advanced statistical methods in Biometric research.* New York: Wiley.

Rao, C.R. (1966). Covariance adjustment and related problems in multivariate analysis. P.E. Krishnaiah, ed., *Multivariate analysis.* New York: Academic Press, pp. 87–103.

Rényi, A. (1965). On the amount of information concerning an unknown parameter in a sequence of observations. *Pub. Math. Inst. Hung. Acad. Sci.*, **9**, 617–624.

Rényi, A. (1976). *Cálculo de probabilidades.* Barcelona: Reverté.

Robert, C.P. (1993). A note on Jeffreys-Lindley paradox. *Statistica Sinica*, **3**, 601–608.

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

Robert, C.P. (2001). *The Bayesian Choice.* New York: Springer-Verlag, 2nd ed.

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.

Rukhin, A.L. (1987). How much better are better estimators of a Normal variance. *J. Amer. Statist. Assoc.*, **82**, 925–928.

Savage, L.J. (1954/1972). *The Foundations of Statistics.* New York: Dover, 2nd ed.

Savage, L.J. (1970). *The Foundations of Statistical Inference.* London: Methuen.

Savage, L.J. (1976). On rereading R. A. Fisher. *The Annals of Statistics*, **6**, 441–500.

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.

Shafer, G. (1982). Lindley's paradox. *J. Amer. Statist. Assoc.*, **77**, 325–351.

Shafer, G. and Olkin, O. (1983). Adjusting p-values to account for selection over dichotomies. *J. Amer. Statist. Assoc.*, **78**, 674–678.

Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Tech. J.*, **27**, 379–423 y 623–656. Reprinted in *The Mathematical Theory of Communication* (Shannon, C. E. and Weaver, W., 1949). Urbana Ill.: Univ. Illinois Press.

Sloan, J. and Sinha, S. (1996). Bayesian estimation for the bivariate normal distribution. D.A. Berry, ed., *Bayesian Analysis in Statistics and Econometrics.* New York: John Wiley, pp. 115–127.

Stein, C. (1955). A necessary and sufficient condition for admissibility. *Ann. Math. Stat.*, **26**, 518–522.

Stein, C.M. (1959). An example of wide discrepancy between fiducial and confidence intervals. *Ann. Math. Stat.*, **30**, 877–880.

Stein, C.M. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Inst. Stat. Math*, **16**, 155–160.

Stein, C.M. (1981). Estimation of the mean of a Multivariate Normal distribution. *The Annals of Statistics*, **6**, 1135–1151.

Stigler, S.M. (1999). *Statistics on the Table.* Cambridge, Mass.: Harvard University Press.

Stone, M. and Dawid, A.P. (1972). Un-Bayesian implications of improper Bayesian inference in routine statistical problems. *Biometrika*, **59**, 369–373.

Strawderman, W.E. (1974). Minimax estimation of powers of the variance of a normal population under squared error loss. *The Annals of Statistics*, **2**, 190–198.

Tamura, N. and Boos, D.D. (1986). Minimum Hellinger distance estimation for multivariate location and covariance. *J. Amer. Statist. Assoc.*, **81**, 223–229.

Tibshiriani, R. (1989). Non-informative priors for one parameter of many. *Biometrika*, **76**, 604–608.

Vidakovic, B. and DasGupta, A. (1992). Lower bounds on Bayes risks for estimating a Normal variance: with applications. *Discussion paper 92-22*, ISDS: Duke University.

Villegas, C. (1977). On the representation of ignorance. *J. Amer. Statist. Assoc.*, **72**, 651–654.

Wald, A. (1939). Contributions to the theory of statistical estimation and testing hypotheses. *Ann. Math. Stat.*, **10**, 299–326.

Wald, A. (1947). An essentially complete class of admissible decision functions. *Ann. Math. Stat.*, **18**, 549–555.

Wald, A. (1950). *Statistical Decision Functions.* New York: Wiley.

Yang, R. and Berger, J.O. (1994). Estimation of a covariance matrix using the reference prior. *The Annals of Statistics*, **22**, 1195–1211.

Yang, R. and Berger, J.O. (1997). A catalog of non-informative priors. *Technical Report 97-42*, ISDS.

Zellner, A. (1971). *An introduction to Bayesian Inference in Econometrics.* New York: Wiley.

Zellner, A. (1984). Posterior odds ratios for regression hypotheses: general considerations and some specific results. *Basic Issues in Econometrics.* Chicago: University Press, pp. 275–305.

Zellner, A. and Siow, A. (1980). Posterior odds ratios for selected regression hypotheses. J.M. Bernardo, M.H.D. Groot, D.V. Lindley and A.F.M. .Smith, eds., *Bayesian Statistics.* Valencia: University Press, pp. 585–604.