

José M. Bernardo

***Comparing proportions: A
modern solution to a
classical problem***



1

Comparing Proportions: A Modern Solution to a Classical Problem

CONTENTS

1.1	Summary	7
1.2	Introduction	8
1.3	Integrated Bayesian analysis	9
1.3.1	Bayesian inference summaries	9
1.3.1.1	Point estimation	10
1.3.1.2	Region estimation	10
1.3.1.3	Precise hypothesis testing	11
1.3.2	Continuous invariant loss functions	12
1.3.3	The intrinsic logarithmic loss	13
1.4	Intrinsic reference analysis	15
1.4.1	Intrinsic reference estimation and testing	15
1.4.2	Example: the normal variance	16
1.5	Comparing Binomial proportions	19
1.5.1	Intrinsic logarithmic loss to test equality	19
1.5.2	Reference posterior distributions	20
1.5.3	The RV144 HIV vaccine efficacy trial in Thailand	21
1.6	Discussion	25

1.1 Summary

All final conclusions of a Bayesian statistical analysis are contained in the joint posterior distribution of the set of parameters included in the assumed model, but it is often necessary to *summarize* its contents for a correct assimilation of its inferential implications. In particular, the comparative analysis of the results obtained with two different strategies applied to the same problem typically focuses on the study of the difference or on the rate of the more relevant parametric functions of the problem, with particular attention to the compatibility of the data with the possibility that the difference might be zero, or that the ratio might be one, a conventional example of *precise hypothesis testing*.

The use of Bayesian decision theory with *continuous* loss functions allows an *integrated* solution to the problems of estimation and hypothesis testing, where the *same* prior distribution may be used in both cases, and where that common prior may ever be improper [6]. A particular continuous loss function, the intrinsic logarithmic loss, is recommended for general use. In this Chapter, that methodology is shortly summarized and then applied to an important problem, the comparison of the parameters associated to two independent binomial populations. This is a relatively elementary problem over which there is however no consensus as shown—for example—by the noticeable polemic in the media on the scientific consequences of the 2009 Thailand trial to assess the possible efficacy of the RV144 vaccine against human immunodeficiency.

1.2 Introduction

From a Bayesian viewpoint, the final outcome of any problem of inference is the posterior distribution of the vector of interest. Thus, given a probability model $\mathcal{M}_z = \{p(z|\omega), z \in \mathcal{Z}, \omega \in \Omega\}$ which is assumed to describe the mechanism which has generated the available data \mathbf{z} , *all* that can be said about any function $\theta(\omega) \in \Theta$ of the parameter vector ω is contained in its posterior distribution $p(\theta|\mathbf{z})$. This is computed using standard probability theory techniques from the posterior distribution $p(\omega|\mathbf{z}) \propto p(\mathbf{z}|\omega)p(\omega)$ obtained by Bayes theorem from the assumed prior $p(\omega)$. To facilitate the assimilation of the inferential contents of $p(\theta|\mathbf{z})$, one often tries to *summarize* the information contained in this posterior by

1. providing θ values which, in the light of the data, are likely to be close to its true value (*estimation*), and
2. measuring the compatibility of the data with one or more possible values $\theta_0 \in \Theta$ of the vector of interest which might have been suggested by the research context (*hypothesis testing*).

One would expect that the *same* prior $p(\omega)$, whatever its basis, could be used to derive both types of summaries. However, since the pioneering work by Jeffreys [13], Bayesian methods have often made use of two *radically different* types of prior, some for estimation and some for hypothesis testing. It is argued that this is certainly *not necessary*, and probably not convenient, and that a coherent solution to both problems using the same prior is possible within the standard framework of Bayesian decision theory.

Section 1.3 specifies a decision theoretic formulation for point estimation, region estimation and precise hypothesis testing, emphasizes that the results are highly dependent on the choices of both the loss function and the prior

distribution, and reviews a set of desiderata for loss functions to be used in stylized non context-specific problems of inference.

Section 1.4 proposes the use of the *average log-likelihood ratio against the null*, abbreviated to *intrinsic logarithmic loss*, as a self-calibrated information-based continuous loss function, which is suggested for general use in precise hypothesis testing.

Section 1.5 applies that methodology to the problem of comparing the parameters of two independent binomial populations, providing a *coherent* set of solutions—using the same prior—to *both* the problem of estimating their ratio, and the problem of testing whether or not data are compatible with the hypothesis that both parameters are equal. This is illustrated with the analysis of results published in the literature on the 2009 RV144 HIV vaccine efficacy trial held in Thailand.

1.3 Integrated Bayesian analysis

1.3.1 Bayesian inference summaries

Let \mathbf{z} be the available data which are assumed to have been generated as one random observation from model $\mathcal{M}_{\mathbf{z}} = \{p(\mathbf{z} | \boldsymbol{\omega}), \mathbf{z} \in \mathcal{Z}, \boldsymbol{\omega} \in \boldsymbol{\Omega}\}$. Often, but not always, data will consist of a random sample $\mathbf{z} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ from some distribution $q(\mathbf{x} | \boldsymbol{\omega})$, $\mathbf{x} \in \mathcal{X}$; then, $p(\mathbf{z} | \boldsymbol{\omega}) = \prod_{i=1}^n q(\mathbf{x}_i | \boldsymbol{\omega})$, and $\mathcal{Z} = \mathcal{X}^n$. Let $\boldsymbol{\theta}(\boldsymbol{\omega})$ be the vector of interest. Without loss of generality, the model may explicitly be expressed in terms of the quantity of interest $\boldsymbol{\theta}$, so that $\mathcal{M}_{\mathbf{z}} = \{p(\mathbf{z} | \boldsymbol{\theta}, \boldsymbol{\lambda}), \mathbf{z} \in \mathcal{Z}, \boldsymbol{\theta} \in \boldsymbol{\Theta}, \boldsymbol{\lambda} \in \boldsymbol{\Lambda}\}$, where $\boldsymbol{\lambda}$ is some appropriately chosen nuisance parameter vector. Let $p(\boldsymbol{\theta}, \boldsymbol{\lambda}) = p(\boldsymbol{\lambda} | \boldsymbol{\theta}) p(\boldsymbol{\theta})$ be the assumed prior, and let $p(\boldsymbol{\theta} | \mathbf{z})$ be the corresponding marginal posterior distribution of $\boldsymbol{\theta}$. Appreciation of the inferential contents of $p(\boldsymbol{\theta} | \mathbf{z})$ may be enhanced by providing both point and region estimates of the vector of interest $\boldsymbol{\theta}$, and by declaring whether or not some context suggested specific value $\boldsymbol{\theta}_0$ (or maybe a set of values $\boldsymbol{\Theta}_0$), is (are) compatible with the observed data \mathbf{z} . A large number of Bayesian estimation and hypothesis testing procedures have been proposed in the literature. It is argued that their construction is better made within a *coherent* decision theoretical framework, making use of the *same* prior distribution in all cases.

Let $\ell\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda})\}$ describe, as a function of the (unknown) parameter values $(\boldsymbol{\theta}, \boldsymbol{\lambda})$ which have generated the available data, *the loss* to be suffered if, working with model $\mathcal{M}_{\mathbf{z}}$, the value $\boldsymbol{\theta}_0$ were used as a proxy for the unknown value of $\boldsymbol{\theta}$. As summarized below, point estimation, region estimation and hypothesis are all appropriately described as specific decision problems using a *common* prior distribution and a *common* loss structure. The results may dramatically depend on the particular choices made for both the prior and

the loss function but, given the available data \mathbf{z} , they all only depend on \mathbf{z} through the corresponding posterior expected loss,

$$\bar{\ell}(\boldsymbol{\theta}_0 | \mathbf{z}) = \int_{\Theta} \int_{\Lambda} \ell\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda})\} p(\boldsymbol{\theta}, \boldsymbol{\lambda} | \mathbf{z}) d\boldsymbol{\theta} d\boldsymbol{\lambda}.$$

As a function of $\boldsymbol{\theta}_0 \in \Theta$, the expected loss $\bar{\ell}(\boldsymbol{\theta}_0 | \mathbf{z})$ provides a *direct* measure of the relative unacceptability of all possible values of the quantity of interest in the light of the information provided by the data. Together with the marginal posterior distribution $p(\boldsymbol{\theta} | \mathbf{z})$, this provides the basis for an *integrated* coherent Bayesian analysis of the inferential content of the data \mathbf{z} with respect to the quantity of interest $\boldsymbol{\theta}$.

1.3.1.1 Point estimation

To choose a point estimate for $\boldsymbol{\theta}$ may be seen as a decision problem where the action space is the class Θ of all possible $\boldsymbol{\theta}$ values. Foundations of decision theory dictate that the best estimator is that which minimizes the expected loss; this is called the *Bayes estimator* which corresponds to this particular loss:

$$\boldsymbol{\theta}^*(\mathbf{z}) = \arg \inf_{\boldsymbol{\theta}_0 \in \Theta} \bar{\ell}(\boldsymbol{\theta}_0 | \mathbf{z}).$$

Conventional examples of loss functions include the ubiquitous quadratic loss $\ell\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda})\} = (\boldsymbol{\theta}_0 - \boldsymbol{\theta})^t(\boldsymbol{\theta}_0 - \boldsymbol{\theta})$, which yields the posterior expectation as the Bayes estimator, and the zero-one loss on a neighborhood of the true value, which yields the posterior mode as a limiting result.

1.3.1.2 Region estimation

Bayesian region estimation is easily achieved by quoting posterior credible regions. To choose a q -credible region for $\boldsymbol{\theta}$ may be seen as a decision problem where the action space is the class of subsets of Θ with posterior probability q . Foundations dictate that the best region is that which contains those $\boldsymbol{\theta}$ values with minimum expected loss. A *Bayes q -credible region* $\Theta_q^*(\mathbf{z}) \subset \Theta$ is a q -credible region where any value within the region has a smaller posterior expected loss than any value outside the region, so that

$$\forall \boldsymbol{\theta}_i \in \Theta_q^*(\mathbf{z}), \forall \boldsymbol{\theta}_j \notin \Theta_q^*(\mathbf{z}), \quad \bar{\ell}(\boldsymbol{\theta}_i | \mathbf{z}) \leq \bar{\ell}(\boldsymbol{\theta}_j | \mathbf{z}).$$

The concept of a Bayes credible region was introduced by Bernardo in [4] under the name of *lower posterior loss* (LPL) credible regions.

The quadratic loss function yields credible regions which contain those values of $\boldsymbol{\theta}$ closest to the posterior expectation in the Euclidean distance sense. A zero-one loss function leads to highest posterior density (HPD) credible regions.

1.3.1.3 Precise hypothesis testing

Consider a value θ_0 of the vector of interest which deserves special consideration, either because assuming $\theta = \theta_0$ would noticeably simplify the model, or because there are additional context specific arguments suggesting that $\theta = \theta_0$. Intuitively, the value θ_0 should be judged to be *compatible* with the observed data \mathbf{z} if its posterior density $p(\theta_0 | \mathbf{z})$ is relatively high. However, a more precise form of conclusion is typically required.

Formally, testing the hypothesis $H_0 \equiv \{\theta = \theta_0\}$ may be described as a decision problem where the action space $\mathcal{A} = \{a_0, a_1\}$ contains only two elements: to accept (a_0) or to reject (a_1) the hypothesis under scrutiny. Foundations require to specify a loss function $\ell_h\{a_i, (\theta, \lambda)\}$ measuring the consequences of accepting or rejecting H_0 as a function of the actual parameter values. By assumption, a_0 means to *act as if* H_0 were true, that is to work with the submodel $\mathcal{M}_0 = \{p(\mathbf{z} | \theta_0, \lambda_0), \mathbf{z} \in \mathcal{Z}, \lambda_0 \in \Lambda\}$, while a_1 means to reject this simplification and to keep working with the full model $\mathcal{M}_z = \{p(\mathbf{z} | \theta, \lambda), \mathbf{z} \in \mathcal{Z}, \theta \in \Theta, \lambda \in \Lambda\}$. Alternatively, an already established model \mathcal{M}_0 may have been embedded into a more general model \mathcal{M}_z , constructed to include promising departures from $\theta = \theta_0$, and it is required to verify whether presently available data \mathbf{z} are compatible with $\theta = \theta_0$, or whether the extension to $\theta \in \Theta$ is really necessary. The optimal action will be to reject the hypothesis if (and only if) the expected posterior loss of accepting (a_0) is larger than that of rejecting (a_1), so that

$$\int_{\Theta} \int_{\Lambda} [\ell_h\{a_0, (\theta, \lambda)\} - \ell_h\{a_1, (\theta, \lambda)\}] p(\theta, \lambda | \mathbf{z}) d\theta d\lambda > 0.$$

Hence, only the difference $\Delta\ell_h\{\theta_0, (\theta, \lambda)\} = \ell_h\{a_0, (\theta, \lambda)\} - \ell_h\{a_1, (\theta, \lambda)\}$, which measures the *marginal advantage of rejecting* $H_0 \equiv \{\theta = \theta_0\}$ as a function of the parameter values, must be specified. The hypothesis H_0 should be rejected whenever the expected marginal advantage of rejecting is positive. Without loss of generality, the function $\Delta\ell_h$ may be written in the form

$$\Delta\ell_h\{\theta_0, (\theta, \lambda)\} = \ell\{\theta_0, (\theta, \lambda)\} - \ell_0$$

where (precisely as in estimation), $\ell\{\theta_0, (\theta, \lambda)\}$ describes, as a function of the parameter values which have generated the data, the non-negative loss to be suffered if θ_0 were used as a proxy for θ . Since $\ell\{\theta_0, (\theta_0, \lambda)\} = 0$, so that $\Delta\ell_h\{\theta_0, (\theta_0, \lambda)\} = -\ell_0$, the constant $\ell_0 > 0$ describes (in the same loss units) the context-dependent non-negative marginal advantage of accepting $\theta = \theta_0$ when it is true. With this formulation, the optimal action is to reject $\theta = \theta_0$ whenever the expected value of $\ell\{\theta_0, (\theta, \lambda)\} - \ell_0$ is positive, *i.e.*, whenever $\bar{\ell}(\theta_0 | \mathbf{z})$, the posterior expectation of $\ell\{\theta_0, (\theta, \lambda)\}$, is larger than ℓ_0 . Thus the solution is found in terms of the *same* expected loss function that was needed for estimation. The *Bayes test criterion* to decide on the compatibility of $\theta = \theta_0$ with available data \mathbf{z} is to reject $H_0 \equiv \{\theta = \theta_0\}$ if (and only if), $\bar{\ell}(\theta_0 | \mathbf{z}) > \ell_0$, where ℓ_0 is a context dependent positive constant.

Using the quadratic loss function leads to rejecting a θ_0 value whenever its Euclidean distance to the posterior expectation of θ is sufficiently large. The use of the (rather naive) zero-one loss function, $\ell\{\theta_0, (\theta, \lambda)\} = 0$ if $\theta = \theta_0$, and $\ell\{\theta_0, (\theta, \lambda)\} = 1$ otherwise, so that the loss advantage of rejecting θ_0 is a constant whenever $\theta \neq \theta_0$ and zero otherwise, leads to rejecting H_0 if (and only if) $\Pr(\theta = \theta_0 | \mathbf{z}) < p_0$ for some context-dependent p_0 . Notice however that, using this particular loss function *requires* the prior probability $\Pr(\theta = \theta_0)$ to be *strictly positive*; if θ is a continuous parameter this forces the use of a non-regular “sharp” prior, concentrating a positive probability mass at θ_0 , which would typically *not* be appropriate for estimation and will obviously depend on the particular θ_0 value to test. Foundations would suggest however that the *same* prior—which is supposed to describe the available knowledge about the parameter values—should be used for *any* aspect of the Bayesian analysis of the problem.

The threshold constant ℓ_0 —which is used to decide whether or not the expected loss $\bar{\ell}(\theta_0 | \mathbf{z})$ is too large—is part of the specification of the decision problem, and should be context-dependent. However, as demonstrated below, a judicious choice of the loss function leads to *self-calibrated* expected losses, where the relevant threshold constant has an immediate, operational interpretation.

1.3.2 Continuous invariant loss functions

The formulation above is *totally general*, and may be used with any loss function $\ell\{\theta_0, (\theta, \lambda)\}$ —which measures the loss to be suffered if a value θ_0 were used as a proxy for the true value θ —and any prior $p(\theta, \lambda)$ —which describes the available knowledge about the parameter values—to provide solutions for both estimation and hypothesis testing. If both the loss function and the prior distribution are *continuous*, precisely the *same* loss and the *same* prior may be used to obtain a coherent, integrated set of solutions for both estimation and testing, which may all be derived from the *joint* use of the corresponding posterior density $p(\theta | \mathbf{z})$, and the posterior expected loss function $\bar{\ell}(\theta_0 | \mathbf{z})$. Moreover the prior used may well be improper, as will typically be the case when an ‘objective’ analysis is required.

For most conventional loss functions, Bayes estimators are *not* invariant under one to one transformations. For example, the Bayes estimator of a variance under quadratic loss (its posterior expectation), is not the square of the Bayes estimator of the standard deviation. This is rather difficult to justify when, as it is the case in pure inference problems, one merely wishes to report an estimate of some quantity of interest.

Similarly, Bayes credible regions are generally *not* invariant under one to one transformations. Thus, HPD regions in one parameterization—obtained from a zero-one loss function—will *not* transform to HPD regions in another.

Rather more dramatically, Bayes test criteria are generally *not* invariant

under one-to-one transformations so that, if $\phi(\theta)$ is a one-to-one transformation of θ , rejecting $\theta = \theta_0$ does *not* generally imply rejecting the—logically equivalent—proposition $\phi(\theta) = \phi(\theta_0)$.

Invariant Bayes point estimators, credible regions and test procedures may all be easily obtained by using *invariant* loss functions, so that

$$\ell\{\theta_0, (\theta, \lambda)\} = \ell\{\phi(\theta_0), (\phi(\theta), \psi(\lambda))\}$$

for any one-to-one transformations $\phi(\theta)$ and $\psi(\lambda)$ of θ and λ , rather than conventional (non-invariant) loss functions such as the quadratic or the zero-one loss functions. A particularly interesting family of invariant loss functions is described below.

Conditional on model $\mathcal{M}_z = \{p(z | \theta, \lambda), z \in \mathcal{Z}, \theta \in \Theta, \lambda \in \Lambda\}$, the required loss function $\ell\{\theta_0, (\theta, \lambda)\}$ should describe, in terms of the unknown parameter values (θ, λ) which are assumed to have generated the data, the loss to be suffered if, in work with model \mathcal{M}_z , the value θ_0 were used as a proxy for θ . It may naively appear that what is needed is just some measure of the discrepancy between θ_0 and θ . However, since all parameterizations are arbitrary, what is really required is some measure of the discrepancy between the *models* labelled by θ and by θ_0 . By construction, such a discrepancy measure will be independent of the particular parameterization used. C. P. Robert [17] coined the word *intrinsic* to refer to these model-based loss functions; by construction, they are always invariant under one-to-one reparameterizations.

1.3.3 The intrinsic logarithmic loss

A particular *intrinsic* loss function with very attractive properties, the *logarithmic intrinsic loss*, is now introduced.

Let $\mathcal{M}_z = \{p(z | \theta, \lambda), z \in \mathcal{Z}\}$ be the model which is *assumed* to have generated the available data $z \in \mathcal{Z}$, where $\theta \in \Theta$ and $\lambda \in \Lambda$ are both unknown, and consider any other model $\mathcal{M}_0 = \{p(z | \omega_0), z \in \mathcal{Z}\}$, for some $\omega_0 \in \Omega_0$, with the same or larger support. The Kullback-Leibler [16] directed divergence of the probability density $p(z | \omega_0)$ from the probability density $p(z | \theta, \lambda)$,

$$\kappa\{p_z(\cdot | \omega_0) | p_z(\cdot | \theta, \lambda)\} = \int_{\mathcal{Z}} p(z | \theta, \lambda) \log \frac{p(z | \theta, \lambda)}{p(z | \omega_0)} dz,$$

is the average (under repeated sampling) log-likelihood ratio *against* the alternative model $p(z | \omega_0)$. This is known to be *nonnegative*, and zero if, and only if, $p(z | \omega_0) = p(z | \theta, \lambda)$ almost everywhere, and it is *invariant* under one-to-one transformations of either the data z or the parameters θ, λ and ω_0 . It is also *additive*, in the sense that if $z = \{x_1, \dots, x_n\}$ is assumed to a random sample from some model, then

$$\kappa\{p_z(\cdot | \omega_0) | p_z(\cdot | \theta, \lambda)\} = n \kappa\{p_x(\cdot | \omega_0) | p_x(\cdot | \theta, \lambda)\}.$$

And it is *invariant* under reduction to sufficient statistics in the sense that, if $\mathbf{t} \in \mathcal{T}$ is a sufficient statistic for both \mathcal{M}_z and \mathcal{M}_0 , then

$$\kappa\{p_z(\cdot | \omega_0) | p_z(\cdot | \boldsymbol{\theta}, \boldsymbol{\lambda})\} = \kappa\{p_{\mathbf{t}}(\cdot | \omega_0) | p_{\mathbf{t}}(\cdot | \boldsymbol{\theta}, \boldsymbol{\lambda})\}.$$

Definition 1 Intrinsic logarithmic loss function. Let z be the available data, let $\mathcal{M}_z = \{p(z | \boldsymbol{\theta}, \boldsymbol{\lambda}), z \in \mathcal{Z}\}$ be the model from which the data are assumed to have been generated, and let H_0 be the hypothesis that the data have actually been generated from a member of the family

$$\mathcal{M}_0 = \{p(z | \omega_0), \omega_0 \in \Omega_0, z \in \mathcal{Z}_0\}, \quad \mathcal{Z} \subseteq \mathcal{Z}_0$$

The intrinsic logarithmic loss function from assuming H_0 is the minimum average under sampling of the log-likelihood ratio against an element of \mathcal{M}_0 ,

$$\delta\{H_0 | \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathcal{M}_z\} = \inf_{\omega_0 \in \Omega_0} \kappa\{p_z(\cdot | \omega_0) | p_z(\cdot | \boldsymbol{\theta}, \boldsymbol{\lambda})\}.$$

Notice the complete generality of this definition. It may be used with either discrete or continuous data models (in the discrete case, the integrals will obviously be sums), and with either discrete or continuous parameter spaces, of any dimensionality.

The particular case which obtains when $H_0 \equiv \{\boldsymbol{\theta} = \boldsymbol{\theta}_0\}$, so that

$$\delta\{H_0 | \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathcal{M}_z\} = \delta_z\{\boldsymbol{\theta}_0 | \boldsymbol{\theta}, \boldsymbol{\lambda}\} = \inf_{\lambda_0 \in \Lambda_0} \int_{\mathcal{Z}} p(z | \boldsymbol{\theta}, \boldsymbol{\lambda}) \log \frac{p(z | \boldsymbol{\theta}, \boldsymbol{\lambda})}{p(z | \boldsymbol{\theta}_0, \lambda_0)} dz,$$

is an appropriate loss function for both point and region estimation of $\boldsymbol{\theta}$, and for testing whether or not a particular $\boldsymbol{\theta}_0$ value is compatible with the observed data.

The intrinsic logarithmic loss function $\delta\{H_0 | \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathcal{M}_z\}$ formalizes the use of log-likelihood ratios against the null to define a general loss function. With this loss structure, a precise hypothesis H_0 will be rejected if, and only if

$$d(H_0 | z) = \int_{\Theta} \int_{\Lambda} \delta\{H_0 | \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathcal{M}_z\} p(\boldsymbol{\theta}, \boldsymbol{\lambda} | z) d\boldsymbol{\theta} d\boldsymbol{\lambda} > \ell_0,$$

that is if, and only if, the posterior expectation of the average log-likelihood ratio loss—which estimates the minimum log-likelihood ratio against H_0 —is larger than a suitably chosen constant ℓ_0 . In particular, if $\ell_0 = \log[R]$, then H_0 would be rejected whenever, given the observed data, the minimum average likelihood ratio against H_0 , may be expected to be larger than about R . Conventional choices for ℓ_0 are $\{\log 20, \log 100, \log 1000\} \approx \{3.0, 4.6, 6.9\}$.

In a multivariate normal model with known covariance matrix the intrinsic logarithmic loss is proportional to the Mahalanobis distance. Thus, if z is a random sample of size n from a k -variate normal distribution $N(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\Sigma})$,

$$\delta_z\{\boldsymbol{\mu}_0 | \boldsymbol{\mu}, \boldsymbol{\Sigma}\} = \frac{n}{2} (\boldsymbol{\mu}_0 - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_0 - \boldsymbol{\mu}),$$

which is $n/2$ times the Mahalanobis distance between $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}$. This result may be used to obtain large-sample approximations to the intrinsic logarithmic loss. In particular, if \boldsymbol{z} is a random sample of size n from the single parameter model $p(x|\theta)$, and $\tilde{\theta}_n = \tilde{\theta}_n(\boldsymbol{z})$ is an asymptotically sufficient consistent estimator of θ whose sampling distribution is asymptotically normal with standard deviation $s(\theta)/\sqrt{n}$, then, for large values of n ,

$$\delta_{\boldsymbol{z}}\{\theta_0|\theta, \mathcal{M}_{\boldsymbol{z}}\} \approx \frac{n}{2}[\phi(\theta_0) - \phi(\theta)]^2,$$

where $\phi(\theta) = \int^\theta s(y)^{-1}dy$ is the corresponding variance stabilization transformation.

1.4 Intrinsic reference analysis

The decision-theoretic procedures described above to derive summaries for Bayesian inference are totally general, so that they may be used with any loss function and any prior distribution. The advantages of using the intrinsic logarithmic loss have been described above: it is invariant under both reparameterization and reduction to sufficient statistics, and—most important—it has a simple operational interpretation in terms of average log-likelihood ratios against the null, so it is self-calibrated in terms of simple log-likelihood ratios.

1.4.1 Intrinsic reference estimation and testing

Foundations indicate that the prior distribution should describe available prior knowledge. In many situations however, either the available prior information is too vague to warrant the effort required to formalize it, or it is too subjective to be useful in scientific communication. An “objective” procedure, where the prior function is intended to describe a situation where there is no relevant information about the quantity of interest, is therefore often required. Objectivity is an emotionally charged word, and it should be explicitly qualified whenever it is used. No statistical analysis is really objective, since both the experimental design and the model assumed have very strong subjective inputs. However, frequentist procedures are often branded as “objective” just because their conclusions are only conditional on the model assumed and the data obtained. Bayesian methods where the prior function is directly derived from the assumed model are objective in this limited, but precise sense. There is a vast literature devoted to the formulation of objective priors. Reference analysis, introduced by Bernardo in [3], and further developed in [1], [2], [10], [11] and references therein, is probably the most popular approach for deriving objective priors.

Reference priors may be numerically obtained (see [10] for details) but, under appropriate regularity conditions, explicit formulae for the reference priors are readily available (see [2] and [9] for details). In particular, if the posterior distribution of θ given a random sample of size n from $p(x|\theta)$ is asymptotically normal with standard deviation $s(\tilde{\theta}_n)/\sqrt{n}$, where $\tilde{\theta}_n$ is a consistent estimator of θ , then the reference prior is $\pi(\theta) = s(\theta)^{-1}$. This includes the well-known one-parameter Jeffreys prior

$$\pi(\theta) \propto i(\theta)^{1/2}, \quad i(\theta) = \mathbb{E}_{\mathbf{x}|\theta}[-\partial^2 \log p(\mathbf{z}|\theta)/\partial\theta^2],$$

as a particular case.

For objective Bayesian solutions to inferential problems the *combined* use of the *intrinsic logarithmic loss* function and the relevant *reference prior* are recommended. The corresponding Bayes point estimators, Bayes credible regions and Bayes test criteria are respectively referred to as *intrinsic reference* estimators, credible regions or test criteria. The basic ideas were respectively introduced in [7], [5] and [8].

All inference summaries depend on the data only through the expected reference intrinsic loss, $d(\boldsymbol{\theta}_0|\mathbf{z})$, the expectation of the intrinsic loss with respect to the appropriate joint reference posterior,

$$d(\boldsymbol{\theta}_0|\mathbf{z}) = \int_{\Theta} \int_{\Lambda} \delta\{\boldsymbol{\theta}_0|\boldsymbol{\theta}, \boldsymbol{\lambda}, \mathcal{M}_{\mathbf{z}}\} \pi(\boldsymbol{\theta}, \boldsymbol{\lambda}|\mathbf{z}) d\boldsymbol{\theta} d\boldsymbol{\lambda}.$$

Most other *intrinsic* loss functions (invariant, continuous loss functions which measure the discrepancy between the models rather than the discrepancy between their parameters) would yield qualitatively similar results, but attention will here be confined to the *intrinsic logarithmic* loss defined above, for this is often easily derived, and—most important—it is self-calibrated in terms of easily interpretable log-likelihood ratios against the null.

The following example is intended to illustrate the general procedure:

1.4.2 Example: the normal variance

Let $\mathbf{z} = \{x_1, \dots, x_n\}$ be a random sample from a normal $N(x|\mu, \sigma)$ distribution whose variance σ^2 is of interest. Since reference analysis is invariant under one-to-one transformations, one may equivalently work in terms of σ , $\log \sigma$, or any other one-to-one transformation of σ . The Kullback-Leibler discrepancy of $p(\mathbf{z}|\mu_0, \sigma_0)$ from $p(\mathbf{z}|\mu, \sigma)$ is given by

$$\begin{aligned} \kappa\{N_{\mathbf{z}}(\cdot|\mu_0, \sigma_0) | N_{\mathbf{z}}(\cdot|\mu, \sigma)\} &= n \int_{\mathfrak{R}} N(x|\mu, \sigma) \log \frac{N(x|\mu, \sigma)}{N(x|\mu_0, \sigma_0)} dx \\ &= \frac{n}{2} \left[\log \frac{\sigma_0^2}{\sigma^2} + \frac{\sigma^2}{\sigma_0^2} - 1 + \frac{(\mu - \mu_0)^2}{\sigma_0^2} \right], \end{aligned}$$

which is minimized when $\mu_0 = \mu$. Hence, the intrinsic logarithmic loss function for $H_0 \equiv \{\sigma = \sigma_0\}$ is

$$\delta\{H_0 \mid \sigma, \mu, \mathcal{M}_{\mathbf{z}}\} = \delta_{\mathbf{z}}\{\sigma_0 \mid \sigma, \mu\} = \frac{n}{2} \left[\log \frac{\sigma_0^2}{\sigma^2} + \frac{\sigma^2}{\sigma_0^2} - 1 \right].$$

Since the normal is a location-scale model, the reference prior is the conventional improper prior $\pi(\mu, \sigma) = \sigma^{-1}$. The corresponding reference posterior density of σ , after a random sample $\mathbf{z} = \{x_1, \dots, x_n\}$ of size $n \geq 2$ has been observed, which is always proper, is

$$\pi(\sigma \mid \mathbf{z}) = \pi(\sigma \mid n, s^2) = \frac{(ns^2)^{(n-1)/2}}{2^{(n-3)/2} \Gamma[(n-1)/2]} \sigma^{-n} \exp\left[-\frac{ns^2}{2\sigma^2}\right],$$

where $s^2 = n^{-1} \sum_{j=1}^n (x_j - \bar{x})^2$ is the MLE of σ^2 .

The corresponding *reference posterior expected loss* from using σ_0 as a proxy for σ , given a random sample of size n , is

$$\begin{aligned} d(\sigma_0 \mid \mathbf{z}) &= \int_0^\infty n \delta_{\mathbf{z}}\{\sigma_0 \mid \sigma, \mu\} \pi(\sigma \mid \mathbf{z}) d\sigma \\ &= \frac{n}{2} \left[\psi\left(\frac{n-1}{2}\right) - 1 + \frac{ns^2}{(n-3)\sigma_0^2} + \log\left(\frac{2\sigma_0^2}{ns^2}\right) \right]. \end{aligned}$$

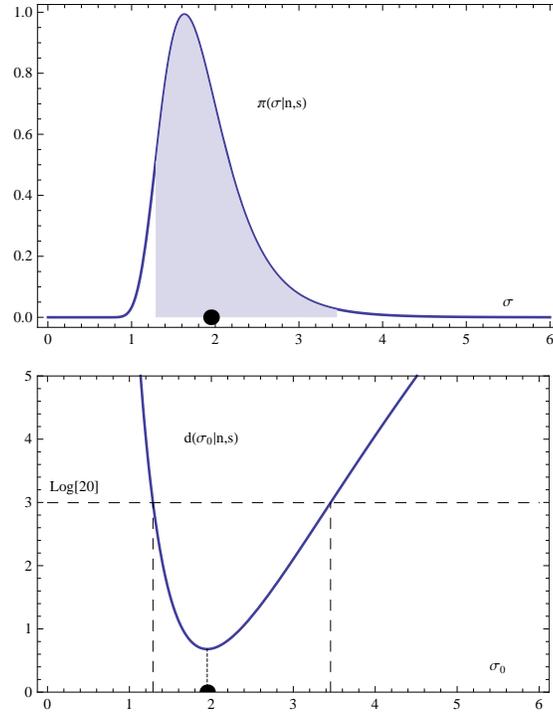
By definition, the Bayes point estimator with respect to this loss function, the intrinsic reference estimator of σ is that value of σ_0 which minimizes $d(\sigma_0 \mid \mathbf{z})$; this is found to be $\sigma^*(\mathbf{z}) = \sqrt{n}/(s\sqrt{n-3})$. Thus, the intrinsic reference estimator of the variance is

$$\sigma^{2*}(\mathbf{z}) = \frac{ns^2}{n-3},$$

an estimator already suggested by Stein [18], which is always larger than both the MLE and the conventional unbiased estimator, that respectively divide the sum of squares ns^2 by n and by $n-1$; for small samples, the differences are noticeable. Since intrinsic estimation is consistent under one-to-one reparametrizations, the intrinsic reference estimator of, say, $\log \sigma$ is simply $\log[\sigma^*(\mathbf{z})]$.

As an illustration, a random sample \mathbf{z} of size $n = 10$ was simulated from a normal distribution with $\mu = 1$ and $\sigma = 2$, yielding $\bar{x} = 0.951$ and $s = 1.631$. Intrinsic reference analysis of σ is well summarized by two complementary functions: (i) the reference posterior density $\pi(\sigma \mid \mathbf{z})$, and (ii) the expected posterior intrinsic logarithmic loss $d(\sigma_0 \mid \mathbf{z})$ of using σ_0 as a proxy for σ . Figure 1.1 represents both $\pi(\sigma \mid \mathbf{z})$ (upper panel) and $d(\sigma_0 \mid \mathbf{z})$ (lower panel), in the same horizontal scale.

The expected intrinsic logarithmic loss $d(\sigma_0 \mid \mathbf{z})$ is minimized at the

**FIGURE 1.1**

Intrinsic reference analysis for the standard deviation of a normal distribution, given a random sample of size $n = 5$, with $s = 1.631$.

reference intrinsic estimate, $\sigma^* = 1.949$, represented in both panels by a black dot. Thus, the intrinsic estimator of the variance is $\sigma^{*2} = 3.80$, which may be compared with the MLE $s^2 = 2.66$, or with the conventional unbiased estimator $\hat{\sigma}^2 = 2.96$.

To test if a particular σ_0 value is supported by the data one simply checks its expected loss. For example, all σ_0 values smaller than 1.29 or larger than 3.46 have an expected intrinsic logarithmic loss larger than $\log(20) \approx 3.0$, and would be rejected if the threshold were set to reject σ_0 values with an expected log-likelihood ratio against the true (unknown) value of σ larger than $\log(20)$, suggesting that the average likelihood under the true model may be expected to be at least 20 times larger than that under any model with $\sigma = \sigma_0$. The corresponding acceptance region, the interval (1.29, 3.46), shaded area in the upper panel) has a posterior probability of 0.92. Since all elements within that region have smaller expected loss than those outside, this is an intrinsic reference 0.92-credible region. By definition, this region is invariant under transformations; thus, the 0.92 reference intrinsic region for σ^2

is simply $(1.29^2, 3.46^2)$. Notice, that these regions are *not* HPD regions. The intrinsic reference 0.95-credible interval is $(1.24, 3.74)$.

The expected logarithmic intrinsic loss of using $\sigma_0 = 5$ as a proxy for σ is $d(5 | \mathbf{z}) = 5.86 \approx \log[350]$, so that the hypothesis $H_0 \equiv \{\sigma = 5\}$ would typically be rejected on the grounds that the likelihood ratio *against* H_0 may be expected to be at least 350.

1.5 Comparing Binomial proportions

Consider two random samples of sizes n_1 and n_2 from independent binomial populations with parameters θ_1 and θ_2 , respectively yielding r_1 and r_2 successes. Thus, the data are $\mathbf{z} = \{(r_1, n_1), (r_2, n_2)\}$, the unknown parameter vector is $\boldsymbol{\theta} = \{\theta_1, \theta_2\}$, and the sampling model is

$$p(\mathbf{z} | \boldsymbol{\theta}) = \text{Bi}(r_1 | n_1, \theta_1) \text{Bi}(r_2 | n_2, \theta_2) \propto \theta_1^{r_1} (1 - \theta_1)^{n_1 - r_1} \theta_2^{r_2} (1 - \theta_2)^{n_2 - r_2}$$

Interest focuses in comparing θ_1 and θ_2 and, more specifically, in deciding whether or not there is evidence against the hypothesis $H_0 \equiv \{\theta_1 = \theta_2\}$ that the two proportions are actually equal.

1.5.1 Intrinsic logarithmic loss to test equality

The Kullback-Leibler discrepancy of a model $p(\mathbf{z} | \alpha)$ with equal parameters $\theta_1 = \theta_2 = \alpha$, so that $p(\mathbf{z} | \alpha) = \text{Bi}(r_1 | n_1, \alpha) \text{Bi}(r_2 | n_2, \alpha)$, from the assumed model $p(\mathbf{z} | \boldsymbol{\theta}) = \text{Bi}(r_1 | n_1, \theta_1) \text{Bi}(r_2 | n_2, \theta_2)$, is the expected value under the true model $p(\mathbf{z} | \boldsymbol{\theta})$ of the corresponding log-likelihood ratio ratio,

$$\begin{aligned} \kappa\{p_{\mathbf{z}}(\cdot | \alpha) | p_{\mathbf{z}}(\cdot | \boldsymbol{\theta})\} &= \mathbf{E}_{\mathbf{z} | \boldsymbol{\theta}} \left[\log \frac{p(\mathbf{z} | \boldsymbol{\theta})}{p(\mathbf{z} | \alpha)} \right] \\ &= n_1 \log(1 - \theta_1) + n_2 \log(1 - \theta_2) \\ &\quad - [n_1(1 - \theta_1) + n_2(1 - \theta_2)] \log[1 - \alpha] \\ &\quad - n_1 \theta_1 \log \frac{\alpha \theta_1}{1 - \theta_1} - n_2 \theta_2 \log \frac{\alpha \theta_2}{1 - \theta_2}, \end{aligned}$$

which is minimized when

$$\alpha = \frac{n_1 \theta_1 + n_2 \theta_2}{n_1 + n_2}.$$

The intrinsic logarithmic loss function for $H_0 \equiv \{\theta_1 = \theta_2\}$ is

$$\delta\{H_0 | \theta_1, \theta_2, n_1, n_2\} = \min_{\alpha \in (0,1)} \kappa\{p_{\mathbf{z}}(\cdot | \alpha) | p_{\mathbf{z}}(\cdot | \boldsymbol{\theta})\},$$

and substitution of the minimizing value of α yields

$$\begin{aligned} \delta\{H_0 | \theta_1, \theta_2, n_1, n_2\} &= n_1 \log(1 - \theta_1) + n_2 \log(1 - \theta_2) \\ &\quad + n_1 \theta_1 \log\left[\frac{\theta_1}{1 - \theta_1}\right] + n_2 \theta_2 \log\left[\frac{\theta_2}{1 - \theta_2}\right] \\ &\quad - (n_1 \theta_1 + n_2 \theta_2) \log\left[\frac{n_1 \theta_1 + n_2 \theta_2}{n_1 + n_2}\right] \\ &\quad - (n_1(1 - \theta_1) + n_2(1 - \theta_2)) \log\left[1 - \frac{n_1 \theta_1 + n_2 \theta_2}{n_1 + n_2}\right], \end{aligned}$$

a non-negative function of (θ_1, θ_2) which is zero if, and only if, $\theta_1 = \theta_2$ and reaches a maximum

$$\delta_{max} = (n_1 + n_2) \log(n_1 + n_2) - n_1 \log n_1 - n_2 \log n_2$$

when $(\theta_1, \theta_2) = (0, 1)$ or $(\theta_1, \theta_2) = (1, 0)$, which reduces to $2n \log[2]$ when $n_1 = n_2 = n$.

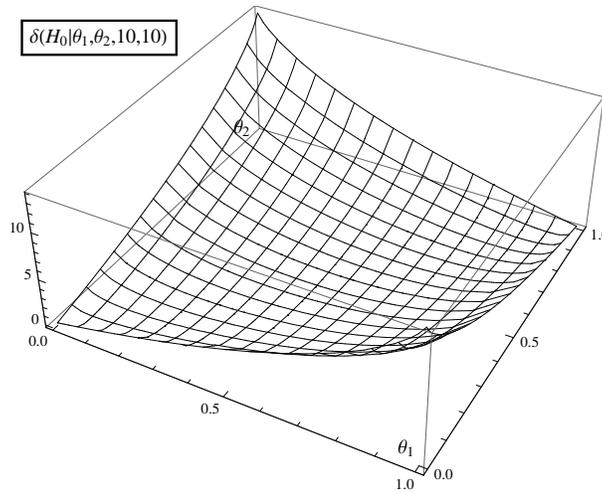


FIGURE 1.2

Intrinsic logarithmic loss function for $H_0 = \{\theta_1 = \theta_2\}$ when $n_1 = n_2 = 10$.

Figure 1.2 represents $\delta\{H_0 | \theta_1, \theta_2, n_1, n_2\}$ when $n_1 = n_2 = 10$. The function $\delta\{H_0 | \theta_1, \theta_2, n_1, n_2\}$ unambiguously and precisely describes, as a function of θ_1 and θ_2 , the discrepancy of the hypothesis $H_0 \equiv \{\theta_1 = \theta_2\}$ from the model $\text{Bi}(r_1 | n_1, \theta_1) \text{Bi}(r_2 | n_2, \theta_2)$.

1.5.2 Reference posterior distributions

Objective solutions to inferences about the possible differences between θ_1 and θ_2 require the use of an objective joint prior $\pi(\theta_1, \theta_2)$. In single parameter

problems, the reference prior is uniquely defined, and it is invariant under reparameterization. In multiparameter models however, the reference prior depends on the quantity of interest.

In this problem there are several possible choices for the quantity of interest. A clear option would be the intrinsic logarithmic loss function itself, for this precisely measures the discrepancy of H_0 for the model, and one is interested in checking whether or not this might be zero. Thus, one could define $\phi_0(\theta_1, \theta_2) = \delta\{H_0 | \theta_1, \theta_2, n_1, n_2\}$ and proceed to derive the reference prior $\pi_{\phi_0}(\theta_1, \theta_2)$ when ϕ_0 is the quantity of interest; this is a non trivial exercise, but it may be done. However other options more easily interpretable by the user suggest themselves, as the difference $\phi_1(\theta_1, \theta_2) = \theta_1 - \theta_2$, or the ratio $\phi_2(\theta_1, \theta_2) = \theta_1/\theta_2$.

Indeed, in models with many parameters, there are many situations where one is simultaneously interested in several functions of them, and it would then be useful to have a *single* objective prior which could safely be used to produce reasonable marginal posteriors for all the quantities of interest. Berger, Bernardo and Sun propose in [12] a criterium to select an *overall* joint prior function which may be considered a good approximate joint reference prior, in the sense that, for all data sets, it may be expected to produce marginal posteriors for all the quantities of interest which are not too different from the relevant reference posteriors. In situations where independent binomial situations are considered this leads to the use of the corresponding reference priors for each of the binomial models considered, which are known to be the relevant (proper) Jeffreys priors, $\pi(\theta_i) = \text{Be}(\theta_i | \frac{1}{2}, \frac{1}{2})$. In the case discussed here, this reduces to

$$\begin{aligned} \pi(\theta_1, \theta_2) &= \text{Be}(\theta_1 | \frac{1}{2}, \frac{1}{2}) \text{Be}(\theta_2 | \frac{1}{2}, \frac{1}{2}) \\ &= \pi^{-2} \theta_1^{-1/2} (1 - \theta_1)^{-1/2} \theta_2^{-1/2} (1 - \theta_2)^{-1/2}, \end{aligned}$$

and this is the *overall* objective prior suggested to analyze this problem. The corresponding joint reference posterior is

$$\begin{aligned} \pi(\theta_1, \theta_2 | \mathbf{z}) &= \pi(\theta_1, \theta_2 | r_1, r_2, n_1, n_2) \\ &= \text{Be}(\theta_1 | r_1 + \frac{1}{2}, n_1 - r_1 + \frac{1}{2}) \text{Be}(\theta_2 | r_2 + \frac{1}{2}, n_2 - r_2 + \frac{1}{2}), \end{aligned}$$

and the expected logarithmic intrinsic loss is

$$d(H_0 | \mathbf{z}) = \int_0^1 \int_0^1 \delta\{H_0 | \theta_1, \theta_2, n_1, n_2\} \pi(\theta_1, \theta_2 | r_1, r_2, n_1, n_2) d\theta_1 d\theta_2,$$

which may easily be numerically evaluated.

1.5.3 The RV144 HIV vaccine efficacy trial in Thailand

In 2009, the RV144 randomized, double-blind, efficacy trial in Thailand reported that a prime-boost human immunodeficiency virus (HIV) vaccine

regimen conferred about 30% protection against HIV acquisition, but different analyses seemed to give conflicting results, and a heated debate followed as scientists and the broader public struggled with their interpretation; see [15] for a detailed description of the issues involved. The main result concerned individuals in the general population in Thailand, mostly at heterosexual risk, 61% of which were men, randomized within the “intention to treat” population, excluding subjects found to be HIV positive at the time of randomization. The press release reported $r_1 = 51$ infected among $n_1 = 8197$ who have taken the vaccine, to be compared with $r_2 = 74$ infected among $n_2 = 8198$ who have taken a placebo. Using conventional frequentist testing, the two corresponding binomial parameters, θ_1 and θ_2 , were said to be significantly different, with a p -value of 0.04. Moreover, the results suggested an *estimated* vaccine efficacy (one minus the relative hazard rate of HIV in the vaccine versus placebo group) of

$$\text{VE}(r_1, n_1, r_2, n_2) = 1 - \frac{r_1/n_1}{r_2/n_2} = 0.31.$$

An objective Bayesian, intrinsic reference analysis of these data will now be provided.

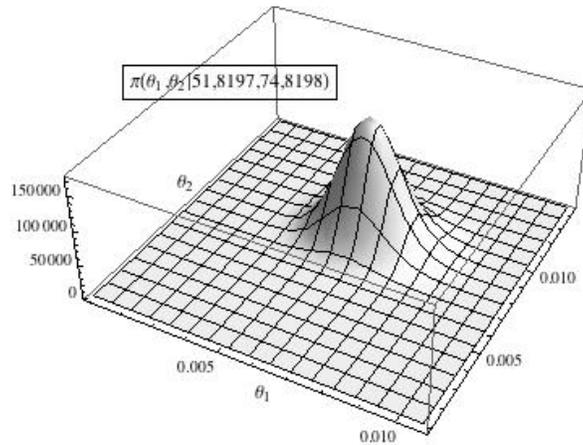


FIGURE 1.3

Joint posterior reference density $\pi(\theta_1, \theta_2 | r_1, r_2, n_1, n_2)$, when $r_1 = 51$, $n_1 = 8197$, $r_2 = 74$ and $n_2 = 8198$.

The joint reference posterior distribution for θ_1 and θ_2 which corresponds to the overall reference prior $\pi(\theta_1, \theta_2) = \text{Be}(\theta_1 | \frac{1}{2}, \frac{1}{2}) \text{Be}(\theta_2 | \frac{1}{2}, \frac{1}{2})$ is $\pi(\theta_1, \theta_2 | r_1, r_2, n_1, n_2) = \text{Be}(\theta_1 | 51.5, 8146.5) \text{Be}(\theta_2 | 74.5, 8124.5)$, represented in Figure 1.3, which has with a unique mode at (0.0063, 0.0091).

The reference posterior probability that the proportion θ_1 of infected among those vaccinated is actually smaller than the proportion θ_2 of infected

among those take the placebo is simply

$$\Pr[\theta_1 < \theta_2 | \mathbf{z}] = \int_0^1 \int_0^{\theta_2} \pi(\theta_1, \theta_2 | \mathbf{z}) d\theta_1 d\theta_2 = 0.981,$$

so there seems to be some—not overwhelming—evidence that indeed, θ_1 may be smaller than θ_2 .

The posterior expectation of the intrinsic logarithmic loss corresponding to the hypothesis $H_0 = \{\theta_1 = \theta_2\}$ that both parameters are actually equal is

$$d(H_0 | \mathbf{z}) = \int_0^1 \int_0^1 \delta\{H_0 | \boldsymbol{\theta}, n_1, n_2\} \pi(\theta_1, \theta_2 | \mathbf{z}) d\theta_1 d\theta_2 = 2.624 = \log[13.8].$$

Thus, given the data, it is estimated that the *average* log-likelihood ratio

$$\delta\{H_0 | \boldsymbol{\theta}, n_1, n_2\} = \inf_{\boldsymbol{\theta}_0 \in H_0} E_{\mathbf{z} | \boldsymbol{\theta}} \left[\log \frac{p(\mathbf{z} | \boldsymbol{\theta})}{p(\mathbf{z} | \boldsymbol{\theta}_0)} \right]$$

for the model who has generated to data against any model within H_0 is at least $2.624 = \log[13.8]$; hence, the observed data may be expected to be about 14 more likely under the true model than under a model within H_0 . This certainly indicates that there is some evidence on the existence of a difference between the two hazard rates and that the vaccine may well be effective, but the evidence is far from strong.

Standard probability calculus may be used to derive the reference posterior distribution of the actual efficacy of the vaccine,

$$\phi(\theta_1, \theta_2) = 1 - \frac{\theta_1}{\theta_2},$$

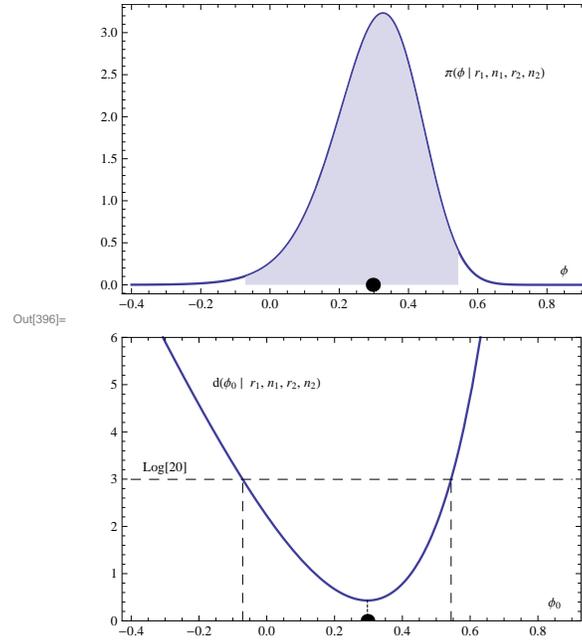
from the joint reference posterior $\pi(\theta_1, \theta_2 | \mathbf{z})$ to obtain

$$\begin{aligned} \pi(\phi | \mathbf{z}) &= \int_0^1 \pi(\theta_1, \theta_2 | \mathbf{z}) \theta_2 |_{\theta_1 \rightarrow \theta_2(1-\phi)} d\theta_2 \\ &= \frac{n_1! n_2! (r_1 + r_2)!}{\Gamma[r_1 + 1/2] \Gamma[r_2 + 1/2] \Gamma[n_1 - r_1 + 1/2]} (1 - \phi)^{r_1 - 1/2} \\ &\quad \times {}_2F_1^R[1/2 - n_1 + r_1, 1 + r_1 + r_2, 3/2 + n_2 + r_1, 1 - \phi], \end{aligned}$$

where ${}_2F_1^R(a, b, c, z)$ is the regularized ${}_2F_1$ hypergeometric function.

The posterior density of $\phi(\theta_1, \theta_2) = 1 - (\theta_1/\theta_2)$ which corresponds to the vaccine trial data is represented in the upper panel of Figure 1.4. The lower panel represents $d(\phi_0 | \mathbf{z})$, the reference posterior expectation of the corresponding intrinsic logarithmic loss,

$$\delta\{\phi_0 | \phi, \theta_2, n_1, n_2\} = \inf_{\theta_{20} \in [0,1]} E_{\mathbf{z} | \phi, \theta_2} \left[\log \frac{p(\mathbf{z} | \phi, \theta_2)}{p(\mathbf{z} | \phi_0, \theta_{20})} \right],$$

**FIGURE 1.4**

Reference posterior analysis of the vaccine efficacy $1 - (\theta_1/\theta_2)$ for the RV144 vaccine efficacy trial data.

which is given by

$$d(\phi_0 | \mathbf{z}) = \int_{-\infty}^{\infty} \int_0^1 \delta\{\phi_0 | \phi, \theta_2, n_1, n_2\} \pi(\phi, \theta_2 | \mathbf{z}) d\theta_2 d\phi.$$

and precisely describes the loss to be expected (in self-calibrated average log-likelihood ratio terms) if a particular value for the vaccine efficacy ϕ_0 were used as a proxy for the true unknown value of ϕ . This is minimized at $\phi^* = 0.297$ which is therefore the intrinsic reference estimate of the vaccine efficacy (represented by a solid point in both panels). The values within the interval $(-0.071, 0.544)$ have all an expected loss smaller than $\log[20]$ so they could possibly be accepted as proxies for ϕ in that the expected log-likelihood ratio against them would be smaller than $\log[20]$. Notice that this includes the value $\phi_0 = 0$ of zero efficacy. The region has a reference posterior probability of 0.97, and thus provides a intrinsic reference 0.97-credible interval for ϕ (shaded region in the upper panel). The intrinsic reference 0.95-credible interval is $(-0.009, 0.514)$, which also contains $\phi_0 = 0$.

All these results elaborate on the basic conclusion already provided by the computation of $d(H_0 | \mathbf{z}) = \log[13.8]$, which clearly indicates a *rather weak*

evidence against the hypothesis H_0 that there is no difference between the two hazard rates.

There is certainly some suggestion of a vaccine efficacy of about 30%, but the true value of the efficacy could really be anywhere between about -1% and 50% , so that—against the “firm conclusion” of an existing difference between the parameters apparently implied by the $p = 0.04$ frequentist significance—more information is really necessary before any final answers could possibly be reached.

1.6 Discussion

As described in Section 1.2, standard Bayesian decision theory provides a unified framework where the problems posed by point estimation, region estimation and hypothesis testing, may all be coherently solved within the same structure. Although the formulation is totally general, there are clear advantages in the use of a *continuous* loss function and a *continuous* prior (which may well be improper) that, for consistency, should both be the same in all those problems.

We have argued that one should preferably make use of *intrinsic* loss functions, for those have the required invariance properties and thus provide solutions which are invariant under reparameterization. Among intrinsic loss functions, there are important arguments to choose that derived from the average log-likelihood ratio against the null, the *intrinsic logarithmic loss*, for this is self-calibrated and has an immediate interpretation.

The *combined* use of the *intrinsic logarithmic loss function* and an *overall reference prior* provides the elements for an *integrated Bayesian reference analysis*, including both the posterior densities of the quantities of interest, and the expected logarithmic intrinsic losses associated to any alternative values. In particular this provides an immediate solution to any precise hypothesis testing problem in terms of the estimated (expected posterior) average log-likelihood ratio against the null.

As illustrated by the analysis of the RV144 HIV vaccine efficacy trial in Thailand, the solutions proposed are not difficult to obtain, they have a simple, very intuitive interpretation, and have far-reaching consequences, which in hypothesis testing problems often contradict conventional statistical practice.



Bibliography

- [1] J. O. Berger and J. M. Bernardo. Estimating a product of means: Bayesian analysis with reference priors. *J. Amer. Statist. Assoc.*, 84:200–207, 1989.
- [2] J. O. Berger and J. M. Bernardo. On the development of reference priors (with discussion). In *Bayesian Statistics 4* (J.M. Bernardo, J.O. Berger, A. P. Dawid and A.F.M. Smith, Eds.), pages 35–60. Oxford University Press, Oxford, UK, 1992.
- [3] J. M. Bernardo. Reference posterior distributions for bayesian inference (with discussion). *J. Roy. Statist. Soc.*, 41:113–147, 1979.
- [4] J. M. Bernardo. Intrinsic credible regions: An objective bayesian approach to interval estimation (with discussion). *Test*, 14:317–384, 2005.
- [5] J. M. Bernardo. Reference analysis. In *Bayesian Thinking: Modeling and Computation, Handbook of Statistics 25* (Dey, D. K. and Rao, C. R., Eds.), pages 465–476. Elsevier, Amsterdam, The Netherlands, 2005.
- [6] J. M. Bernardo. Integrated objective bayesian estimation and hypothesis testing (with discussion). In *Bayesian Statistics 9* (J.M. Bernardo, M.J. Bayarri, J.O. Berger, A. P. Dawid, D. Heckerman, A.F.M. Smith and M. West, Eds.), pages 1–68. Oxford University Press, Oxford, UK, 2011.
- [7] J. M. Bernardo and M. A. Juárez. Intrinsic estimation. In *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.), pages 465–476. Oxford University Press, Oxford, UK, 2003.
- [8] J. M. Bernardo and R. Rueda. Bayesian hypothesis testing: A reference approach. *Internat. Statist. Rev.*, 70:351–372, 2002.
- [9] J. M. Bernardo and A. F. M. Smith. *Bayesian Theory*. Wiley, Chichester, UK, 1994.
- [10] J. M. Bernardo J. O Berger and D. Sun. The formal definition of reference priors. *Ann. Statist.*, 37:905–938, 2009.
- [11] J. M. Bernardo J. O Berger and D. Sun. Objective priors for discrete parameter spaces. *J. Amer. Statist. Assoc.*, 107:636–648, 2012.

- [12] J. M. Bernardo J. O Berger and D. Sun. Overall reference priors. *Tech. Rep.*, pages Duke University, USA, 2013.
- [13] H. Jeffreys. *Theory of Probability* (3rd ed.). Oxford University Press, Oxford, UK, 1961.
- [14] R. E. Kass and L. Wasserman. The selection of prior distributions by formal rules. *J. Amer. Statist. Assoc.*, 1996.
- [15] P. B. Gilber J. O. Berger D. Stablein S. Becker M. Essex S. M. Hammer J. H. Kim and V. G. DeGruttola. Statistical interpretation of the rv144 hiv vaccine efficacy trial in thailand: A case study for statistical issues in efficacy trials. *J. Infect. Dis.*, 203:969–975, 2011.
- [16] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22:79–86, 1951.
- [17] C. P. Robert. Intrinsic loss functions. *Theory and Decision*, 40:192–214, 1996.
- [18] C. Stein. Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Instr. Statist. Math.*, 1964.